



This is a repository copy of *A systematic review of the validity and responsiveness of EQ-5D and SF-6D for depression and anxiety.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/74659/>

Article:

Peasgood, T., Brazier, J. and Papaioannou, D. (2012) A systematic review of the validity and responsiveness of EQ-5D and SF-6D for depression and anxiety. HEDS Discussion paper 12/15. (Unpublished)

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



HEDS Discussion Paper

No.12.15

A systematic review of the validity and responsiveness of EQ-5D and SF-6D for depression and anxiety

¹Tessa Peasgood, ¹John Brazier, ¹Diana Papaioannou

1. Health Economics and Decision Science, School of Health and Related Research, University of Sheffield

Disclaimer:

This series is intended to promote discussion and to provide information about work in progress. The views expressed in this series are those of the authors, and should not be quoted without their permission. Comments are welcome, and should be sent to the corresponding author.

White Rose Repository URL for this paper: <http://eprints.whiterose.ac.uk/74659>

White Rose Research Online
eprints@whiterose.ac.uk

A systematic review of the validity and responsiveness of EQ-5D and SF-6D for depression and anxiety

Tessa Peasgood, John Brazier, Diana Papaioannou

The University of Sheffield

October 2012

Abstract

Background: Generic preference based measures (PBM) such as the SF-6D and EQ-5D are increasingly used to inform health care resource allocation decisions. They aim to be generic in the sense of being applicable to all physical and mental health conditions. However, their applicability has not been demonstrated for all mental health conditions.

Aims: To assess the construct validity and responsiveness of EQ-5D and SF-6D measures in depression and anxiety.

Method: A systematic review of the literature was undertaken. Eleven databases were searched in December 2010 and reference lists scrutinised to identify relevant studies. Studies were appraised and data extracted. A narrative synthesis was performed of the evidence on construct validity including known groups validity (detecting a difference in PBM scores between different groups such as different levels of severity of depression), convergent validity (strength of association between generic PBM and other outcome measures) and responsiveness (the ability to detect relevant health changes in health status and the absence of change where there is none).

Results: 26 studies were identified that provided data on the validity and/or responsiveness of the EQ-5D and SF-6D. Both measures demonstrate good construct validity and responsiveness for depression. One study, however, suggests EQ-5D may lack responsiveness in the elderly. These measures are more highly correlated with depression scales in patients with anxiety than they are clinical anxiety scales suggesting known group validity in patients with anxiety may be driven by aspects of depression within anxiety disorder and the presence of co-morbid depression. Direct comparisons between the measures find that the EQ-5D gives lower utility levels for severe depression hence greater health improvement for this group and SF-6D shows more sensitivity to mild depression and performs better in terms of ES and SRM. The comparison between EQ-5D and SF-6D is similar to that found in other conditions.

Conclusion: The evidence base supports the use of EQ-5D and SF-6D in patients with depression and anxiety. More work is needed on the true utility level for severe depression.

List of abbreviations

ACQ	14-item self-report instrument measuring the frequency of fearful cognitions associated with panic attacks and agoraphobia (scores from 0 to 4)
BAI	Beck Anxiety Inventory Anxiety-specific measure of psychopathology. 21-item measure designed to assess the severity of self-reported anxiety. The total score ranges from 0 to 63, with higher scores indicating higher levels of anxiousness. Patient completed.
BDI-II	Beck Depression Inventory 21-item, self-report measure of severity of depression. Scored 0 to 63, with a high score indicating severe depression. Patient completed.
BRMS	Bech-Rafaelsen Mania Rating Scale – Modified version (BRMS). Clinician rated.
BRAMES	Severity of depression 11 items, rated 0-44, Clinician rated.
BSQ	Body Sensation Questionnaire Anxiety-specific measure of psychopathology. 17-item self-report instrument to evaluate fear of the physical sensations generally associated with a panic attack. Patient completed. Scored 0 to 4.
CES-D	Centre for Epidemiological Studies Depression Scale 20 item questionnaire on feelings of depression, scored 0-60. Patient completed
CGI-S	Severity of illness scale, rated 1-7. Clinician rated.
CV	Convergent validity
DD	Depressive Disorder
DE	Depressive Episode
ES	Effect size (mean assessment – mean baseline)/SD pooled SD at baseline
EPDS	Edinburgh Postnatal Depression Scale
EQ-VAS	The VAS question asked alongside the EuroQol EQ-5D measure.
GAD	Generalized Anxiety Disorder
GAD-Q-IV	9 item diagnostic measure of GAD. Clinician rated.

GAF	Overall occupational functioning. Clinician rated.
HADS	Hospital Anxiety and Depression Scale Scored 0 (no anxiety) to 21 (many complaints of anxiety). Clinician rated. Subscales: HADS-D (depression) HADS-A (anxiety)
HAM-D or HRSD	Hamilton Depression Rating Scale The original scale has 17 items. Scores range 0 to 62, higher scores indicating more severe symptoms. Clinician rated.
HRQL	Health related quality of life
KGV	Known group validity
MADRS	Montgomery-Asberg Depression Rating Scale. Clinician rated.
MBI	Maslach Burnout Inventory Work related stress. 3 subscales. Patient completed.
M-CIDI	Munich version of Composite International Interview
MDD	Major Depressive Disorder
MDE	Major Depressive Episode
MDI	Major Depressive Inventory 12 items used to calculate scores on 10 ICD-10 symptoms of depression. Patient completed.
MI	Mobility Inventory Anxiety-specific measure of psychopathology. The MI is a 29-item self-report instrument measuring the severity of behavioural avoidance. The MI is divided into two subscales, Avoidance Alone (MIA) and Avoidance Accompanied (MIB). Patient completed. Scores range 0 to 4.
NICE	National Institute of Health and Clinical Excellence
OLS	Ordinary Least Squares
PC	Primary Care
PHQ	Patient health questionnaire Includes a 9 item depression scale
PROQSY	Computerized assessment of minor psychological morbidity based on the Clinical Interview Schedule
QALY	Quality adjusted life year
QLDS	Quality of Life in Depression Scale 34 item depression specific HRQL instrument. Scores range from 0-34, with 34 indicating worst possible case. Patient completed.
Q-LES-Q-SF	Quality of Life Enjoyment and Satisfaction Questionnaire – short form 15 general activity items and one overall life satisfaction

	Sheehan disability scale – patient reported 3 item questionnaire to assess mental health functional impairment
QWB	Quality of Well-being. Preference based measure of utility.
R	Responsiveness
RCT	Randomised controlled trial
SCL-A	Symptom Checklist 10 questions scored from 10 (no anxiety) to 50 (many complaints of anxiety). Clinician rated.
SD	Standard Deviation.
SCID	Structure Clinical Interviews for DSM Disorders
SIGH-A	Structured Interview Guide for the Hamilton Anxiety Scale. Clinician rated.
SF-36	Short-form 36. Generic HRQL measure consisting of 8 dimensions assessing physical functioning, role limitations due to physical problems, bodily pain, general health, vitality, mental health, role limitation due to emotional problems, and social functioning. Two summary scores assess physical (PCS) and mental (MCS) facets. Scores range from 0 to 100.
SG	Standard gamble
SRMs	Standardised response means. (mean at assessment – mean baseline) / SD of differences in mean scores
SSI-28	Somatic Symptom Inventory
TAU	Treatment as Usual
TTO	Time trade-off
VAS	Visual analogue scale
VAS-pain	Visual analogue scale for pain
WHOQOL-BREF	World Health Organization Quality of Life-Brief questionnaire. Patient completed.
WHO-CIDI	World Health Organisation’s 12 month Composite International Diagnostic Interview
YMRS	Young mania rating scale. Clinician rated.

INTRODUCTION

Generic preference-based health status measures such as the EuroQol-5D (EQ-5D) are increasingly being used to inform health policy. The last decade has seen the increased use of economic evaluation, particularly the use of cost effectiveness analyses by agencies such as NICE to inform resource allocation decisions (NICE, 2008) where interventions are assessed in terms of their cost per Quality Adjusted Life Year (QALY). The QALY provides a way of measuring the benefits of health care interventions, including improvements in HRQL usually measured using a generic measure like EQ-5D. However, there has been only a limited use of generic measures of health in mental health (Gilbody et al, 2003).

The EQ- 5D and other generic preference-based measures such as the SF-6D (Brazier et al, 2002) aim to be applicable to all interventions and patient groups. For many physical conditions these instruments have passed psychometric tests of reliability and validity (e.g. for rheumatoid arthritis patients (Marra et al, 2005)), but not all (e.g. visual impairment in macular degeneration (Espallargues et al, 2005) and hearing loss (Barton et al, 2004). Doubts have also been raised about the appropriateness of generic measures in mental health (Brazier, 2010) and whether they are “sufficiently sensitive to the kinds of symptoms, functioning and quality of life change important for people with mental health problems.” (Knapp and Mangalore, 2007: 292).

One solution would be to use disease-specific preference-based measures (PBM), for example, there have been attempts to derive PBM from the PANSS and CORE-OM in mental health (Mavranouzouli et al, 2011). However, there are concerns about the comparability of such disease specific scales and in the UK, health technology assessment submissions to NICE are expected to follow the details outlined in the ‘reference case’ analysis described by the NICE methods guide (NICE, 2008). This clearly stipulates that wherever possible and appropriate, the EQ-5D is the favoured measure for generating utility values, thus allowing a common metric to assess health care interventions. Alternative measures may be used where the EQ-5D has been empirically demonstrated to be inappropriate in terms of their validity and responsiveness.

To assess the appropriateness of generic PBM in patients with depression and anxiety, we have undertaken a systematic review to investigate the construct validity

and responsiveness of generic PBM in depression and anxiety. This forms part of a wider project funded by the Medical Research Council (MRC) exploring the appropriateness of using generic PBM for mental health.

The review here will consider whether there is evidence to support the construct validity (or the degree to which an instrument measures what it claims to be measuring) and responsiveness (or the extent to which a measure can detect a clinically significant or practically important change over time (Walters, 2009) of generic utility measures within patient's with depression and/or anxiety.

METHODS

Utility measures being evaluated

EQ-5D

The EQ-5D questionnaire comprises a five dimensions: mobility, self-care, usual activities, pain and anxiety/depression. Respondents are asked to report their level of problems (no problems, some/moderate problems or severe/extreme problem) on each dimension to provide a position on the EQ-5D health state classification. Responses can be converted into one of 243 different health state descriptions (ranging from no problems on any of the dimensions [11111] to severe problems on all five dimensions [33333]) which each have their own preference-based score. Preference-based scores are determined by eliciting preferences i.e. establishing which health states are preferred from a population sample. To derive preferences a method such as time trade off (TTO) is used which involves asking participants to consider the relative amounts of time (for example, number of life-years) they would be willing to sacrifice to avoid a certain poorer health state. Utility values for each state have been elicited from respondents in various countries (see www.Euroqol.org). The scoring algorithm, or social tariff, for the UK is based on TTO responses of a random sample (n=2,997) of non-institutionalised adults. Values are anchored by '1' representing full health and '0' representing the state 'dead' with states 'worse than death' bounded by '-1'. Utility values from the UK EQ-5D tariff range from -0.59 to 1 (Dolan, 1997). The EQ-5D is often administered with the EQ-VAS requiring a direct valuation of the respondent's health state on a scale from worst health imaginable to best imaginable. Whilst this is also a reflection of individual preferences (Parkin and Devlin, 2006), it is not normally used to derive

QALYs, in part due to concerns that the VAS scale does not explicitly involve choice, nor provide a cardinal measure that is needed for QALYs.

SF-6D

The SF-6D provides a means of translating the widely used general health measure the SF-36 (Ware and Sherbourne, 1992) or the SF-12 into a preference-based single index (Brazier et al, 2002). The SF-6D reduces the eight dimensions of the SF-36 into six: physical functioning, role limitations, social functioning, pain, mental health and vitality. Each dimension has 4, 5 or 6 levels, giving a total of 18,000 possible health states. The values attached to each level and dimension generated by the classification system were derived from standard gamble (SG) valuations for a sample of 249 of these health states. Face-to-face interviews were conducted with a representative sample of 611 members of the UK population (Brazier et al, 2002).

Respondents initially ranked five SF-6D health states, plus the best and worst states from the SF-6D and immediate death. The SG questions then asked respondents to choose between each of five certain SF-6D states (imagining remaining in those states for the rest of their lives), versus a gamble between the best and 'pits' health states. Respondents were then asked to value the 'pits' state in relation to immediate death. The form of this valuation varied depending upon whether the respondent had ranked the 'pits' state as better or worse than dead. The result of the 'pits' valuation was used to 'chain' the health states such that they could be placed on the 0 (dead) to 1 (full health) scale. The valuations for the SF-6D were derived from a linear random effects model, and ranged from 0.29 to 1.0. (Brazier et al, 2002; Brazier et al., 2008).

Inclusion and exclusion criteria

Studies were eligible for inclusion if they contained data on any preference based health related quality of life measure within adults with depression or anxiety. This included studies that used a standardised utility measure within a trial setting, or as part of studies looking at the burden of illness of depression and anxiety. The outcomes had to include data that allowed measurement of the construct validity (i.e. known groups or convergent) or the responsiveness of the preference-based measure(s). Studies in which depression was not the primary diagnosis, but was co-morbid to another condition, were excluded. Those studies which contained only the VAS part of the EQ-5D were also excluded.

Identification of studies

For this review, 11 databases¹ were searched for published research, with searches limited to the English Language. (Search strategies are available from the authors). All searches were conducted in December 2010. The reference lists of relevant studies were searched for further papers. Citations, and where necessary full papers, identified by the searching process were screened by one reviewer (TP) using the inclusion criteria.

Data extraction

Data from all included studies were extracted (by one reviewer (TP)) using a form designed specifically for the broader project, and piloted on a sample paper. Data extracted included: country of publication, type of disorder, study sample characteristics (numbers, age, gender), outcome measures used, mean values for utility measures, and validity and responsiveness data. Where publications reported on similar data, this is highlighted and only recorded where different aspects of analysis are conducted.

Quality Assessment

The overall quality of a study does not necessarily determine whether it can provide useful evidence on the validity and responsiveness of the preference-based measures it contains. For example, to assess effectiveness of an intervention data should be analysed on an intention to treat basis. However, this is not necessary to be able to judge whether the utility measure is responsive to a change in health. As there is no formal method for assessing the quality of studies for this purpose (i.e. there are no quality assessment checklists) we draw on the methods described by Fitzsimmons et al (2009) to evaluate health-related quality of life data in their systematic review on the use and validation of quality of life instruments within older cancer patients. This includes whether tests of statistical significance were applied, whether differences between treatment groups were reported, whether clinical significance was discussed and whether missing data were documented.

The extent of missing data is important to know in order to judge how representative findings might be for patients with depression and anxiety. Missing item data and

¹ Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials, NHS Economics and Evaluations Database, Health Technology Database, Database of Abstracts of Reviews of Effects, MEDLINE, PreMEDLINE, CINAHL, EMBASE, Web of Science and PsycInfo.

completion rates for utility measures is also an important aspect of their practicality. However, that is not the focus of this review. How researchers have dealt with missing items within a scale is also important yet is not always reported. Studies may report that values have been imputed, although useful to know, does not in itself give sufficient information to assess appropriateness of how missing items have been dealt with.

Appropriate tests of significance between groups, and for changes over time, should be applied and discussed. It will not be possible to make judgments overall about whether the utility measure can identify groups which the population or patients consider to have a different health state, without such tests. If particular studies lack significance tests, they can however, help our general understanding and contribute towards a broader picture.

Studies should provide sufficient information to enable the reader to know exactly what utility measure is being used (for example, where a tariff based on public preferences is used this should be clearly referred to).

The overall aim of this review is to see if there is an accumulation of evidence that does or does not support the use of utility values in depression and anxiety. If studies identify findings which contradict an overall picture, knowing why this is the case, will contribute towards our understanding. We do not want to exclude studies based on any strict quality criteria, unless specific information on utility scores can not be extracted or there is a danger of misinterpretation of the findings.

Evidence synthesis and meta-analysis

Due to the large degree of heterogeneity between studies (including types of study designs, outcome measures, population characteristics and methods of determining construct validity and responsiveness) it was not appropriate to perform meta-analysis. Analysis was by narrative synthesis and data were tabulated.

Defining validity and responsiveness

Construct validity is defined as the extent to which an instrument measures the construct it is designed to measure and in the settings it is designed for (Streiner, 2003). Support for construct validity of health measures in the psychometric literature is typically taken from: Firstly, showing that the measure distinguishes between groups which we would expect to have different levels of the construct (*known group*

validity), such as the presence or absence of a disease or different levels of disease severity; Secondly, showing that the measure correlates highly with alternative, preferably validated, measures which are designed to measure the same construct (*convergent validity*). Evidence for known group validity will come from showing statistically significant differences in the average utility score by subgroup of another outcome which may be a measure of disease severity, functioning, disease specific quality of life or generic quality of life. Outcome measures may be judged either by the clinician or patient themselves. Patient measures are usually given more weight when measuring quality of life. Evidence for convergent validity will come from showing significant, preferably high, correlation to other outcome measures. Regression analysis can also be used to explore whether the generic utility measure, (or change in that measure), is related to factors which are identifying the construct we are trying to pick up (e.g. disease severity) and not to external factors unrelated to that construct (e.g. personal characteristics).

Support for the responsiveness of a measure is typically taken from showing that the measure responds to a change in health status, possibly following an intervention. If the measure changes when we expect it to, or changes in line with other measures, this also provides additional support for construct validity. Evidence for responsiveness will come from significant correlation with change scores on clinical outcome measures, significant change in the utility measure before and after an intervention and significant differences between patients classified as responders or non-responders by clinical or self-report measures. The performance of different outcome measures can be compared using effect sizes (ES) that compare the size of the effect or change relative to variability in the population. Common measures include the standardized response mean (SRM) which is computed by dividing the mean change in score (i.e. follow-up minus baseline) by the standard deviation of the change (Terwee et al, 2003) and the Cohens' D which is calculated by dividing the mean change in score by the standard deviation at baseline. Effect sizes of 0.2 are defined as small, 0.5 defined as moderate, and 0.8 defined as large (Cohen, 1988).

Traditional psychometric methods for considering construct validity and responsiveness need to be adapted to deal with utility scales. Generic multi-attribute utility scales are comprised of three elements: choice of dimensions; the levels for each dimension; the weights or preference attributed to each level/dimension. The validity and responsiveness of the first two can be assessed by traditional means

through considering data disaggregated into each dimension and comparing this to other quality of life and clinical measures. However, judging the validity and responsiveness of the combined utility score is less straightforward as this incorporates public preferences towards each state in addition to a description of the state, consequently, the application of these psychometric criteria to preference-based measures requires some adaptation (Brazier and Deverill, 1999). A generic utility measure may fail to identify change in an aspect of health which is identified by a disease severity measure, but if this change is not important to patients and not valued by them or the general population, then this is not a weakness of the utility measure.

For construct validity of utility measures tests of known group validity must be between groups which patients would report as different and the general population would value differently (Brazier and Devrill, 1999). We would like to know if the utility measures can identify differences in health which society would like to take into consideration in resource allocation decisions.

It may be possible to validate one utility measure against another. Where different utility measures have used different methods to generate the weights and use different dimensions this may be particularly useful, however, it is not clear which of the utility scales could be taken as a gold standard. Where differences exist between utility measures considering the methodology of the development process of the utility scale may shed light on this. For example, did the measure incorporate mental health in the development of the dimensions? Did those who valued the states have a good understanding of how different levels on a mental health dimension would impact upon quality of life?

Measures that pick up quality of life from a patient perspective and those which focus on functioning, are likely to have a stronger relationship to preferences than symptom based, and disease severity measures. Consequently, greater emphasis will be put on those comparisons.

Comparisons between utility scales and non-preference based quality of life scales or clinical measures do not necessarily support/or show lack of support for construct validity as they not designed to measure the same construct. However, these comparisons may highlight interesting differences between scales or parts of scales, which helps to build a picture of how useful utility measures are for this patient group.

For assessing responsiveness of utility measures, we require that the utility measure can identify a change in health where the before and after health states would be valued differently by the patient or the general public. Change which is not valued by society and/or the individual would not be expected to be picked up by a utility measure.

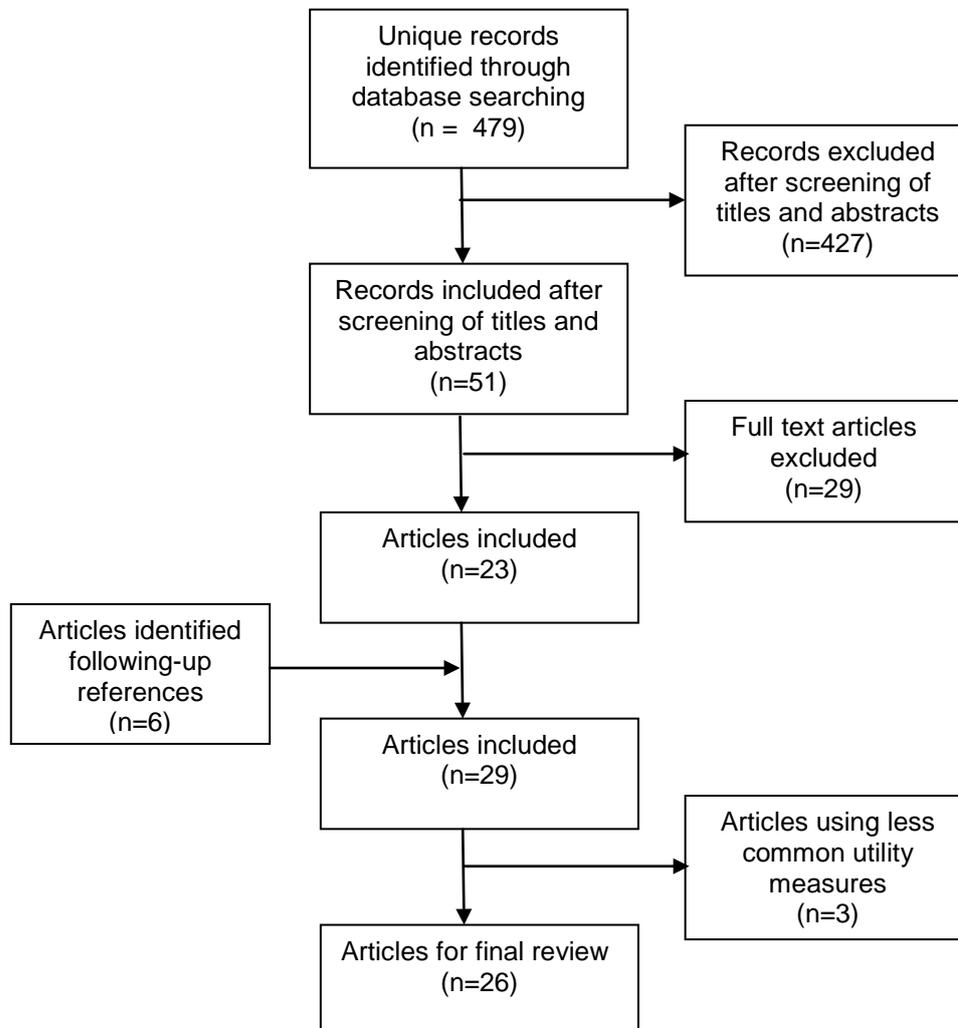
Assessing health measurement scales draws on an accumulation of evidence that suggests converging results, rather than single experiments (Streiner and Norman, 2003). Given the additional complexity of needing to judge utility measures by how closely they reflect preferences towards health states rather than just health states, this need for converging evidence from a number of different perspectives is even more important.

FINDINGS

Study characteristics

The search identified 479 studies. On reading titles and abstracts 427 were excluded and 29 were excluded on reading the full paper, leaving 23 papers. Following up on references gave an additional 6 papers. In some cases less commonly used preference-based HRQL measures were used. One study used the HUI2 and HUI3 (Revicki et al, 2008), two studies used the Quality of Well-Being (Mittal et al, 2006 and Pyne et al, 1997), one used the 15D (Saarni et al, 2007), and one used the SF-12 with utility weights derived from a convenience sample (Wells et al, 2007). As this gives insufficient evidence to draw conclusions about the validity and effectiveness of these five measures, the focus of the review became on EQ-5D and SF-6D only. A further three papers were therefore excluded, leaving 26 papers (see Figure 1).

Figure 1: Paper Identification



Included papers can be categorised into three. Those which explicitly look at the usefulness of utility measures in depression and anxiety (of which there are 8, see Table 1); those which use utility measures to consider the burden of depression and anxiety (of which there are 7, see Table 2) and those which use utility measures in clinical trials of depression and anxiety (of which there are 11, see Table 3). Further details of the papers can be found in Appendix 1.

Table 1: Validity and responsiveness of EQ-5D or SF-6D

Study	Patient group. Country	Utility measure	Contribution
-------	------------------------	-----------------	--------------

Gunther, 08	DE. Germany	EQ-5D	CV, R
Konig, 10	Anxiety disorder. Germany	EQ-5D	KGV, CV, R
Lamers, 06	Mood and anxiety disorders. Netherlands	EQ-5D, SF-6D	KGV, R
Mann, 09	Depression. UK	EQ-5D, SF-6D	KGV, CV, R
Petrou, 09	Post-natal depression. UK	EQ-5D, SF-6D	KGV
Revicki, 08	GAD. US	SF-6D	KGV, CV
Sapin, 04	MDD. France	EQ-5D	KGV, CV, R
Supina, 07	Population survey. Canada	EQ-5D	KGV

KGV = known group validity, CV = convergent validity, R = responsiveness

Table 2: Burden of depression and anxiety as measured by EQ-5D or SF-6D

Study	Patient group, country	Utility measure	Contribution
Aydemir, 09	MDE patients. Turkey	EQ-5D	KGV, CV
Fernandez, 10	Survey PC patients. Spain	SF-6D*	KGV
Mychaskiw, 08	GAD patients. US	EQ-5D	KGV, R
Saarni, 07	Population survey. Finland	EQ-5D	KGV
Sobocki, 07	Depressed patients. Sweden	EQ-5D	KGV, R
Stein, 05	Anxiety disorder patients. US	SF-6D*	KGV
Zivin, 08	Veterans with depression, US	SF-6D*	KGV

*SF-6D derived from the SF-12

Table 3: Trials on depression and anxiety using EQ-5D or SF-6D

Study	Patient group (n)	Utility measure	Contribution
Bosmans, 08	Patients with depression in PC. Netherlands	EQ-5D	R
Caruso, 10	DE in PC. Italy (FINDER)	EQ-5D	R
Ergun, 08	MDD. Turkey	EQ-5D	CV, R
Fernandez, 05	MDD outpatients. Europe	EQ-5D	R
Reed, 09	DE in PC. Europe. (FINDER)	EQ-5D	KGV
Konig, 09	Anxiety disorder in PC. Germany	EQ-5D	R
Peveler, 05	DE. UK	EQ-5D	R
Pyne, 10	Depressed patients. USA	SF-6D*	R
Serfaty, 09	Geriatric depression. UK	EQ-5D	R

van Straten, 08	Selected into self-help for depression, anxiety and stress. Netherlands	EQ-5D	R
Swan, 04	DD and moderate-severe episode. UK	EQ-5D	R

*SF-6D derived from the SF-12

Quality of included studies

Quality assessment of the studies was restricted to items relating to utility measures. All but 3 studies (Caruso et al, 2010; Ergun et al, 2007, Reed et al, 2009) reported tests for statistical significance relevant to tests of validity and responsiveness. As can be seen in Appendix 1, many studies do not completely report details on how missing outcome measure data was dealt with.

Validity and responsiveness of the EQ-5D

Known group validity of the EQ-5D

The EQ-5D is able to identify a utility detriment for patients with depression and anxiety disorders. In a Finnish population survey, controlling for somatic and psychiatric comorbidity, depressive disorders reduced EQ-5D by -0.091, anxiety disorders by -0.114, GAD by -0.110, MDD by -0.058, dysthymia by -0.122, social phobia by -0.102 (Saarni et al, 2007). A Canadian population survey found EQ-5D values for those with MDE only (recurrent and current) of 0.83, those with anxiety only of 0.84, those with anxiety and MDE of 0.70 and those with neither of 0.92 (Supina et al, 2007). More of the population experience both conditions (5.2%) than MDE alone (2.6%) emphasising the interconnectedness of these conditions.

The EQ-5D also shows significant differences by severity group for MDD patients (Sapin et al, 2004 and Sobocki et al, 2007), those with general mood and anxiety disorders (Lamers et al, 2006), and those with GAD (Mychaskiw et al, 2008). For example, Sobocki et al (2007) find an average EQ-5D score of 0.6 for mild depression (95% CI 0.54-0.65), 0.46 (95% CI 0.30-0.48) for moderate, and 0.27 (95% CI 0.21–0.34) for severe. Between group differences are not always significant (for example, the data above does not find a significant difference between average values for moderate and severe depression), often due to the high standard deviation of the EQ-5D. Aydemir et al (2009) is an exception as they do not find that the EQ-5D

significantly identifies MDD single episode versus recurrent episodes for patients in Turkey. The mental health summary component of the SF-36 also does not identify this group difference. Interestingly, they do find a significant difference between single and recurrent MDD episode in the physical functioning score of the SF-36, the physical health summary score and general health perception.

For patients with anxiety disorder, Konig et al (2010) find that almost all EQ-5D dimensions response levels (but particularly anxiety and depression) are associated with significant differences in scores of WHOQOL domains and measures of psychopathology, such as the BAI score.

To further understand the ability of the EQ-5D to identify patients with depression and anxiety it is useful to consider where health loss is identified across EQ-5D domains. For depressed patients this is in the domains of depression and anxiety, pain and discomfort, usual activities, and to a lesser extent mobility and self-care (see Table 4). The picture is remarkably similar across different studies conducted in different countries, supporting the reliability of the EQ-5D. Some differences arise due to different exclusion criteria across studies, particularly where comorbid physical conditions are excluded (e.g. Aydemir et al, 2010), which leads to less health loss on the pain dimension. Anxiety and affective disorder give a similar pattern of domain problems, but with less reporting of problems in the anxiety and depression domain.

Table 4: Health loss by dimension on the EQ-5D. % reporting moderate or extreme problems

Patient group	Mobility	Self-care	Usual Activities	Pain / Discomfort	Anxiety / Depression
MDE (DSM-IV), excluding comorbid condition (Aydemir, 10)	28.4	16.3	64.8	43.2	98.7
MDD (SCID) (Mann, 09)	27	16.2	75.7	64.2	94.5
MDD (DSM-IV) (Sapin, 04)	26.5	16.2	75.2	64.1	99.1

DE (ICD-10) (Gunther, 08)	28.8	26.9	66.4	66.0	78.8
Anxiety disorder (Konig, 10)	23	3.9	40.8	71.5	77.4

Comorbidity is very prevalent in patients with common mental health problems. For example, in Swedish patients diagnosed with depression in primary care 59% have one comorbidity (56% physical and 9% psychiatric) (Sobocki, 2007). A separation of mental and physical health does not therefore fit well for this patient group as the health impact of depression and anxiety is connected to both mental and physical health.

What is not clear, however, is whether the EQ-5D is picking up the impact of depression and anxiety on health domains beyond the anxiety and depression domain or whether the impact arises from other somatic or psychological comorbidities.

Convergent validity of the EQ-5D

The EQ-5D shows good correlation to clinician-rated measures of depression severity (-0.539 to -0.77) for depressed patients. Correlations to functioning (0.492), patient rated severity (-0.451 to -0.638) and patient rated quality of life (0.43 to 0.63) are also moderately good (see Table 5).

For patients with anxiety, Konig et al (2010) find a stronger correlation between EQ-5D and the physical health component of the WHOQOL-BREF than with the mental health component and moderately good correlations to depression measures (0.54 for the BDI-II). The Beck Anxiety Inventory correlates at 0.53, however, other self-complete measures of anxiety show correlations of 0.4 and below. This suggests a general pattern whereby the EQ-5D is best at identifying mental health conditions which also impact upon physical health, then those which impact upon depression, but less effective at picking up anxiety.

Table 5: Correlations of EQ-5D and clinical/quality of life measures

Scale	Type	Correlation	Patient group
-------	------	-------------	---------------

HAM-D. Depression severity	CS	-0.77	MDD (Aydemir, 09)
BRAMES. Depression severity	CS	-0.576	DE (Gunther, 08)
CGI. Severity of illness	CS	-0.539	
GAF. Occupational functioning	CF	0.492	
EQ-VAS	PQoL	0.440	
PHQ-9. Depression symptoms	PS	-0.451 baseline -0.638 follow up	Patients with depression (Mann, 09)
SF36 MHC	PQoL	0.49 baseline 0.56 day 28 0.63 day 56	MDD (Sapin, 04)
QLDS	PQoL	-0.43 baseline -0.68 day 56	
SF36 MHC	PQoL	0.49 baseline 0.63 day 56	
WHOQOL physical health, mental health	PQoL	0.7, 0.5	Anxiety disorder (Konig, 10)
Anxiety scales: BSQ, ACQ, MIA, MIB, BAI	PS	-0.40, -0.32, -0.35, -0.36, -0.53	
Depression scale. BDI-II	PS	-0.54	

CS = Clinician rated symptoms, CF = Clinician rated functioning, PF = patient completed functioning, PS = Patient completed symptoms, PQoL = patient assessed quality of life

Interestingly, correlations with patient quality of life (Sapin et al, 2004) and patient completed symptom scales (Mann et al, 2009) are stronger at endpoints than baseline. This suggests a stronger correlation between EQ-5D and patient reported depression outcome measures for milder states.

Regression analysis shows EQ-5D to be related to expected variables for depressed patients (Caruso et al, 2010, Reed et al, 2009, Soboki et al, 20002). For example, it has a significant negative relationship between the number of previous depressive episodes, the duration of the current episode, and somatic symptoms (Reed et al, 2009). Soboki et al (2002) find that clinical severity variables explain 23% of the variation in EQ-5D for depressed patients, with demographic variables not being significant. Models including patient rated quality of life find 40% of the variation in EQ-5D explained (Sapin et al, 2004).

Responsiveness of the EQ-5D

In general the EQ-5D is very responsive to improvement in both depressed (Caruso et al, 2010, Ergun et al, 2007, Fernandez et al, 2005, Sapin et al, 2004; Sobocki et al, 2007; Swan et al, 2004; Reed et al, 2009) and anxious patients (Konig et al, 2010) and performs as well as symptom based, functioning and quality of life measures. In some studies, despite substantial change, improvement is not significant due to the high standard deviation (Peveler et al, 2005). Studies also find substantial differences between patients identified as in remission versus those who are not (Mann et al, 2009, Sapin et al, 2004)

For depressed patients, effect size, and SRM are broadly in line with other measures, but lower in some studies due to higher standard deviation of EQ-5D relative to other measures. Van Straten (2008) find a Cohen's D of 0.44 for self-selected members of the public who complete a self-help course for depression, anxiety and stress. This compares with a Cohens D of 0.67 for the CES-D and 0.56 for the MDI (both patient completed symptom measures of depression), 0.51 for the SCL-A (a clinician rated symptom list for anxiety) and 0.48 for the HADS (clinician rated symptom list for depression). Lamers et al (2006) follow patients with a diagnosis of mood and anxiety disorders (major depression, dysthymic, social phobia, generalised anxiety) over an 18 month period. Despite a greater increase in the EQ-5D than the SF-6D they find an SRM about half that of the SF-6D.

Konig et al (2010) find that for anxiety patients the t statistic, ES and SRM of the EQ-5D are higher than for other measures (WHOQoL, BSQ, ACQ) for patients who become more anxious but for those who become less anxious the relative performance of the EQ-5D is mixed, being lower than BSQ and ACQ but higher than the WHOQoL. Konig et al (2009) find the EQ-5D to be in line with other anxiety measures (BAI and BDI-II) in showing no difference between and intervention and control group. Similarly, Bosmans et al (2008) find no significant difference between intervention and control group for depressed patients in primary care, in line with the MADRS depression scale.

Increases in the EQ-5D are positively related to disease severity for depression (Gunther et al, 2008, Lamers et al, 2006; Sobocki et al, 2007) which is indicative of a ceiling effect.

The findings of Serfaty et al (2009) are an exception to the general picture of responsiveness of the EQ-5D. Here the EQ-5D is less responsiveness than the BDI-

II. The patient group in this study has a mean age of 74.1, suggesting the EQ-5D may lack responsiveness for older patients.

Table 6: Responsiveness of the EQ-5D

Patient group	Responsiveness evidence	Significant difference
Depressed patients (Bosmans, 08)	No significant difference between intervention and control group. In line with MADRS measure.	No
MDD (Ergun 07)	Mean score increased from 0.44 to 0.91 at 6 weeks.	NA
Depressed patients (Caruso, 10)	Improvement of 0.26 at 3 months, 0.33 at 6 months.	NA
Severe MDD (Fernandez, 05)	Improvement at 8 weeks on Escitalopram and Venlafaxine	Yes
DE (Gunther, 08)	EQ-5D showed deterioration for those in worst health according to patient perceptions and BRAMES score and improvement for those in better health, but the later less so than other measures. T statistic, ES and SRM find greater responsiveness of EQ-5D (UK and German index) to deteriorating health than clinical measures (almost twice as large), but less responsive to health improvement: half the ES of CGI.	Yes
Anxiety disorder (Konig, 09)	No difference between intervention and control group. BAI and BDI also showed no differences.	No
Anxiety disorder (Konig, 10)	Effect size for more anxiety (a BAI increase of more than 0.5 of SD) -0.99, which was twice as big as other measures (WHOQoL, BSQ, ACQ). Effect size for less anxiety, 0.39. SRM -0.54 for more anxiety, again higher than other measures.	

	SRM 0.46 for less anxiety (BSQ -0.72, WHOQoL 0.35).	
Mood disorder (Lamers, 06)	Improvement of 0.167 at 1.5 years. Mean improvement in EQ-5D increased with severity. SRM was 0.466 (about half that for SF-6D)	Yes
Depressed patients (Mann, 09)	Mean score increased by 0.147 at 3 months. Median score increased by 0.069. Those classed as in remission (62% of sample) showed an increased in EQ-5D of 0.243.	Yes
Depressed patients (Mychaskiw, 08)	Those in functional remission 0.26 higher than those not in remission. Those in symptomatic remission 0.24 or 0.26 higher, depending on HAMA cut off.	Yes
Depressed patients (Peveler, 05)	Improvement of 0.22 at 12 months.	No
Depressed patients (Reed, 09)	Improvement at 3 months	NA
MDD (Sapin, 04)	Improvement of 0.35 at 4 weeks and 0.45 at 8 weeks. Only 9.3% extreme problem with anxiety / depression after 77.9% at baseline. Able to distinguish responder-remitters, responder non-remitters and non-responders based on MADRS score.	
Depressed patients (Serfaty, 09)	Mixed evidence on responsiveness. BDI-II found clearer improvement from baseline to 4 and 10 months and found CBT intervention superior to TAU. EQ-5D did not show superiority of CBT over	No

TAU.		
Depressed patients (Sobocki, 2007)	EQ-5D increased by 0.23 at 6 months (or last followup). Increase in EQ-5D positively related to disease severity (CGI-S).	Yes
Public recruited for self-help (van Straten, 08)	Improvement pre/post intervention. Effect sizes: Cohens D (course completers) CES-D 0.5 (0.67); MID 0.33, (0.56); SCL-A 0.42, (0.51) HADS 0.33, (0.48); EQ-5D 0.31, (0.44) MBI work stress not significant.	Yes
Depressed patients previous inadequate response (Swan, 04)	Of those that attended follow up, improvement found at week 12 and 26. In line with changes in GSI and BDI.	Yes

Validity and responsiveness of the SF-6D

Known group differences of the SF-6D

The SF-6D shows significant differences between disease severity groups (SCL subgroups) for mood disorder patients (Lamers et al, 2006) and for subgroups based on HAM-A scores for GAD patients (Revicki et al, 2008).

The utility detriment for depression and anxiety has been identified using the SF-6D (estimated from the SF-12) in a number of population surveys. Analysis of US survey data shows that the SF-6D identifies significantly lower utility for veterans with depression than without depression (0.57 versus 0.63) (Zivin et al, 2008). US outpatient data shows a drop in the SF-6D of -0.122 for anxiety disorder and -0.087 for major depression (Stein et al, 2005). Fernandez et al (2010) conduct quantile regressions on SF-6D values from a sample of patients from Spanish primary care. At the median they find a drop in utility of -0.20 for mood disorder and -0.04 for anxiety disorder (Fernandez et al, 2010).

In a sample of 114 patients with depression in the UK health loss is identified by the SF-6D in the domains of mental health (100%), vitality (98.8%), role limitation (98.8%), social functioning (89.1%), pain (78.7%) and physical functioning (22.1%) (Mann et al, 2009). As with the EQ-5D the SF-6D is picking up either the impact upon health of comorbidities, or a more holistic impact of depression and anxiety.

Convergent validity of the SF-6D

One study looked at the convergent validity of SF-6D for patients with GAD (Revicki et al, 2008). The SF-6D correlates -0.38 with GAD-Q-IV (a diagnostic measure of GAD), -0.52 with HAM-A (a severity score for GAD) and -0.64 with the PHQ-9 (a patient completed depression scale). Symptom measures explain 46% of the variance of SF-6D, suggesting a close relationship. The stronger correlation with the depression measure than the anxiety measures suggests that either public preferences give greater weight to changes in depression severity than anxiety or that the SF-6D measure is not as sensitive to changes in anxiety as it is to changes in depression. This pattern reflects that found for the EQ-5D.

Responsiveness of the SF-6D

Two studies compare the responsiveness of the SF-6D versus the EQ-5D, in MDD (Mann et al, 2009) and general mood disorder patients (Lamers et al, 2006). Although the SF-6D shows significant change over time, and distinguishes those patients in remission, the absolute improvement in both studies is higher for the EQ-5D. Mann et al (2009) find that mean improvement is higher for the SF-6D in the low severity group but lower than the EQ-5D in the two high severity groups. However, due to its lower SD the SRM is at least twice as high as that for the EQ-5D (0.833 for SF-6D versus 0.466 for EQ-5D at 1.5 years follow up).

SF-6D versus EQ-5D

Both the EQ-5D and the SF-6D perform reasonably well in terms of convergent validity with other measures, known group validity and responsiveness. However, evidence suggests they are not substitutes.

Lamers' study in the Netherlands includes both the EQ-5D and the SF-6D (Lamers et al, 2006). At baseline, more respondents report having no limitations when using the EQ-5D than when using the SF-6D and less report problems at the severe end of the scale. For example, 78% report no mobility problems and 93.5% report no problems

with self-care according to EQ-5D yet only 18% report no limitations in physical functioning in SF-6D. Fewer respondents report the most severe level of mental health problems with the EQ-5D than the SF-6D: 65% report 4 or 5 out of 5 for mental health responses on the SF-6D yet only 33% report 3 out of 3 for the EQ-5D.

This pattern is replicated in UK patients with depression (Mann et al, 2009). 73% and 83.8% report no mobility or self-care problems on the EQ-5D respectively, yet only 27.9% report no physical problems on the SF-6D. 86.6% of patients report feeling tense/downhearted or low most or all of the time using the SF-6D but only 29.4% report extreme problems with anxiety and depression on the EQ-5D. 57.1% report the most severe level on vitality for which there is no comparable measure in the EQ-5D.

The greater mental health loss reported on the SF-6D may be due to the fact that SF-36 and SF-12 asks questions about feelings whereas the EQ-5D domain 'depression or anxiety' sounds more clinical (Mann et al, 2009: p574). Alternatively, this may be a consequence of using 5 rather than 3 levels.

Despite the fact that SF-6D appears to identify more health loss, in the study by Mann et al (2009) the EQ-5D shows greater responsiveness with larger health gains for all patients at follow up and for those in remission. This arises in part through the lower average score on the EQ-5D for severely depressed patients at baseline (0.337 versus 0.544 for the SF-6D). Lamers et al (2006) also find mean improvement in EQ-5D to be higher than the SF-6D for the two most severe subgroups, although lower than the SF-6D for low severity groups.

The SF-6D generally outperforms the EQ-5D in terms of effect size and SRM in part as a consequence of the lower standard deviation, and the more normal distribution of the SF-6D. Consistently, EQ-5D has higher (by 2-3 times) standard deviation.

The study by Petrou et al (2009) which looks at levels of health for women six months postpartum has been included in this review because it offers another comparison the performance of the EQ-5D and the SF-6D in identifying levels of health these women, some of whom may have post-natal depression. They find the SF-6D to have better discriminatory ability when women are compared across self-reported health status groups, and by two alternative cut off scores on the Edinburgh Post Natal Depression Scale; the SF-6D generating higher area under the ROC

scores. The mean EQ-5D is significantly higher than the SF-6D and the minimum EQ-5D in the sample was 0.077, much lower than that of the SF-6D at 0.374. 177 women (35.9%) had full health according to the EQ-5D yet had an SF-6D score of below 1 (29.2% and 34.1% identifying problems with mental health and vitality, respectively). Whereas only one woman had an SF-6D of 1 yet identified moderate pain/discomfort on EQ-5D. The authors suggest four possible reasons for the greater sensitivity of the SF-6D to maternal health. First, the SF-6D taps into broader aspects of health and quality of life. Secondly, the SF-6D has a greater number of response items. Thirdly, the wording on SF-6D includes positive and negative items. And lastly, the SF-6D refers to a longer time frame (past 4 weeks) versus the EQ-5D which refers to today.

The SF-6D therefore appears better at picking up mild mental health problems, whereas EQ-5D gives greater weight to those with severe mental health problems. This pattern is not unique to mental health and has been identified in a number of conditions (Brazier et al, 2004). This difference arises due to differences in classification system and valuation technique; TTO for the EQ-5D and SG for the SF-6D. Tsuchiya et al (2006) find a cross over relationship where the SG SF-6D protocol generates values which are higher than the EQ-5D TTO protocol, yet for milder states TTO values are higher than SG values. This cross over point has been estimated to be 0.754 on the EQ-5D (Barton et al, 2008).

The worst state which can be described by the descriptive system is worst for EQ-5D than SF-6D, as the SF-6D does not cover very severe states (Brazier et al 2004). This may in part explain why the lowest value on the SF-6D is +0.291 whereas for the EQ-5D the lowest value is -0.59.

If the SF-6D is an accurate representation of preferences then the EQ-5D risks missing health change at the top end and overstating the value of change at the severe end. Alternatively, if EQ-5D is better reflection of preferences the SF-6D identifies change at the top end that is not meaningful for resource allocation decisions and understates change at the severe end.

Further evidence on whether EQ-5D or SF-6D most closely reflects the utility loss from depression and anxiety may be sought by making comparisons with direct utility valuations of depression and anxiety. These comparisons may come either from utility values derived directly from patients with depression or anxiety, from patients

who have previously been in those health states, or from valuations from the general public who are given more detailed scenarios describing these health states.

The range of severity and episodic nature of common mental health problems presents a challenge for valuation. Ideally, we require values for different levels of severity of depression and anxiety in addition to states of remission. Combining utility scores for patients with general depressive or anxiety disorder will disguise much of these differences. Furthermore, severity is likely to be related to study involvement and completion, suggesting values may be an overestimate of average patients.

Some studies which have conducted trade off exercises with patients with depression or anxiety suffer from failure of participants to make any sacrifice of length of length in TTO exercises or risk of death in SG exercises. Konig et al (2009) conducted TTO exercises with patients with affective disorder in psychiatric hospital in Germany. 29.4% of patients did not trade in the TTO exercise and the likelihood of being a non-trader was related to quality of life. Failure to trade is particularly problematic for postal surveys with this patient group (see Wells et al (2007) and Donald-Sherbourne et al (2001)). As these non-traders effectively rate their health state as full health, this leads to higher average utility value, suggesting lack of confidence in findings of postal surveys for this patient group.

Table 7 gives a summary of direct health state utility values for patients experiencing depression and anxiety (excluding those based on postal surveys (Wells et al, 1999, Donald-Sherborne et al, 2001 and Isacson et al, 2005). It is difficult to compare those from Bennett et al (39) since they incorporate a unique McSad descriptive system, the other studies show values ranging from 0.60 for moderate-severe depression and 0.74 for mild or general depressive disorder.

Direct valuations with patients with depression and anxiety are only slightly correlated with generic utility scores. Konig et al (2009) find that TTO scores correlate 0.31 with EQ-5D UK index and 0.24 with the EQ-5D German index in patients with anxiety disorder. Revicki and Wood (1998) find SG responses from 70 patients diagnosed with depressive disorder correlate at 0.29 with EQ-5D. This relatively low correlation may arise if public valuation studies give different weights to health attributes compared with depressed or anxious patients.

Lenert et al (2000) compare SG values for self-rated health of 71 patients with depressive symptoms with utility values from the SF-12 for the same state and do not find a significant difference. Depressed patients with near normal health tended to rate their current health as less preferable than the matched state, those with poorer health rated current health as more preferable than the matched state. This might suggest that utility scores overstate the health loss from severe depression states yet understate that from mild depression. The SF-12 utility score used here is not the SF-6D, so a direct comparison is not possible, but this does indicate that public preferences might undervalue mild mental health loss and over-value severe mental health loss relative to patient valuations, which would favour use of the SF-6D over the EQ-5D.

Table 7: Direct valuation on own current health

Study	Condition	Method	Value
Bennett 00, Canada	Depressed patients in primary care with at least one unipolar episode of major depression	SG McSad	0.79
Fryback 93, USA	Major depression	TTO	0.70
Pyne 09, USA	Mild to moderate depression (PHQ9 of 5-14)	SG	0.74
	Moderate to severe depression	SG	0.60
Revicki & Wood 98, USA	Depressive disorder	SG	0.74
Konig 09, Germany	Patients in psychiatric hospital with affective disorder	TTO	0.66

Earlier studies on valuations of hypothetical states for depression have found much lower values, for example, Sackett and Torrance (1978) find a value of 0.44 using TTO (and a duration of 3 months) on a population survey in Canada. Table 8 shows utility values drawn from hypothetical valuations from: members of the general public; those with a recent history of depression; and those currently experiencing a depressive episode.

Values for severe depression range from 0.04 to 0.68. The values from Bennett et al (2000), which uses the McSad descriptive system, are far lower than any other value. This makes direct comparisons problematic, however, they do point to a substantial

health loss from severe depression states that may not be being reflected in other utility scores possibly because severe states are not adequately described.

Currently depressed patients rate depression as lower than either general public or previously depressed patients (Pyne et al, 2009; Schaffer et al, 2002) and severely depressed patients rate depression states lower than mildly depressed patients (Pyne et al, 2009). The proportionate differences between population and patients valuations increases with hypothetical depression severity. However, those patients who have a history of depression, but are not currently depressed, rate depression similarly to the general public (Pyne et al, 2009; Schaffer et al, 2002)

Table 8: SG valuations of hypothetical scenarios

Study	Who values	Mild	Moderate	Severe
Bennett 00 Canada	105 depressed primary care patients in remission	0.59	0.32	0.09
	As above (life time)			0.04
Pyne 09 USA	Population	0.87	0.77	0.63
	History of depression (PHQ-9 < 5)	0.89	0.80	0.68
	Current Mild-moderate depression (PHQ-9 5 to 14)	0.87	0.74	0.63
	Current Severe depression (PHQ-9 >=15)	0.79	0.69	0.58
Revicki & Wood, 98 USA/Canada	Patients with depressive disorder			0.30
Schaffer 02 Canada	Currently depressed patients (HRSD >=16)	0.59	0.51	0.31
	Patients with DD not currently depressed (HRSD <16)	0.79	0.67	0.47
	Population	0.80	0.69	0.46

The lower valuations of currently depressed patients have been explained by negative cognitive distortions (Pyne et al, 2009) including all or nothing thinking and over-generalising negative events. However, when patients assessed their own health imagining not having depressive symptoms no differences were found (Pyne

et al, 2009). Where some depressed patients are suicidal this might lead to lower valuations because they are happy to risk death in SG and shorten their life in TTO. However, Schaffer et al (2002) find no difference in utility valuations between depressed patients with suicidal thoughts and those without.

Differences may arise through changes in preferences between depressive episode and remission. Alternatively, they may arise from a different understanding or interpretation of the condition. Those patients who are currently in remission may have more limited experience of severe states of depression compared with those currently in a depressive episode.

One of the difficulties of valuing depressive states is that, for most SG questions respondents are required to imagine a permanent state, yet, depression is generally a time limited state. Those in remission may be more aware of the episodic nature of the condition than those currently experiencing the state. Furthermore, the cognitive demands of TTO and SG exercises may be problematic for patients currently experiencing depression or anxiety.

Looking at the evidence on direct utility valuations for depression and anxiety presents a mixed picture. Evidence from patients valuing their own health using direct TTO or SG questions supports higher utility values for severe and moderate depression found by the SF-6D. However, TTO/SG trade-off exercises are subject to high numbers of currently depressed/anxious patients not being willing to trade.

Evidence from patients valuing scenarios describing depressive states finds that patients currently experiencing severe depression valuing severe depression as lower (0.3, 0.31, 0.58) and more in line with lower values found from the EQ-5D. Yet those who are no longer currently in a depressive state value severe depression higher and are more supportive of the SF-6D values.

Summary

The SF-6D and EQ-5D demonstrate good construct validity and responsiveness for patients with depression. They can both distinguish between groups that are known to vary according to severity of depression, and across differences in quality of life of depressed patients. Both measures respond to clinical and quality of life improvement and deterioration. Indeed, in many cases they are more responsive

than depression specific measures. This is likely to be due to integrated nature of mental and physical health problems and potential simultaneous improvement in co-morbid conditions. More evidence is required on the validity and responsiveness of generic measures for older people with depression, since this group may be an exception.

The performance of EQ-5D and SF-6D for patients with anxiety is a little more mixed. The EQ-5D and SF-6D are more highly correlated with depression scales in patients with anxiety than they are clinical anxiety scales. The ability of utility measures to distinguish between sub-groups of patients with anxiety may be driven by aspects of depression within anxiety disorder and the presence of co-morbid depression. More work is required to test whether EQ-5D and SF-6D are sensitive to changes in anxiety alone.

For both anxiety and depression the EQ-5D is better at identifying those who increase in severity compared with those who show improvement. Those patients moving into a level 3 on any dimension on the EQ-5D incur an additional detriment (0.269) according to the UK tariff (the N3 term). For example, moving from health state 11222 to health state 11223 brings a change in utility score from 0.689 to 0.255. Deterioration of already depressed or anxious patients may result in patients moving to level 3 for depression and anxiety.

The relationship between the EQ-5D and the SF-6D reflects that found for other conditions. The EQ-5D shows a lower level of utility at the most severe end for depression, the SF-6D shows equal or greater detriment at the milder end. The SF-6D identifies utility loss in patients that report full health on the EQ-5D, however, patient averages for mild depression and anxiety are still able to show lower than normal population utility using the EQ-5D.

Turning to direct utility estimation for depression and anxiety states does not help distinguish between validity of EQ-5D over SF-6D. Whilst severely depressed patients rating scenarios of severe depression give values closer to those found by the EQ-5D, depressed patients rating their own health and ex-depressed patients rating severe depression give values that are significantly higher. More research is needed to understand the source of these different values. Recent developments in bringing five dimensions to the EQ-5D (Pickard et al, 2008) may help illuminate the cause of difference in utility values between EQ-5D and SF-6D.

References

- Aydemir R, Ergun H, Soygur H, Kesebir S, Tulunay C (2009) Quality of life in major depressive disorder: A cross-sectional study, *Turkish Journal of Psychiatry*, 20: 205–212.
- Barton GR, Sach TH, Avery AJ, Jenkinson C, Doherty M, Whyne DK, Muir KR (2008) A comparison of the performance of the EQ-5D and SF-6D for individuals aged ≥ 45 years. *Health Economics*, 17(7): 815-32.
- Barton GR, Bankart J, Davis AC, Summerfield QA (2004) Comparing utility scores before and after hearing-aid provision. *Applied Health Economic Policy*, 3(2): 103-5.
- Bennett, K.J., Torrance, G.W., Boyle, M.H., Guscott, R. (2000) Cost-Utility Analysis in Depression: The McSad Utility Measure for Depression Health States *Psychiatric Services*, 51 (9): 1171-1176.
- Bosmans JE, Hermens MLM, de Bruijne MC, van Hout HPJ, Terluin B, Bouter LM, Stalman WAB, van Tulder MW (2008) Cost-effectiveness of usual general practitioner care with or without antidepressant medication for patients with minor or mild-major depression, *Journal of Affective Disorder*, 111(1): 106-112.
- Brazier J (2010) Is the EQ-5D fit for purpose in mental health? *British Journal of Psychiatry*, 197(348): 349.
- Brazier J, and Deverill M (1999) A checklist for judging preference-based measures of health related quality of life: learning from psychometrics, *Health Economics*, 8(1): 41-51.
- Brazier J, Roberts J, Deverill M (2002) The estimation of a preference based measure of health from the SF-36, *Journal of Health Economics*, 21: 271-92.
- Brazier J, Roberts J, Tsuchiya A and Busschbach J (2004) A comparison of the EQ-5D and SF-6D across seven patient groups, *Health Economics*, 13(9): 873-84.
- Brazier JE, Rowen D, Hanmer J (2008) Revised SF6D scoring programmes: a summary of improvements, *Patient Reported Outcomes Newsletter*, 40: 14-15.

- Caruso R, Rossi A, Barraco A, Quail D, Grassi L (2010) The Factors Influencing Depression Endpoints Research (FINDER) study: Final results of Italian patients with depression. *Annals of General Psychiatry*. Vol.9.
- Cohen J (1988) *Statistical Power Analysis for the Behavioural Sciences*. 2nd ed. USA: Lawrence Erlbaum Associates.
- Dolan P (1997) Modelling valuations for EuroQol health states. *Medical Care*, 11: 1095-1108.
- Donald Sherbourne C, Unützer J, Schoenbaum M, Duan N, Lenert LA, Sturm R, Wells KB (2001) Can utility-weighted health-related quality-of-life estimates capture health effects of quality improvement for depression? *Medical Care*, 39(11):1246-59.
- Ergun H, Aydemir O, Kesebir S, Soygur H, Tulunay FC (2007) SF-36 and EQ-5D quality of life instruments in major depressive disorder patients: Comparisons of two different treatment options, *Value in Health*, 10(6): A303.
- Espallargues M, Czoski-Murray C, Bansback N, Carlton J, Lewis GM, Hughes LA, et al. (2005) The impact of age related macular degeneration on health state utility values. *Investigative Ophthalmology and Visual Science*, 46: 4016-23.
- Fernandez JL, Montgomery S, Francois C. (2005) Evaluation of the cost effectiveness of escitalopram versus venlafaxine XR in major depressive disorder. *Pharmacoeconomics*, 23(2): 155-67.
- Fitzsimmons D, Gilbert J, Howse F, Young T, Arrarras J, Bredart J, et al. (2009) A systematic review of the use and validation of health-related quality of life instruments in older cancer patients, *European Journal of Cancer*, 45: 19-32.
- Gilbody S, House A, Sheldon T (2003) Outcome measures and needs assessment tools for schizophrenia and related disorders, *Cochrane Database of Systematic Reviews*, Issue 1.
- Gunther OH, Roick C, Angermeyer MC, König HH (2008) The responsiveness of EQ-5D utility scores in patients with depression: A comparison with instruments measuring quality of life, psychopathology and social functioning, *Journal of Affective Disorders*, 105(1-3): 87-91.
- Isacson D, Bingefors K, von Knorring L (2005) The impact of depression is unevenly distributed in the population, *European Psychiatry*, 20: 205-212.

Konig HH, Born A, Heider D, Matschinger H, Heinrich S, Riedel-Heller SG, et al. (2009) Cost-effectiveness of a primary care model for anxiety disorders, *British Journal of Psychiatry*, 195(4): 308-17.

Konig HH, Born A, Gunther O, Matschinger H, Heinrich S, Riedel-Heller SG, et al. (2010) Validity and responsiveness of the EQ-5D in assessing and valuing health status in patients with anxiety disorders. *Health & Quality of Life Outcomes*, 8: 47.

Knapp and Mangalore (2007) The trouble with QALYs, *Epidemiologia e Psichiatria Sociale*, 16, 4: 289-293.

Lamers LM, Bouwmans CAM, Van Straten A, Donker MCH, Hakkaart L (2006) Comparison of EQ 5D and SF 6D utilities in mental health patients, *Health Economics*, 15(11): 1229-36.

Lenert LA, Rupnow MF, Elnitsky C, Lenert LA, Rupnow MFT, Elnitsky C. (2005) Application of a disease-specific mapping function to estimate utility gains with effective treatment of schizophrenia, *Health & Quality of Life Outcomes*, 3: 57.

Mann R, Gilbody S, Richards D. Putting the 'Q' in depression QALYs: a comparison of utility measurement using EQ-5D and SF-6D health related quality of life measures. (2009) *Social Psychiatry & Psychiatric Epidemiology* 44(7):569-78.

Marra CA, ahidi AA, uh D, opec JA, Abrahamowicz M, sadaille JM, et al. (2005) Are indirect utility measures reliable and responsive in rheumatoid arthritis patients? *Quality of Life Research*, 14(5): 1333-44.

Mavranouzouli I, Brazier J, Young TA, Barkham M. Using Rasch analysis to form plausible health states amenable to valuation: the development of the CORE-6D from a measure of common mental health problems (CORE-OM) (2011) *Quality of Life Research*, 20(3): 321-33.

Mittal D, Fortney JC, Pyne JM, Edlund MJ, Wetherell JL (2006) Impact of comorbid anxiety disorders on health-related quality of life among patients with major depressive disorder, *Psychiatric Services*, 57(12): 1731.

Mychaskiw M, Hoffman D, Dodge W (2008) EQ-5D index scores by remission status in patients with generalized anxiety disorder, *International Journal of Neuropsychopharmacology* 11: 279.

NICE (2008) Guide to the methods of technology appraisal. London: National Institute for Health and Clinical Excellence.

Parkin D, Devlin N (2006) Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics*, 15: 653–664.

Petrou S, Morrell J, Spiby H. (2009) Assessing the empirical validity of alternative multi-attribute utility measures in the maternity context, *Health and Quality of Life Outcomes*, 7(1): 40.

Peveler R, Kendrick T, Buxton M, Longworth L, Baldwin D, Moore M, et al. (2005) A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine, *Health Technology Assessment* (Winchester, England), 9(16):1-134.

Pickard AS, De Leon MC, Kohlmann T, Cella D, Rosenbloom S (2008) Comparing the standard EQ-5D three level system with a five level version, *Value in Health*, 11(4): 589-99.

Pyne JM, Patterson TL, Kaplan RM, Gillin JC, Koch WL, Grant I (1997) Assessment of the quality of life of patients with major depression, *Psychiatric Services*, 48(2): 224.

Pyne JM, Fortney JC, Tripathi S, Feeny D, Ubel P, Brazier J. (2009) How bad is depression? Preference score estimates from depressed patients and the general population, *Health Services Research*, 44(4): 1406-23.

Pyne JM, Fortney JC, Tripathi SP, Maciejewski ML, Edlund MJ, Williams DK. (2010) Cost-effectiveness analysis of a rural telemedicine collaborative care intervention for depression, *Archives of General Psychiatry*, 67(8): 812-821.

Reed C, Monz BU, Perahia DG, Gandhi P, Bauer M, Dantchev N, et al. (2009) Quality of life outcomes among patients with depression after 6 months of starting treatment: results from FINDER, *Journal of Affective Disorders*, 113(3): 296-302.

Revicki DA, Brandenburg N, Matza L, Hornbrook MC, Feeny D (2008) Health-related quality of life and utilities in primary-care patients with generalized anxiety disorder, *Quality of Life Research*, 17(10): 1285-94.

Revicki DA, Wood M. (1998) Patient-assigned health state utilities for depression-related outcomes: Differences by depression severity and antidepressant medications, *Journal of Affective Disorders*, 48(1): 25-36.

Saarni SI, Suvisaari J, Sintonen H, Pirkola S, Koskinen S, Aromaa A, Lonnqvist, J (2007) Impact of psychiatric disorders on health-related quality of life: General population survey, *British Journal of Psychiatry*, 190(4): 326-32.

Sapin C, Fantino B, Nowicki ML, Kind P (2004) Usefulness of EQ-5D in assessing health status in primary care patients with major depressive disorder, *Health & Quality of Life Outcomes*, 2:20.

Serfaty MA, Haworth D, Blanchard M, Buszewicz M, Murad S, King M (2009) Clinical effectiveness of individual cognitive behavioral therapy for depressed older people in primary care: a randomized controlled trial, *Archives of General Psychiatry*, 66(12): 1332-40.

Sobocki P, Ekman M, Agren H, Krakau I, Runeson B, rtensson B, et al. (2007) Health-related quality of life measured with EQ-5D in patients treated for depression in primary care, *Value in Health*, 10(2): 153-60.

Stein MB, Roy-Byrne PP, Craske MG, Bystritsky A, Sullivan G, Pyne JM, et al. (2005) Functional impact and health utility of anxiety disorders in primary care outpatients. *Medical Care*, 43(12): 1164.

Streiner DL, Nomran GR (2003) *Health Measurement Scales: A practical guide to their development and use*. USA: Oxford University Press.

Supina AL, Johnson JA, Patten SB, Williams JVA, Maxwell CJ (2007) The usefulness of the EQ-5D in differentiating among persons with major depressive episode and anxiety. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care & Rehabilitation*, 16(5):749-54.

Swan, J., Sorrell, E., MacVicar, B., Durham, R., Matthews, K. (2004) "Coping with depression": an open study of the efficacy of a group psychoeducational intervention in chronic, treatment-refractory depression, *Journal of Affective Disorders*, 82(1): 125-9.

Tsuchiya A, Brazier JE., Roberts J (2006) Comparison of valuation methods used to generate the EQ5D and the SF6D value sets in the UK, *Journal of Health Economics*, 25(2): 334-346.

Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM (2003) On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation, *Quality of Life Research*, 12: 349-62.

Walters SJ (2009) *Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation: A Practical Guide to analysis and interpretation*, WileyBlackwell.

Ware JE, Sherbourne CD (1992) The MOS 36-item short-form health survey (SF36): I. Conceptual framework and item selection, *Medical Care*, 30: 473-483.

Wells KB, Schoenbaum M, Duan N, Miranda J, Tang LQ, Sherbourne C (2007) Cost-effectiveness of quality improvement programs for patients with subthreshold depression or depressive disorder, *Psychiatric Services*, 58(10): 1269-78.

Zivin K, McCarthy JF, McCammon RJ, Valenstein M, Post EP, Welsh DE, Kilbourne AM (2008) Health-related quality of life and utilities among patients with depression in the Department of Veterans Affairs, *Psychiatric Services*, 59(11): 1331-4.

Table A.2: Validity and responsiveness assessment methods used in studies, trials, health burden, and validity/responsiveness

Study Author, Year Location	Sample	Descriptive system (i.e. EQ-5D, SF- 36)	Utility values at baseline	Validity results	Responsiveness results	Authors' conclusions & comments
Aydemir et al, 2009 Turkey See Ergun et al for related trial data	74 patients, 18- 65 years, diagnosed major depressive episode according to DSM-IV criteria. Exclusion: other psychiatric disorder, comorbid condition. Mean age 39.6 years, 63.5% female. 32.4% recurrent depression	HAM-D, SF-36, EQ-5D UK EQ-VAS	Highest mean SF-36 score in physical function (79.2) and lowest in vitality (23.9). EQ-5D levels (no difficulties/moderat e/extreme) Mobility (71.6/27.0/1.4) Self-care (83.8/14.9/1.4) Usual care (35.1/40.5/24.3) Pain/discomfort (56.8/37.8/5.4) Anxiety/depression (1.3/36.5/62.2) EQ-5D index 0.4 (SD 0.3) EQ-VAS 38.2 (SD 22.3)	Single: EQ-5D 0.45 (SD 0.29) Recurrent: EQ-5D 0.41 (SD 0.31) Not sig. different The physical component summary of the SF-36 found patients with recurrent depression to be in significantly poorer health. The mental health summary component summary of the SF-36 showed no significant differences between single and recurrent depression. EQ-5D correlated at -0.77 with HAM-D.		

<p>Bosmans et al 2008</p> <p>The Netherlands</p>	<p>Patients with minor or mild-major depression in primary care.</p> <p>Exclusions: Currently receiving antidepressants (AD) or psychological therapy</p> <p>RCT: n=44 Usual care no AD. Mean age 48, 73% female</p> <p>n=45 Usual care plus AD, mean age 46, 76% female</p>	<p>MADRS EQ-5D UK</p>	<p>EQ-5D No AD 0.64 (SD 0.26) AD 0.66 (0.23)</p>		<p>Mean difference in QALYs gained between the two groups - 0.00045 (95% CI -0.093; 0.084)</p> <p>Difference in improvement in MADRS score -0.81 (95% CI -5.6; 4.0)</p>	
<p>Caruso et al, 2010</p> <p>Italian data from FINDER, 6 month observation study in 12 European countries</p>	<p>N=513 patients in primary care with clinically diagnosed episode of depression requiring pharmacological treatment.</p> <p>Mean age 49.2, 72.9% female.</p>	<p>HADS-D HADS-A SSI-28 VAS pain</p> <p>SF-36 EQ-5D</p>		<p>Regression analysis explored predictors of EQ-5D (n=328).</p> <p>EQ-5D at 6 months significantly related to:</p> <ul style="list-style-type: none"> - Switching antidepressants - EQ-5D at baseline, - SSI-somatic at baseline - no of episodes of depression, 	<p>EQ-5D Baseline: 0.40 (SD 0.01) 3 month: 0.66 (SD 0.26) 6 month 0.73 (SD 0.23)</p> <p>EQ-VAS Baseline: 45.7 (SD 19.6) 3 month: 61.3 (SD 17.9) 6 month: 69.3 (SD 17.0)</p>	

	Data recorded at baseline, 3 & 6 months.			<ul style="list-style-type: none"> - chronic medical condition - VAS pain at baseline - number of dependents - HADS-A at baseline <p>VAS scores at 6 months significantly related to</p> <ul style="list-style-type: none"> - EQ-VAS at baseline, - any psychiatric illness in last 2 years - switching antidepressants - occupational status - age - VAS pain at baseline 		
Ergun Turkey Abstract only published	74 patients with major depressive disorder in RCT.	EQ-5D UK HAM-D	Mean at baseline EQ-5D 0.44	(As above) HAM-D correlates 0.77 with EQ-5D.	EQ-5D increase from mean 0.44 at baseline to 0.91 at 6 weeks follow up.	Possible typo in SD
Fernandez et al, 2005 8 European countries RCT of Escitalopram vs venlafaxine	293 outpatients (aged 18-85) fulfilling DSM-IV criteria for severe MDD, without suicidal tendencies. Exclusion: history mania, bipolar,	EQ-5D UK QLDS MADRAS	At baseline >2/3 rd s had some or severe problems in the dimensions for pain, anxiety/depression and usual activities.		EQ-5D Baseline to Week 8 Escitalopram arm: 0.52 to 0.78 (p<0.001) Venlafaxine arm: 0.54 to 0.77 (p<0.001)	

XR in MDD	<p>schizophrenia or psychotic disorder, evidence of OCD, eating disorder, mental retardation, pervasive development disorder.</p> <p>Escitalopram (n=126) Mean age 48.4, 75.4% female</p> <p>Venlafaxine (n=125) Mean age 46.5, 71.2% female</p>					
Fernandez, 2010 Spain	Representative sample of 3815 primary care patients	SF-6D from SF-12	<p>Utility loss, adjusting for socio-demographic characteristics and morbidity</p> <p>Regression model with inference at the median 12-month mood disorder -0.20 (-0.21 to -0.18) 12-month anxiety</p>	<p>Median SF-6D score Any mood disorder: 0.547 (n=476) MDD: 0.527 (n=332) Dysthymia: 0.603 (n=115) Any anxiety disorder: 0.660 (n=659) Panic disorder: 0.606 (n=253) GAD: 0.681 (n=129) Social phobia: 0.599 (n=64) Specific phobia: 0.698 (n=232) Any substance misuse: 0.723 (n=122) ns</p>		

			disorder -0.04 (-0.06 to -0.03) Quantile regression shows impact of mood disorder is greater at higher levels of utility.	Alcohol misuse: 0.737 (n=81) ns		
Garcia-Cebrian et al, 2008 FINDER European	Patients (>= 18) with clinical depression enrolled prior to commencing antidepressant treatment, presenting to primary care or specialists. Of 3468 at baseline, 343 had no follow up data, 271 data at 3 months only, 2854 had 3 & 6 months. Mean age 46.8, 68.2% female	EQ-5D EQ-VAS SF-36 HADS-D HADS-A SSI-28 VAS-pain	Baseline scores: 51.5% experienced anxiety disorder or panic disorder in last 2 years. 42.5% had at least one co-morbid physical condition. (n=3432) % with any problems Anxiety/ Depression: 96.6 Pain / Discomfort: 73.3 (of which 12.1% extreme) Usual activities: 72.5 Mobility: 33.8 Self-care: 18.3 Pain 49% headache 46% soreness in muscles	Baseline scores: HADS-D score increased with increasing pain severity No pain 11.4 (SD 4.6) Moderate 12.3 (SD 4.3) Extreme 14.4 (SD 4.2)		

			<p>45.8% lower back 44.9% back pain 43.5% pain in joints 30.1% pain abdomen 26.3% pain chest</p> <p>EQ-5D 0.44 (0.31) European population 0.85 (0.23)</p> <p>EQ-VAS 44.8 (20.4) European 77 (20)</p>			
<p>Gunther et al 2008 Germany</p>	<p>Patients with a depressive episode according to the ICD-10 Classification (F32.1, F32.2, F33.1, F33.2), recruited consecutively from 3 departments of psychiatry in Germany.</p> <p>Patients with psychotic or manic symptoms were excluded</p>	<p>EQ-5D UK and German (based on TTO) EQ-VAS WHOQoL-BREF CGI-S, GAF BRAMES</p>	<p>At baseline patients mainly report problems in: Depression and anxiety (78.8%), Usual activities (66.4%), Pain (66.0%), Mobility (28.8%) and Self care (26.9%)</p> <p>Improvement found in Usual activities and Anxiety/depression.</p>	<p>Correlation (Spearman rank due to non-normal distribution of EQ-5D) of EQ-5D index and VAS significant with all other measures.</p> <p>VAS strongest correlation with WHO-BREF (0.642)</p> <p>EQ-5D correlations BRAMES -0.576 WHO-BREF 0.545 CGI -0.539 GAF 0.492 EQ-VAS 0.440 EQ-5D German 0.945</p>	<p>Disaggregating into those who think health (at assessment 18 months after baseline) is worst (16), same (39) or better (49), than at baseline.</p> <p>Also disaggregated into 3 groups based on BRAMES score.</p> <p>For those with perceived worst health EQ-VAS shows slight improvement in mean scores. The EQ-5D indexes both show deterioration for those in worst health (-0.290 for UK index) larger than the improvement for those in better health (0.155). Other measures show a greater change for those with better health. When patients were</p>	<p>N3 term 0.269 in UK and 0.323 in Germany, hence moving away from extreme problems represents a large change in score.</p> <p>Concern regarding the regression estimating meaningful difference.</p> <p>Missing values on the assessments (<1.3%) were replaced by baseline or assessment values.</p>

					<p>disaggregated into 3 groups based on BRAMES score this gives similar results.</p> <p>For t statistic, ES & SRM. EQ-5D (UK and German) t stat, ES and SRM find greater responsiveness to deteriorating health (almost twice as large as clinical measures). Although sig. differences for better health, clinical measures show greater responsiveness.</p> <p>ES for health improvement: CGI most responsive (-0.98 patient-based anchor, -1.35 clinician based anchor), VAS (0.84, 1.19), EQ-5D UK (0.55, 0.65), EQ-5D German (0.41, 0.45).</p> <p>Regression analysis showed sig impact of baseline score (not of age or time since diagnosis), such that the smaller the baseline score the larger the difference between assessments. For EQ-5D the coefficient for EQ-5D UK on self-perceived 'better' was 0.104 and clinically judged better was 0.167 (for the German index 0.059 NS and 0.106) For EQ-5D the coefficient for EQ-</p>	
--	--	--	--	--	---	--

					5D UK on self-perceived 'worst' was -0.470 and clinically judged worse was -0.223 (German - 0.397 and -0.139 NS, respectively)	
Kessing et al, 2006	Patients with depressive disorder identified from the Danish Psychiatric Central Research Register 55.3% female, Mean age 43.8 Number of admissions 2.39.	EQ-5D dimensions only EQ-VAS WHO (Five) well-being index a subscale of PGWB) BDI-21 BDI-Anxiety BDI-Mania	Some problems: Mobility 25.4% Self-care 14.2% Usual activities 58.6% Pain/discomfort 52.0% Anxiety/depression 63.5% EQ-VAS 65.1 (SD 23.2)			
Konig et al 2009 Germany	389 patients with anxiety disorder from 46 GP practices. 23 practices allocated intervention - training - group (n=201), and 23 to control group usual care (n=213)	EQ-5D UK EQ-VAS PHQ-D BAI BDI-II			No significant differences between control and intervention group. BAI, BDI also showed no significant differences.	

<p>Konig et al 2010</p> <p>Germany</p>	<p>389 patients with anxiety disorder</p>	<p>EQ-5D UK EQ-VAS WHO-QoL- BREF BSQ ACQ BAI BDI-II MI (MIA and MIB)</p>	<p>Some problems for Anxiety/Dep 77.4% Pain 71.5% Usual activities 40.8% Mobility 23% Self care 3.9% 21% of sample reported 11122 VAS: Mean (median) was 63.8 (70) EQ5D: Mean (median) 0.66 (0.73) 1% reported 1.0 on VAS, 11.6% on EQ5D.</p> <p>Problems in all dimensions reported more frequently than German population.</p>	<p>At baseline. EQ-5D correlated 0.7 with physical health WHO-QoL, 0.5 with mental health WHO-QoL, 0.58 overall WHO-QoL. -0.58 with BAI, -0.54 with BDI-II 0.4 and below for BSQ, ACQ, MIA, MIB.</p> <p>Almost all EQ-5D dimensions response levels were associated with significant differences in scores of WHO-QoL domains and measures of psychopathology.</p> <p>Those with no problems on anxiety have a mean BAI score of 23.7, those with some problems a score of 9.2 giving an ES d = 1.53. Also strongly associated with mental health WHOQoL d = 1.44 and BDI-II d=1.52</p>	<p>More anxiety if BAI increased more than 0.5 of SD of BAI (n=43) Less anxiety if BAI dec > 0.5SD (n=83), others no shift (n=124)</p> <p>EQ-5D, EQ-VAS, WHO-QoL, BSQ, ACQ all show significant differences between more anxiety, constant and same (t stat). Effects size for EQ-5D -0.99 for more anxiety (more than twice that for other measures). Effect size 0.39 for less anxiety. SRM -0.54 for more anxiety EQ-5D again higher than other measures. 0.46 for less anxiety (BSQ -0.72, WHO-QoL 0.35).</p>	<p>Ceiling effect of EQ-5D not pronounced.</p> <p>Missing items on EQ5D described and only complete scores used.</p>
<p>Lamers 2006</p> <p>Part of RCT in the Netherlands</p>	<p>Patients aged 18-65 with a diagnosis of a major depressive disorder, dysthymic disorder, panic disorder, social phobia,</p>	<p>ED-5D SF-36</p>	<p>At baseline more respondents report having no limitations on the EQ-5D compared to the SF-6D. e.g. 78% no mobility problems and 93.5% no self-care problems</p>	<p>Comparing SCL subgroups</p> <p>Both measures show expected pattern, EQ-5D shows large drop in utility in most severe quartile – not found in SF-6D. SD greater in EQ-5D. Difference between mean and median greater for EQ-5D.</p>	<p>Mean EQ-5D utilities increased from 0.513 to 0.680 at 1.5 years and SF-6D from 0.577 to 0.701. Mean improvement in EQ-5D utilities was lower than for SF-6D in the low severity group and higher for the 2 subgroups with highest severity. Mean improvement in EQ-5D</p>	<p>EQ-5D gives larger health gains – would be translated into lower cost-utility ratios, especially for high severity sub-groups.</p>

	generalised anxiety (classified by DSM-IV)		<p>according to EQ-5D BUT only 18% no limitations in physical functioning in SF-6D. e.g. 65% report 4/5 out 5 for mental health SF-6D only 33% 3/3 for EQ-5D.</p> <p>Mean scores EQ-5D 0.518 (SD .29) SF-6D 0.575 (SD .10) Median EQ-5d 0.689 SF-6D 0.567</p>		<p>increased with increased severity. The SD of differences were lower for SF-6D resulting in SRMs for SF-6D that were at least twice as high as those for EQ-5D for all severity groups. At 1.5 year follow up SRM was 0.466 for EQ-5D and 0.833 for SF-6D.</p>	Missing items on outcome measures not discussed
Mann et al 2009 UK	114 patients with depression participating in RCT on collaborative care (MMD according to SCID), Mean age 42.5, 77% female.	<p>SCID PHQ-9 SF-6D EQ-5D</p> <p>PHQ-9 score Mild (5-9) Moderate (10-14) Mod. Severe (15-19) Severe (20-27)</p>	<p>More patients responded no problems on EQ-5D than SF-6D.</p> <p>Level 1: EQ-5D Mobility 73% Self-Care 83.8% Usual activities 24.3% Pain: 35.8% Anxiety/dep: 5.5%</p> <p>Level 1: SF-6D Physical: 27.9 Role limit: 1.8 Social funct: 10.1 Pain: 22.3</p>	<p>EQ-5D correlated -0.451 with PHQ-9 at baseline and -0.638 at 3 month follow up SF-6D correlated -0.351 with PHQ-9 at baseline and -0.833 at 3 month follow up.</p> <p>EQ-5D Mild 0.645 (SD 0.23) Moderate 0.656 (SD 0.21) Mod severe 0.558 (SD 0.27) Severe 0.337 (SD 0.29)</p> <p>SF-6D Mild 0.642 (SD 0.09) Moderate 0.601 (SD 0.09) Mod severe 0.548 (SD 0.07) Severe 0.544 (SD 0.06)</p>	<p>Change from baseline EQ-5D increase 0.147 (change in median scores 0.069) SF-6D increased 0.082 (change in median scores 0.07) Both significant</p> <p>62% assessed as in remission at follow-up according to SCID Remission at follow up: EQ-5D 0.759 (SD 0.25), SF-6D 0.707 (0.12) MDD at follow up: EQ-5D 0.506 (0.37), SF-6D 0.550 (0.09).</p> <p>Mean improvement between baseline of overall study population and follow up utility</p>	<p>Describes direct TTO or SG with patients with depression as “impractical” (p574)</p> <p>Despite reporting more problems with SF-6D, EQ-5D shows greater change – may arise due to differences in valuation, at low end EQ-5D much lower.</p>

			<p>Mental: 0 Vitality: 1.8</p> <p>Highest level mental health EQ-5D: 5.5/65.1/29.4 SF-6D: 0/1.8/11.6/54.5/32.1</p> <p>86.6% of patients reported feeling tense / downhearted or low most or all of the time in the SF-6D but only 29.4% reported extreme problem with anxiety and depression on EQ-5D</p> <p>Suggests more people reporting problems, and severe problems with SF-6D. SF-6D descriptors are "feelings" compared to anxiety / depression which "may be perceived as directly related to mental disorders"</p>		<p>for those with remission was EQ-5D 0.243, SF-6D 0.140.</p> <p>EQ-5D showed larger health gains at follow up for all patients, and for those in remission.</p>	
--	--	--	--	--	--	--

			(p574) 57.1 report highest level on vitality. No comparable measure for EQ-5D.			
Mittal et al 2006 USA	324 veterans receiving primary care for depression (score 12 or more on PHQ). Mean age 59.8, 9% female.	MINI SCL-20, (symptom checklist to identify depression severity) SF-12v QWB-SA (utility score from published weights – Kaplan et al 1997) Duke social support and stress scale	69% were comorbid with anxiety disorder 58% GAD 29% PTSD 11% Panic disorder Mean QWB-SA was 0.39 (SD 0.11)	Regression analysis found that GAD and PTSD independently predicted QWB-SA but panic disorder was not significant.		Missing data (2%) were impute, sensitivity analysis dropped missing data found similar results.
Mychaskiw et al, 2008 USA Abstract only	374 non-depressed patients with GAD treated with pregabalin, venlafaxine-XR or placebo	EQ-5D UK HAM-A		EQ-5D scores decreased as anxiety symptom severity increased. Normal (HADS 0-7) 0.83 Mild (HADS 8-10) 0.78 Moderate (HADS 11-14) 0.60 Severe (HADS 15-21) 0.30	Those achieving Functional remission at 8 weeks (SDS < 5): EQ-5D 0.87, those not in remission: EQ-5D 0.61 Symptomatic remission at 8 weeks HAMA score 7: Remission EQ-5D	

	n at endpoint ranged from 323 to 360.			All significant except normal versus mild.	0.84, No remission EQ-5D 0.60 HAMA score 10: Remission EQ-5D 0.83, No remission EQ-5D 0.57	
Petrou et al 2009 UK	623 women in an RCT of postnatal support. 1046 declined. 312 allocated to control group receiving usual care. 311 offered 10 visits community postnatal support worker. 493 had complete data Aged >=17 given birth to live baby.	EQ-5D SF-6D EPDS Self-rated (SR) health status (excellent, v. good, good, fair, poor) Taken six months postpartum	Mean EQ-5D 0.861 SD 0.181 (95CI 0.844 - 0.877) SF-6D 0.809 SD 0.140 (CI 0.796-0.822) (significantly different from EQ-5D). Median EQ-5D 0.848 (IQR 0.796-1) SF-6D 0.830 (IQR 0.706-0.938) Minimum EQ-5D - 0.077 Minimum SF-6D 0.374 177 women (35.9%) had EQ-5D of 1.0 and SF-6D of < 1 (29.2% and 34.1% identifying problems with mental health and vitality respectively). Only 1 women had	Both show monotonically decreasing scores in line with SR health status. Relative efficiency statistic - how well can they detect differences in SR health status and EPDS. Ratio of the square of the t-statistic of the comparator instrument over the square of the t statistic of the reference instrument. Found SF-6D more efficient by 29% to 423.6%. When sample restricted (dropped 12 with low EQ-5D) SF-6D still more efficient. Also more efficient using EPDS profiles (between 129.8% and 161.7%). Receiver operating characteristics (ROC) curves find area under curve greater for SF-6D hence same conclusion - SF-6D better at discriminating. But all but one analysis differences not significant.		Why 1. SF-6D taps into broader aspects of health and QOL 2. SF-6D greater number of response items - possibly greater sensitivity 3. Wording on SF-6D, which includes positive and negative items gives greater sensitive to maternal health 4. Longer time frame (past 4 weeks) v EQ-5D which is today, might increase sensitivity. Missing items on outcome measures not discussed

			SF-6D of 1 and identified moderate pain/discomfort on EQ-5D. No differences in EQ-5D by demographic characteristics.			
Peveler et al, 2005 UK RCT to receive a TCA or SSRI or lofepramine.	Of 388 patients with new episode of depression referred to study 67.3% female, mean age 42.5. n=327 randomised	EQ-5D HAD-D CIS-R PROQSY SF-36 Depression free week			EQ-5D of 3 groups showed improvement of about 20 points, most of which occurred in first 3 months. Baseline (n=261) EQ-5D 0.5586 (SD 0.275) Month 2 (n=172) EQ-5D 0.763 (SD 0.195) Month 12 (n=162) EQ-5D 0.777 (SD.194) No sig. differences between 3 groups.	
Pyne et al 2010 USA RCT rural telemedicine-based collaborative care for depression vs usual care	395 primary care patients screened positive for depression using PHQ-9, 360 completed 6 month follow up, 335 completed 12 month. Excluded schizophrenia, suicide intention,	SF-6D from SF-12 QWB Depression free days	Baseline SF-6D Usual care (n=179) 0.53 (SD 0.12) Intervention (n=141) 0.54 (SD 0.14) Baseline QWB Usual care (n=179) 0.42 (SD 0.11) Intervention (n=141)		Depression free days find no significant differences QWB showed no difference SF-6D showed sig. difference between intervention and usual care different. (ICER \$85,634/QALY)	

	pregnancy, substance dependence, bipolar		0.43 (SD 0.13)			
Pyne et al 1997 US	<p>100 patients with primary diagnosis of major depression, 60 outpatients and 40 inpatients (from Veteran Affairs Medical Centre). Control group (n=61) identified by VA medical centre staff, without current or past diagnosis of mental illness</p> <p>Diagnosed by SADS or SCID criteria</p> <p>Patients: mean age 48.5, 18% female Control group: mean age 47.4, 1.6% female</p>	BDI HRSD QWB		<p>QWB scale Control group: 0.813 (n=61)</p> <p>Rated using HRSD (n=95) Mild: 0.676 Moderate: 0.645 Severe: 0.554</p> <p>Rated using BDI (n=87) Mild: 0.698 Moderate: 0.643 Severe: 0.597</p> <p>Regression analysis found no sig. relationship with QWB and age, gender, or family history of mental illness</p> <p>BDI and HRSD strong predictors of QWB, even when controlling for presence of comorbid axis III diagnosis</p>		

<p>Reed 09 FINDER 12 European countries</p>	<p>Patients (>= 18) with clinical depression enrolled prior to commencing antidepressant treatment. Of 3468 at baseline, 343 had no follow up data, 271 data at 3 months only, 2854 had 3 & 6 months. Age not given. Gender not given.</p>	<p>EQ-5D EQ-VAS SF36, HADS-D, HADS-A, SSI-28-item Somatic Symptom Inventory, Pain VAS</p>	<p>Baseline EQ-5D 0.44 and 44.8 VAS, showed improvement at 3 & 6 months (other data on a graph)</p>	<p>Regression analysis found significant negative relationship between number of previous depressive episodes and duration of current episode. Also negatively related to somatic symptoms and VAS pain.</p>		<p>Age, country and SSI-somatic score at baseline related to probability of dropout. Missing items on outcome measures not discussed</p>
<p>Revicki et al 2008 USA KPNW study</p>	<p>297 patients with GAD, 72% female, mean age 47.6 HAM-A score Asymp. <= 9 Mild 10-15 Moderate 16-24 Severe >=25</p>	<p>SCID SIGH-A GAD-Q-IV, Q-LES-Q-SF, PHQ, SF-12, HUI2, HUI3</p>	<p>Baseline HAM-A score 16.7 HUI2 0.54 (SD 0.2) HUI3 0.46 (SD 0.3) SF-6D 0.62 (SD 0.1)</p>	<p>Correlations with HAM-A HUI2 -0.52 HUI2 -0.54 SF-6D -0.52 Correlations with GAD-Q-IV HUI2 -0.43 HUI2 -0.44 SF-6D -0.38 Correlations with PHQ HUI2 -0.52 HUI2 -0.57 SF-6D -0.64 Suggest utility measures are</p>	<p>Presents outcomes by HRQL at 3 months by anxiety severity – shows similar picture. HUI2 and HUI3 not able to sig. distinguish between asymptomatic and mild groups.</p>	<p>Authors note that the greater sensitivity of HUI3 compared to HUI2 may be due to content of the emotional domain which focuses on happiness and depression rather than worry and anxiety.</p>

				<p>most strongly correlated with depression scores.</p> <p>HUI2 Asymp. 0.70 (SD 0.2) Mild 0.59 (SD 0.2) Moderate 0.5 (SD 0.2) Severe 0.36 (SD 0.2)</p> <p>HUI3 Asymp. 0.68 (SD 0.2) Mild 0.54 (SD 0.3) Moderate 0.39 (SD 0.3) Severe 0.17 (SD 0.3)</p> <p>SF-6D Asymp. 0.72 (SD 0.1) Mild 0.64 (SD 0.1) Moderate 0.60 (SD 0.1) Severe 0.53 (SD 0.1)</p> <p>HUI3 most sensitive to increased anxiety All compare favourably to other clinical measures.</p> <p>All differences significant</p> <p>Regression analysis found that symptom measures explained 38% of variance of HUI2, 42% of HUI3 and 46% of SF-6D</p>		
--	--	--	--	--	--	--

Saarni 2007 Finland	Population survey, aged 30 and over. Included assessment of 12-month prevalence of depressive anxiety or alcohol disorders (DSM-IV).	EQ-5D UK 15D measure with Finnish valuations Munich version of the Composite International Diagnostic Interview (M-CIDI) used to assess 12-month prevalence of depressive, anxiety or alcohol use disorders	5219 had data on EQ5D and M-CIDI, 65% of sample had EQ5D and M-CIDI data.	47% scored full health on EQ-5D (30%) of those with psychiatric disorder. Unadjusted scores for population were 0.83 for EQ5D and 0.72 for those with any psychiatric diagnosis. Controlling for socio-economic status, somatic comorbidity and psychiatric comorbidity Depressive disorders reduces EQ-5D -0.091 (CI-0.114 to -0.068) Anxiety disorders reduced EQ-5D -0.114 (-0.144 to -0.085) GAD reduced EQ-5D -0.110 (-0.158 to -0.061) MDD -0.058 (-0.079 to -0.036) Dysthymia -0.122 (-0.167 to -0.077) Panic disorder NS Social phobia -0.102 (-0.166 to -0.039) Agoraphobia NS		Only fully completed EQ-5D were included. For 15D only those with 12 more responses included, and missing data were imputed.
Sapin 2004 France	Outpatient population consulting at GP for new episode of major depressive disorder (MDD)	Patient reported: EQ-5D SF-36 QLDS Clinical/phy	At baseline mean EQ-5D was 0.33 (+/- 0.25) range -0.59 to 0.85. 8% had EQ-5D worst than death. Baseline - no difficulties in	Significant differences in EQ-5D by disease severity level (CGI-s) e.g. at baseline 0.12 difference between slightly/moderately ill and markedly ill. Slightly ill and markedly ill scores differed by	4 weeks mean EQ-5D 0.68 (+/- 0.24 range -0.11-1) 8 weeks mean EQ-5D 0.78 (+/- 0.21 range -0.08 to 1) Extreme difficulties on anxiety & depression was 77.9% at baseline moved to 9.3% at D56.	

	<p>according to DSM-IV, aged 18 and over, not treated with any antidepressants prior to inclusion.</p> <p>Exclusion Symptoms suggests schizophrenia or psychotic symptoms</p>	<p>sician reported: MADRS CGI-S</p>	<p>mobility 73.5%, self-care 82.3% usual activities 24.8% pain discomfort 23.9%, anxiety depression 0.9%.</p> <p>MADRS score at D56 <= 12 classed as remitters others as non-remitters</p> <p>If decrease of at least 50% from baseline classed as responder, others non-responders</p> <p>Responder remitters Responder non-remitters Non-responder</p>	<p>0.3.</p> <p>Correlation between SF-36 MHC and EQ-5D was 0.49, 0.56 and 0.63 at baseline, D28 and D56 respectively.</p> <p>Correlations between QLDS range -0.43 to -0.68 at D56.</p> <p>EQ-5D by response group at baseline, D28 and D56</p> <p>Non-responders 0.30 (SD 0.27), 0.54 (SD 0.30), 0.58 (SD 0.28)</p> <p>Responder-remitters 0.35 (SD 0.24), 0.76 (SD 0.18), 0.85 (SD 0.13)</p> <p>Responder-non-remitters 0.21 (SD 0.25), 0.54 (SD 0.26), 0.72 (SD 0.20)</p>	<p>Regression analysis dependent variable the difference (baseline – assessment) EQ-5D independent other measures explained 40% of the variation in EQ-5D.</p>	
<p>Serfaty et al, 2009</p> <p>UK</p> <p>RCT of CBT for older people with depression</p>	<p>N=204</p> <p>Age 65 or older, with depression screened by 15-item geriatric depression scale or BDI-ii score 14 or more. Mean age 74.1, 79.4% female</p>	<p>BDI-II BAI-II SFQ EQ-5D (no reference to scoring system)</p>			<p>CBT: EQ-5D: Baseline: 0.50 (0.32) n=70 4 months: 0.53 (0.34) n=61 10 months: 0.54 (0.33) n=56</p> <p>Taking Control treatment: EQ-5D Baseline: 0.52 (0.31) n=67 4 months: 0.55 (0.39) n=57 10 months: 0.52 (0.32) n=53</p>	<p>Missing data were imputed.</p>

	Measurement: baseline, 4 months, 10 months				<p>Treatment as usual: EQ-5D Baseline: 0.46 (0.29) n=67 4 months: 0.47 (0.38) n=55 10 months: 0.52 (0.31) n=50</p> <p>CBT: BDI-II Baseline: 27.3 4 months: 18.4 10 months: 18.3</p> <p>TAU: BDI-II Baseline: 27.7 4 months: 20.3 10 months: 20.8 Significant difference in improvement in CBT versus TAU for BDI-II</p> <p>When adjusted for baseline BDI- ii GP numbers, baseline outcome value, time period, BDI-II significantly better than TAU or CBT. EQ-5D found CBT significantly better than TC but not significantly different to TAU.</p>	
Sobocki et al, 2007 Sweden	Subjects attending GP at 56 centres in Sweden. N=447. Baseline data on n=394.	EQ-5D UK CGI-S (1-7) (Severity) CGI-I (Improv.)	24% had mild depression, 69% moderate (CGI-S score of 4 or more), 16% severe (CGI-S score 5-7)	Sig. differences in EQ-5D by disease severity as assessed by CGI-S Mild 0.6 (0.54-0.65) Moderate 0.46 (0.30-0.48), Severe 0.27 (0.21-0.34)	Baseline EQ-5D 0.47 (0.44-0.5) range 0-1. (n=394) First follow up visit EQ-5D 0.60 (0.57-0.62) (n=378) At about 6 months EQ-5D 0.66 (0.53-0.75) (n=28)	Numbers in study at 6 months very low.

	Mean age 47, 67% female. Aged > 18 diagnosis of depression (according to centres practice), initiating new treatment with antidepressants		59% had a least one comorbidity 56% physical comorbidities 9% psychiatric. Cut off EQ-5D at zero but 8% rated as worst than dead. (authors note that recalculating with SWD allowed does not substantially effect the results).	Sig. difference between mild/moderate but not between moderate and severe. Pattern similar at follow up (0.76/0.65/0.52) Regression analysis – explanatory variables explained 23% of EQ-5D variation. Demographic variables not significant.	At last follow up visit EQ-5D 0.69 (0.67-0.72), corresponding to increase in utility of 0.23 (p< 0.0001) EQ-5D increased 40 to 63 at about 6 months. Increase by severity classification Mild: 0.16 (0.11-0.23) Moderate: 0.22 (0.18-0.26) Severe: 0.35 (0.25-0.44)	
Stein et al, 2005 USA CCAP trial	N=480 outpatients with anxiety disorder, 63% female	SF-6D from the SF-12 WHO Disability scale	Adjusting for covariates any anxiety disorder lowered utility values by -0.122 and co-morbid major depression by -0.087. Adjusting for covariates (comorbidities, socio-economic factors) utility primary care patients without anxiety or depressive disorder			

			0.80 (0.78-0.82): With anxiety disorder alone 0.68 (0.66-0.70), with depressive disorder alone 0.72 (0.66-0.79), both 0.59 (0.57-0.61)			
van Straten et al, 2008 The Netherlands	N=213 recruited via media N=107 web self help intervention for depression, anxiety and work-related stress N=106 control group	EQ-5D CES-D MDI HADS SCL-A MBI, work related stress - 3 subscales.			EQ-5D Control Pre:0.61 Post: 0.66 Intervention all: Pre 0.62 Post 0.73 Intervention complete: Pre 0.63 Post 0.8 Effect size (Cohens d) All (n=107), course completers (n=59) CES-D 0.5 (0.22-0.79), 0.67 (0.32-1.02) MDI 0.33 (0.03-0.63), 0.56 (0.22-0.9) SCL-A 0.42 (0.14-0.70), 0.51 (0.18-0.84) EQ-5D 0.31 (0.03-0.60), 0.44 (0.11-0.77) HADS 0.33 (0.04-0.61), 0.48 (0.15-0.82) MBI not sig.	Missing data were imputed by regression analysis

Supina et al 2007 Canada	Alberta Mental Health Survey, stratified random sample. Sample size n=5,410 (77% return), n=5,383 successful data. Mean age 40.8. Female 61.2%	EQ-5D EQ-VAS MINI	MDE (recurrent and current) alone 2.6% Anxiety disorders only 11.2% MDE and anxiety 5.2% Neither 80.9%	Anxiety only (n=601) EQ-5D 0.84 (0.83-0.85). EQ-VAS 76.68 MDE only (n=140) 0.83 (0.81-0.85), VAS 70.82 Anxiety and MDE (n=280) 0.70 (0.69-0.72), VAS 64.17 Neither (n=4338) 0.92 (0.91-0.92), VAS 84.68		
Swan et al, 2004. RCT of Coping with Depression (CWD) course. UK	Inclusion: Primary diagnosis of chronic or recurrent depressive disorder; current depressive episode of at least moderate severity (ICD=10 F32.1-F32.2, F33.1-F33.2); inadequate or poor response to previous treatments. Aged 18-65 N=76 entrants, 31 completed CWD, n= 26 (34%) attended	BDI-II. BSI which generates the GSI EQ-5D (no reference to scoring system) Allocated to worse, unchanged, improved or recovered based on BDI-II and GSI	Baseline EQ-5D 0.44 (SD 0.41, range -0.24 to 1.0) (n=76)		EQ-5D (n=26) Baseline: 0.49 (SE 0.07) (0.34-0.64) Week 12: 0.65 (SE 0.06) (0.52-0.79) Week 26 0.68 (SE 0.06 (0.55-0.82) Significant improvement in BDI and GSI (baseline- 12; baseline-26).	

	follow up. No differences in clinical or demographic characteristics between completers and drop outs.					
Wells et al 2007 Partners in Care data USA	<p>Patients with recent depressive disorder or sub-threshold depression (current depressive symptoms but no disorder)</p> <p>Usual care n=214 Quality improvement n=532 (of which Medication n=249 Therapy n=283)</p> <p>Usual care: mean age 43, 71% female Quality improvement 44, 77% female</p>	SF-12 with weights derived from a convenience sample of primary care patients using SG (see Lenert et al 2000 above)			<p>Incremental effect of quality improvement over usual care</p> <p>For depressive disorder (n=746) Days of depression burden over 24 months -46 (95% CI -84; 8) p = 0.02</p> <p>Days of employment over 24 months 23 (5 41) p= 0.1</p> <p>QALY gain 0.02 (0 to 0.4) p=0.1</p> <p>For sub-threshold (n=502) QALY gain 0.02 (0;0.4) p=0.06</p> <p>Days depression burden -31 (-71;9) p= 0.13</p> <p>Days employment 15 (-1;31) p = 0.07</p>	

<p>Zivin et al 2008</p> <p>USA</p>	<p>n=87,797 Veterans, mean age 60, 10% female</p> <p>Identified from VA depression registry and VA outpatients</p> <p>n=58,442 with depression</p>	<p>SF-6D from the SF-12</p>	<p>VA with depression Utility 0.57 (SD 0.13)</p> <p>VA without depression 0.63 (SD 0.14)</p>			
--	--	---------------------------------	--	--	--	--