EEPRU

Economic Evaluation of Health and Social Care Interventions Policy Research Unit

RESEARCH REPORT

Supporting the routine collection of patient reported outcome measures in the National Clinical Audits for assessing cost-effectiveness

Work Package 1

What patient reported outcome measures should be used in the 13 health conditions specified in the 2013/14 National Clinical Audit programme?

APPENDIX H, PSYCHOLOGICAL THERAPIES

Authors: Roberta Ara, Ana Duarte, Sue Harnan, Jo Leaviss, Steve Palmer, Mark Sculpher, John Brazier

Correspondence to: Roberta Ara, HEDS, ScHARR, University of Sheffield, Regent Court, 30 Regent Street, Sheffield, S1 4DA. Email: r.m.ara@sheffield.ac.uk

RR0034

March 2015

The University Of Sheffield.

University of York

The Policy Research Unit in Economic Evaluation of Health and Care interventions is funded by the Department of Health Policy Research Programme. It is a collaboration between researchers from the University of Sheffield and the University of York.

The Department of Health's Policy Research Unit in Economic Evaluation of Health and Care Interventions is a 7 year programme of work that started in January 2011. The unit is led by Professor John Brazier (Director, University of Sheffield) and Professor Mark Sculpher (Deputy Director, University of York) with the aim of assisting policy makers in the Department of Health to improve the allocation of resources in health and social care.

This is an independent report commissioned and funded by the Policy Research Programme in the Department of Health. The views expressed are not necessarily those of the Department.

INDEX

1 BACKGROUND

2 OVERVIEW

3 METHOD

3.1 Psychometric properties (WP1.1)
3.2 Alternative measures (WP1.2)
3.3 Evidence required for economic evaluations (WP1.3)

4 RESULTS FOR PSYCHOLOGICAL THERAPIES

4.1 Evidence of appropriateness of EQ-5D in psychological therapies
4.2 Alternative measures in psychological therapies
4.3 Evidence for economic evaluations in psychological therapies
4.4 Recommendations for psychological therapies

5 SUMMARY

6 REFERENCES

7 APPENDIX

TABLES

Table 1 Selection of most appropriate review for psychological therapies

Table 2 Summary of evidence on EQ-5D for patients receiving psychological therapies

Table 3 Summary of existing models used in psychological therapy TAs

Table 4 Recommendations and associated future research for psychological therapies

Table 5 Summary of evidence currently available for recommended measure(s)

FIGURES

Figure 1 Modelling approach used in psychological therapy HTAs

Acronym	Definition
AE	Adverse events
ACQ	Agoraphobic cognitions questionnaire
AMSTAR	Assessing the quality of systematic reviews
ARM	Agnew relationship measure,
BAI	Beck Anxiety Inventory
BCVA	Best corrected visual activity
BDI	Becks depression index
BRAMES	Bech-Rafaelsen melancholia scale
BSQ	Body shape questionnaire
CBT	Cognitive based therapy
CES-D	Center for epidemiologic studies - depression
CG	Clinical guideline
CGI-S	Clinical global depression scale - severity
DH	Department of Health
EQ-5D	EuroQol 5 dimensions
EPDS	Edinburgh postnatal depression scale
FR	Future research
GAD-7	Generalised Anxiety Disorder Assessment-7
GAF	Global assessment of functioning
HADS	Hospital anxiety and depression scale
HAM-A	Hamilton anxiety scale
HAM-D	Hamilton depression scale
HRQoL	Health related quality of life
HS	Health states
HTA	Health technology assessment
MADRS	Montgomery-Asberg depression rating scale
MBCT	Mindfulness-based cognitive therapy
MDD	Major depressive disorder
MDE	Major depressive episode
MIA	Mobility inventory – avoidance alone
MIB	Mobility inventory – avoidance accompanied
MTA	Multiple technology assessment
NAPT	National Audit of Psychological Therapies for Anxiety and Depression
NCA	National Clinical Audit
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
PHQ-9	Patient health questionnaire
PQWB	Psychological General Well-Being Index
PR	Potential recommendations
PROM(s)	patient reported outcome measure(s)
QALY	Quality adjusted life year
QLDS	Quality of life in depression scale
R&D	Research and development
RCT	Randomised controlled trial
SCL	Symptom checklist
SF-6D	Short form 36
SMD	Standardised mean difference
SR	Systematic review

SRM	Standardised response mean
TA	Technology Appraisal
TAU	Treatment as usual
UK	United Kingdom
VAS	Visual analogue scale
WHOQOL-BREF	WHO quality of life - BREF
WP	Work package
WSA	Work and Social Adjustment scale
WTE	Whole time equivalent

1. BACKGROUND

EEPRU was approached by Jason Cox (R&D Division) to prepare a programme of research to support the appropriateness of, and use of, patient reported outcome measures (PROMs) collected for the National Clinical Audit (NCA). The EEPRU programme was informed by a Research and Development (R&D) template prepared by Simon Bennett, Steve Fairman and Keith Willett at NHS England.

The purpose of introducing PROMs into the NCA programme is to be able to 1) compare performance between providers and commissioners in the National Health Service (NHS), 2) compare the cost-effectiveness of alternative providers in delivering the specific services (i.e. linking outcomes and resource use), and 3) assess the cost-effectiveness of alternative interventions and other changes in the NHS. The intention is to introduce PROMs across a range of conditions over the next 3 years commencing with 13 conditions in the 2014/15 NCA programme.

The agreed research programme consists of 3 concurrent work packages (WP) as described in the document submitted to the DH (8th November 2013). The current document provides details on the objectives, methodology and results for Work Package 1 (WP1): to determine what PROMS should be used in the 13 health conditions specified in the 2014/15 NCA programme.

2. OVERVIEW

WP1 is split into three separate components consisting of:

- WP1.1 To examine whether the Euro-QoL 5 dimensions (EQ-5D) is appropriate in the 13 health conditions specified in the 2013/14 NCA programme.
- WP1.2 To identify what measure could be used when the EQ-5D is not appropriate in the 13 health conditions, taking into account that the proposed measure would be used to generate preference-based utility measures (either directly through existing preference-based weights, or indirectly through existing mapping functions suitable for the proposed measure).
- WP1.3 To identify the evidence required to address questions of cost-effectiveness using the NCA data.

Each component consists of a series of reviews of the literature.

This Appendix provides the detailed results for recipients of psychological therapies and should be read in conjunction with both the main report and the methods/search strategy appendices.

3. METHOD

The full detailed methodology used is provided in Appendix A, including the search strategy, selection criteria for studies included, and data extraction etc. In summary, a review of the literature was undertaken to assess the appropriateness of the EQ-5D in terms of classic psychometric criteria (WP1.1); where the EQ-5D was not considered appropriate, additional searches were undertaken to identify alternative measures (WP1.2); and finally, existing health technology appraisals (HTAs) were reviewed and data requirements were compared with variables currently collected in the psychological therapy audit (WP1.3).

3.1 **Psychometric properties** (WP1.1)

Assessments reported in the included studies were categorised according to the following definitions:

Acceptability

Data relating to how acceptable the measure was to the person completing it, expressed as the proportion of completed surveys, or the proportion of missing data.

Reliability

There are two main definitions for reliability, a) the degree to which a measure reproduces the same results in an unchanged population and b) the degree to which a measure reproduces the same results when completed by different assessors (e.g. patient and proxy report). In both cases, reliability can be assessed by re-testing, and calculating the correlations or difference between tests. In case a) the comparison may be between the same populations separated by time, where no change in health state was observed (as compared to using an alternative condition specific or generic measure). In case b) the measure may be completed by multiple people (proxies) on the patient's behalf and their responses compared with those of the patient. Where the outcome measure is specifically designed for self-report by patients, this test of reliability may be expected to produce less agreement.

Construct validity

This is an assessment of how well an instrument measures what it intends to measure. Two main definitions are used in this review.

a) Known group validity, where estimates for groups that are known to differ in a concept of interest are compared either qualitatively or statistically. The known groups may be defined using other measures, according to clinical categorisation.

b) Convergent validity assesses the extent to which a measure correlates with other measures of the same or similar concepts. Correlation coefficients were considered low if <0.3, moderate if between 0.3 and 0.5, and strong when >0.5.

Responsiveness

a) Change over time. This is an assessment of whether measurements using the instrument can detect a change over time, where a change is expected. This may be before and after an intervention, or through progression of a disease. Evidence was considered to be good where a t-test was significant, though weaker evidence to support responsiveness was considered where there was a change in the expected direction, but was not statistically significant or not tested. Effect size and standardised response mean were also acceptable assessments of responsiveness.

b) Ceiling and floor effects were also considered to be indicators of responsiveness. Assessments of ceiling effects include the proportion of patients who score full health within a group of patients with known health detriments. A ceiling or floor effect can affect the sensitivity of the measure in detecting changes over time in patients at the extremes of the measure (for example those with severe disease activity and those with just minor symptoms of the condition).

3.2 **Alternative measures** (WP1.2)

Searches were conducted to identify exsting reports and guidelines relating to other measures that could be used in depression and anxiety. The results of WP1.1 suggested that the EQ-5D was appropriate for depression, with less certainty about its use in patients with anxiety. WP1.2 therefore concentrated on other measures for anxiety.

3.3 **Evidence required for economic evaluations** (WP1.3)

The existing health technology assessments (HTAs) were reviewed alongside the variables currently collected in the NCA to determine if clinical or PROM data routinely collected in the NCAs would suffice to address questions of cost-effectiveness, and to identify any gaps in the evidence that would be required to compare providers, or the cost-effectiveness of interventions or policies.

4. RESULTS FOR PSYCHOLOGICAL THERAPIES

4.1 Evidence of appropriateness of EQ-5D in psychological therapies (WP1.1)

4.1.1 Selection of systematic review

Two systematic reviews were identified through expert sources.(1),(2) The process of selection of the most appropriate review is documented in Table 1. Peasgood et al. was selected as it provides more detail about the psychometric properties of the EQ-5D, and is more recent than the Oxford review.(2)

Table 1: Selection of most appropriate review for psychological therapies

Review	Search date	Relevance of review	Quality of search	Quality of review	Selection
Oxford review (2)	Unclear	Question relevant, but too little psychometric data provided	Reliance on pre-existing database, with additional searches, but full strategy not provided.	No QA; no search numbers; other methods (DE, SS) unclear	Exclude – less recent than Peasgood, less psychometric detail than Peasgood
Peasgood et al. 2012(1)	December 2010	Question relevant, psychometric data provided	Several databases searched, reference lists also searched. Search strategy NR but available from authors.	QA performed; search numbers provided; DE and SS methods described (one reviewer)	Include

QA, quality assessment; DE, data extraction; SS, study selection.

4.1.2 Structured abstract for Peasgood et al.(1)

Purpose of review

The review aimed to assess the construct validity and responsiveness of EQ-5D and Short-Form 6 dimensions (SF-6D) measures in anxiety and depression.

Methods of review

Search and study selection: Eleven databases were searched. Searches were limited to English language. No details of databases searched were provided in the review. Electronic searches were conducted in December 2010. The full search strategies were not reported but were available from the authors. Reference lists of included papers were hand searched.

Inclusion criteria: Studies were included in the review if they satisfied the following criteria: they contained health related quality of life (HRQoL) data as measured by any preference-based health

measure for adults with depression or anxiety. Study design could include controlled trials or studies examining the burden of illness in depression or anxiety. Studies had to contain data from the HRQoL instrument that allowed measurement of construct validity (convergent or known groups), or responsiveness of the measure.

Exclusion criteria: Studies were excluded from the review if the study population did not have a primary diagnosis of depression, i.e. it was comorbid to another condition. Studies that only contained data relating to the visual analogue scale of the EQ-5D were excluded.

Data extraction and synthesis: Data were extracted by one reviewer using a newly developed form, designed for specific use in a wider review. Due to heterogeneity between studies, a narrative synthesis was performed and data tabulated according to the psychometric quality assessed, namely construct validity and responsiveness. These were defined by the review authors as follows: *Construct validity*, the degree to which an instrument measures the construct it is designed to measure and in the settings it is designed to measure. This can be measured by one of two methods. Known or extreme groups: where in two groups who differ in a trait or behaviour , one group is expected to score significantly higher or lower compared with the other group (definition from Streiner 2003);(3) Convergent validity: where the relationship between two instruments measuring the same construct is assessed by Pearson's product movement correlation or Spearman's rank correlation. Secondly, r*esponsiveness* was defined as the extent to which an instrument can detect a clinically significant or practically important change over time (definition from Walters 2009).(4) Effect sizes for responsiveness were most commonly calculated using the standardised response mean (SRM) or Cohen's D statistic. Effect size thresholds for Cohen's D were: 0.2 was defined as small, 0.5 was defined as moderate, and 0.8 was defined as large.

Results of the review

A total of 26 studies were identified that provided data relating to the construct validity or responsiveness of the EQ-5D or the SF-36. 21 of these studies evaluated the construct validity or responsiveness of the EQ-5D. The remaining studies did not report data relating to the construct validity or responsiveness of the EQ-5D and therefore do not meet the inclusion criteria of WP 1.1 and will not be discussed here.

Studies were conducted in a wide range of countries. One study used the German EQ-5D,(5) and one used the UK EQ-5D 15D measure with Finnish valuations.(6) Whilst several of the remaining studies

specified use of the EQ-5D UK, no further details of which version of the EQ-5D are provided for many studies, it is therefore unclear whether the UK version was used. Five studies were conducted in the UK.(7-11) Three studies were multinational(12-14), one was conducted in France(15); two in Germany(16;17); three in the Netherlands(18-20); one in Canada(21); one in Sweden(22), two in Turkey(23;24), and one in the USA(25).

A range of measures were used to assess the construct validity and/or responsiveness of the EQ-5D. Measures used for comparison included symptom measures such as: Center for epidemiologic studies – depression (CES-D)(20); Hamilton depression rating scale (HAM-D) (23;24); Beck anxiety inventory (BAI)(11;16;17); Beck depression inventory (BDI)(9;11;16;17); Hamilton anxiety rating scale (HAM-A)(25); Patient health questionnaire - depression (PHQ-D)(7;16); Montgomery-Asberg depression rating scale (MADRS)(13;15;18); hospital anxiety and depression scale (depression) HADS-D(10;12;14;20); hospital anxiety and depression scale (anxiety) (HADS-A)(12;14); Edinburgh postnatal depression scale (EPDS)(8). Functional measures were also used as comparators: Clinical global impression scale – severity (CGI-S)(5;15;22), and global assessment of functioning (GAF)(5). Other generic quality of life measures were also included: World Health Organisation quality of life – bref (WHOQOL-BREF)(5), Bech-Rafaelsen melancholia scale (BRAMES)(5), quality of life in depression scale (QLDS)(13;15), and short form – 36 (SF-36)(7;10;19) and short form – 6 dimensions (SF-6D)(8;14;15).

Population characteristics differed across studies. Mean ages ranged from 39.6 years(23) to 74.1 years.(11) Ten studies focused on individuals with depression(5;9;10;12-15;18;23;24), 3 studies focused on individuals with anxiety(16;17;25), 3 focused on individuals with either depression or anxiety(6;19;20). The remaining 2 studies were surveys of the general population, aiming to identify individuals with postnatal depression(8) or depression or anxiety(21). The number of withdrawals was not reported.

Study designs were not always reported. Where reported, most studies consisted of controlled trials.(7-11;13;16-18;24;25) Four studies were population surveys.(6;15;21;22)

Construct validity (known group): Ten studies reported data on construct validity using the known-groups method.(6-8;14;15;17;19;21-23;25) Patients were grouped by disease severity(7;15;17;25), single/recurrent depression groups(23), symptom checklist (SCL) subgroups(19), or self-report health status group. (8) Most studies found that the EQ-5D was able to distinguish between groups,

although Aydemir et al. found no significant differences between individuals having a single episode compared to those having recurrent depressive episodes, with effect sizes of 0.45 and 0.41 respectively. (23) Saarni et al. showed that the EQ-5D was able to distinguish between depression and anxiety disorders, with lower EQ-5D scores for people with depression (-0.091 lower than the general population), anxiety (-0.114 lower), generalised anxiety disorder (GAD) (-0.110 lower), major depressive disorder (MDD) (-0.058 lower), dysthymia (-0.122 lower), and social phobia (-0.102 lower). No significant differenences were found for panic disorder or agoraphobia.(6) Supina et al. also showed that the EQ-5D could distinguish between groups of patients with anxiety or depression, with mean EQ-5D scores for individuals with major depressive episode (MDE) of 0.83, scores for individuals with anxiety 0.84, scores for individuals with both conditions reducing to 0.70, compared with 0.92 for individuals with neither condition. The EQ-5D was also shown to distinguish between groups by severity of depression.(21) Sobocki et al. reported mean EQ-5D scores of 0.60 for mild depression, 0.46 for moderate depression, and 0.27 for those with severe depression, as assessed by CGI-S.(22) Konig et al. compared the mean BAI scores for groups who indicated they had a problem on the EQ-5D anxiety/depression health dimension with those who had no problem. They found that most of the EQ-5D dimension response levels (especially for anxiety and depression) were associated with significant differences in scores of WHOQoL and measures of psychopathology such as BAI scores.(17) Mann 2009 showed that EQ-5D distinguished between depression severity groups, with mean scores of 0.645 for mild depression, 0.656 for moderate, 0.558 for moderate/severe, and 0.337 for severe.(7) Mycheski et al. also showed decreases in EQ-5D scores by severity group, this time for anxiety, with mean EQ-5D scores reducing with increasing anxiety severity (normal 0.83, mild 0.78, moderate 0.60, severe 0.30).(25) Petrou 2009 also showed that EQ-5D scores decreased in line with self-reported health status, although the SF-36 was found to be more efficient (described in Peasgood as "Relative efficiency statistic – how well can they detect differences in self reported (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.").(8)

Construct validity (convergent): Five studies tested the convergent validity of the EQ-5D compared to a variety of other measures (Ayedemir 2009, Gunther 2008, Konig 2010, Mann 2009, Sapin 2004). Correlations with symptom measures showed differences between studies in the strength of these relationships, ranging from not significant for the body sensation questionnaire (BSQ), agoraphobic cognitions questionnaire (ACQ), mobility inventory – avoidance alone (MIA), mobility inventory – avoidance accompanied (MIB) (17), to very strong (-0.77) on HAM-D,(23) and 0.7 on physical health WHO-QOL (17). Moderate correlations were found between EQ-5D and functional measures (r=0.49

with GAF, -0.58 with BRAMES)(5) For symptom measures, Konig 2010 showed significant correlations with BAI (r=-0.58) and BDI-II (r=-0.54).(17) A significant correlation was also reported between EQ-5D and other generic quality of health measures, with Sapin 2004 finding moderate to strong correlations between EQ-5D and SF-36 (0.49 at baseline, increasing to 0.63 at day 56). (15) Finally, Mann 2009 showed significant correlations between EQ-5D and PHQ-9, increasing from -0.451 at baseline to -0.638 at 3 month follow-up. (7)

Responsiveness (change over time): Sixteen studies assessed responsiveness through changes in EQ-5D scores over time (table A6). Significant responsiveness was found for studies of depression. Fernandez 2005 showed responsiveness to improvement after treatment from baseline to 8 week follow-up, with mean differences in EQ-5D scores for patients with severe MDD on escitalopram 0.52 to 0.78 ($p<0.001$), and those on venlaxafine 0.54 to 0.77 ($p<0.001$).(13) Gunther 2008 demonstrated responsiveness of the EQ-5D to deterioration over time. EQ-5D scores were found to deteriorate for those in the worst health (-0.290) more than they improved for those in better health (0.155).(5) Mann 2009 showed in an RCT of collaborative care for major depressive disorder that the EQ-5D was responsive over time, with scores increasing by 0.147 between baseline and follow-up at 3 months.(7) Sapin 2004 demonstrated responsiveness to improvement in anxiety/depression over time. EQ-5D scores showed improvement of 0.35 at 4 weeks and 0.45 at 8 weeks. After 8 weeks, 9.3% of individuals reported extreme problems with anxiety / depression, compared with 77.9% at baseline. EQ-5D was also able to distinguish responder-remitters, responder non-remitters and non-responders based on MADRS score.(15) Caruso 2010 also showed that EQ-5D was responsive over time for individuals with depression, with mean scores improving from 0.40 at baseline to 0.66 (0.73) at 3 (6) months.(12) Swan 2004 showed responsiveness of the EQ-5D in patients with depression following a Coping with Depression course, with scores increasing from 0.49 at baseline, to 0.65 (0.68) at week 12 (26). These improvements aligned with scores on the GSI and BDI.(9) Van Straten 2008 found significant responsiveness for EQ-5D in patients with depression taking part in a trial of a web-based self-help intervention. Using post-intervention improvement scores for those who completed the course as their outcome measure, the authors report effect sizes (Cohen's D) of EQ-5D (ES=0.31), compared to CES-D (ES=0.5), MID (ES=0.33) and SCL-A (ES=0.42). (20)

Bosmans et al. showed no significant difference in quality adjusted life years (QALYs) gained between antidepressant and usual care control groups (standardised mean difference (SMD) -0.00045), although nor was there a significant difference in improvement on the MADRS score (SMD -0.81).(18) Peveler et al. reported a numerical change from baseline for each of three treatment

groups for improvement in depression symptoms at 12 months (10). Peasgood et al. report that whilst mean scores for all groups showed improvement, this was non-significant because of high standard deviation (Peasgood). Serfaty et al. failed to find significant responsiveness of EQ-5D in patients with depression taking part in an RCT of CBT versus treatment as usual (TAU). Mean EQ-5D scores remained similar from baseline (0.50), at 4 months (0.53) and 10 months (0.54) for those in the CBT group, and baseline (0.52), 4 months (0.55) and 10 months (0.52) for those receiving TAU. The findings of Serfaty et al. are an exception to the general picture of responsiveness of the EQ-5D. In this study, the EQ-5D was less responsiveness than the BDI-II. A possible explanation is that EQ-5D may lack responsiveness for older patients, as the patient group in this study had a mean age of 74.1 years.(11) For individuals with anxiety, although Konig et al. 2009 found no significant difference for the EQ-5D between intervention and control groups in a trial of training versus treatment as usual (TAU), scores on the BDI and BAI did not detect any differences either.(16) However for the same sample, Konig et al. 2010 reported effect sizes for anxiety by severity group. Results showed effect sizes for the EQ-5D (ES=-0.99) for 'more anxiety', was twice as large as for the comparator measures (WHOQoL, BSQ ACQ), and the corresponding SRM (SRM=-0.54) was also more responsive than the other measures, suggesting the EQ-5D may overestimate improvements(17) Mychawski et al. also demonstrated the responsiveness of the EQ-5D between those in functional remission from anxiety and those not in remission (mean EQ-5D: 0.87 vs. 0.61 at 8 week follow-up).(25) Finally, Lamers et al. showed the EQ-5D was responsive over 1.5 years follow-up in a group of patients with anxiety, with EQ-5D scores increasing from 0.513 to 0.680, compared to the SF-6D, where scores increased over time from 0.577 to 0.701.(19)

Authors' conclusions: See Section 4.1.4 below.

4.1.3 Assessment of the review in relation to objectives of work package 1.1

Relevance of review question: The aim of Peasgood et al 2011(1) is convergent with the aims of WP1.1.

Assessment of review quality: Assessment of the quality of the review was conducted using a modified version of the AMSTAR tool (Shea et al 2007) and also by considering the strength and quantity of the evidence. The adequacy of the reported data in the context of work package 1.1 was also assessed. A summary of the quality assessment is shown in the Appendix.

Peasgood et al. 2011(1) scored well against most of the relevant AMSTAR criteria. Reference is made to a published protocol to evidence an a priori design, thus reducing the possibility of changes to the protocol in response to results. Quality assessment of the included studies was conducted and whilst no formal method for assessing the quality of this type of study has been previously validated, methods published elsewhere were followed.(2) Study selection was carried out by only one reviewer, and double data extraction or data-checking was not conducted, leaving the study at higher risk of errors. Inclusion/exclusion criteria are clearly defined.

Acceptability of the search: A comprehensive search of a wide range of sources was carried out including reference tracking.

Acceptability of study selection: Study selection criteria were clearly defined.

Adequacy of available data and synthesis: The review only provided a small amount of data relating to each study, however this was adequate for the purposes of WP 1.1.

4.1.4 Conclusion of appropriateness of EQ-5D for psychological therapies (anxiety and depression)

The authors concluded that while the evidence base supports the use of the EQ-5D in patients with anxiety and depression, there is evidence to suggest the EQ-5D may lack responsiveness in the elderly.(1) They also noted a stronger correlation with depression scales than anxiety scales in patients with anxiety which suggests the known group validity results may be driven by the presence of comorbid depression or the depression aspect within anxiety disorders. Comparing the results of the EQ-5D with the SF-6D, as has been found elsewhere,(26) the authors noted that the EQ-5D showed greater improvements than the SF-6D for those at the lower end of the HRQoL spectrum (e.g. severe depression) while the SF-6D appeared to be more sensitive to changes at the top end of the HRQoL spectrum (e.g. mild depression).

The evidence suggests the EQ-5D is appropriate in patients with depression, but additional research is required to confirm its appropriateness in patients with anxiety conditions.

Table 2: Summary of evidence on EQ-5D for patients receiving psychological therapies

Measure (N)	Acceptability	Reliability	Construct (KGV; Convergent)	Responsiveness (Change over time; Ceiling effects)
Adults				
EQ-5D (21)	Not reported	Not reported	Good; Good	Mixed; not reported
	Authors note that EQ-5D correlations were higher when compared against depression than anxiety scales in patients with anxiety (study n=1), that there may be a lack of responsiveness in older adults (study n=1) and that the EQ-5D showed greater changes at the lower end of the HRQoL spectrum. The EQ-5D is appropriate in patients with depression, but additional research is needed to confirm its appropriateness in patients with anxiety.			

EQ-5D: EuroQol 5 dimensions; HRQoL: health related quality of life.

4.2 Alternative measures in psychological therapies (WP1.2)

Evidence from WP1.1 for psychological therapies suggests that the EQ-5D is appropriate for use in depression, though there may be a lack of responsiveness in older adults, and additional research is needed to confirm its appropriateness in patients with anxiety. Both the latter conclusions were based on single studies.(11;17)

To investigate other measures that may be appropriate for use in psychological therapies, searches for WP1.2 were conducted and six reports of potential relevance were identified. Characteristics and recommendations of these reports are given in the Appendix. One was a report to the DH from the Oxford PROMS group,(2) and had similar aims to WP1.2. Two reports were from the Royal College of Psychiatrists with the aim of providing recommendations to clinicians. Both covered several mental health topics; one concentrated on adults and one on older adults. The report for older adults was intended to aid the improvement of care and the assessment of individual patient outcomes. The three remaining reports were research recommendations from the EMA, and were intended as recommendations for clinical research into interventions for anxiety in consideration of a submission to the EMA for licensing in Europe. Each covered a different anxiety condition, namely: generalised anxiety disorder (GAD); panic disorder; and social anxiety disorder. These latter three reports were deemed too specific in their topics to be useful in an audit context where a wide range of anxiety disorders would be encountered, and were not considered further.

Only the Oxford PROMS group review(2) had the same aim as WP1.2, with most others recommending measures for use in clinical practice or clinical research. With the exception of the Oxford review, none reported robust methods of assessment which included evaluation of psychometric evidence; the majority used an expert panel or working group to conclude which if any

measure was suitable in the specific conditions. None of the reviews aimed to provide a comprehensive list of validated measures.

The most relevant report for WP1.2 came from the Oxford PROMS group.(2) This report did not, however, come up with one single recommendation and did not consider all measures available. It concluded that "Compared to the debates about screening and use in the context of individual patient care, the debate about PROMs and quality in mental health services is still in its earliest stages."(2) Of most relevance to WP1.2 were the report's considerations about choosing a measure that covers both anxiety and depression, for which Clinical Outcomes in Routine Evaluation — Outcome Measure (CORE-OM) was preferred over HADS, as it had a better level of evidence to support it and included social function. The report stated that if a preference measure is required, EQ-5D is preferred over the SF-6D.

For adult populations, the Royal College of Psychiatrists give as examples the patient health questionnaire (PHQ-9), a 9-item depression scale intended to diagnose and monitor depression; the Generalised Anxiety Disorder Assessment -7 (GAD-7), a seven-item questionnaire used as a screening tool and severity measure for generalised anxiety disorder; CORE-OM, used before and after therapy; and also recommends both the EQ-5D and the SF-6D. In older adults, the Royal College recommends GAD-7 and the hospital anxiety and depression scale (HADS), but notes that the latter may miss somatic symptoms.

It is also worth noting that a new measure is in development that is intended to be suitable for use across the spectrum of psychotic and non-psychotic mental health conditions.(Department of Health 2013) The measure, Recovering Quality of Life (ReQoL) is currently under development by the Policy Research Unit in Economic Evaluation of Health and Care Interventions and is due to be available around July 2015.(27) Once the measure is available and has been validated in people with depression and anxiety, the ReQoL may become a candidate measure for inclusion in the NCA.

4.3 Evidence for economic evaluations in psychological therapies (WP1.3)

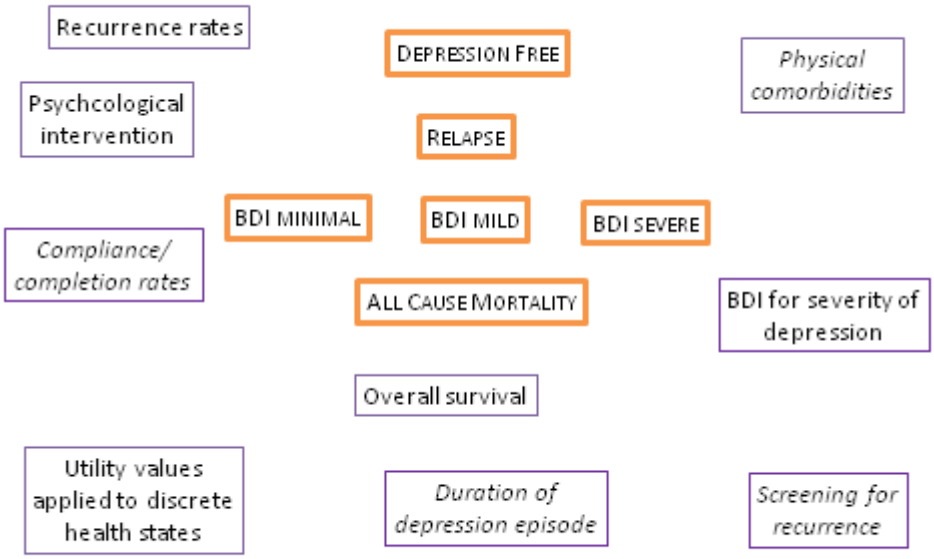
4.3.1 Cost-effectiveness modelling approach used in recent HTAs in psychological therapies

Just one multiple technology appraisal (MTA) in psychological therapies for depression and anxiety (published in 2002) was identified from the searches.(28) This was superseded by a later TA,(29) which was subsequently withdrawn, leaving no valid TAs for this condition. As a consequence the

most recent CG in anxiety and depression was identified.(30) The CG encompassed a broad decision space covering pharmacological and physical interventions, services (organisation of care, development of staff roles, introduction of mental health specialists into primary care), and psychological and psychosocial interventions. The following text reviews the economic evidence relating to psychological therapies only (i.e. Sections 6 to 8 of the CG). The guideline team identified two studies describing economic evaluations for (low-intensity psychological interventions),(28;31) and two studies describing economic evaluations for (high-intensity psychological interventions).(32;33)

The low-intensity studies both evaluated CCBT software packages (Beating the Blues (31;34), Overcoming Depression and Cope(34)) compared to standard care in patients with depression in the UK (Table 3). McCrone conducted an economic evaluation alongside a clinical trial. The main results were reported in terms of cost per point reduction in Becks Depression Index (BDI).(35;36) Although they also reported a cost-utility evaluation, the methodology used to weight the survival for the QALYs was not considered to be robust.(34) Kaltenthaler *et al.* used a decision tree to compare the interventions under evaluation. The clinical pathway was described using discrete health states based on severity of depression (Figure 1) using well-established cut-offs relating to the BDI: minimal (≤9), mild (10-18), moderate (19-29), and severe (30-63). The intervention specific severity evidence (i.e. the proportion in each severity category whilst on treatment) was sourced from clinical trials and the rate of relapse (assumed equivalent for both interventions) was sourced from the literature. Mean EQ-5D scores were assigned to the discrete health states within the model. The analysts reported the relationship between EQ-5D and severity of depression (measured using the BDI mapped onto the CORE-OM) was non-linear and rather than using a statistical model to predict changes in EQ-5D in the economic model, mean EQ-5D scores were estimated for the individual discrete health states.

Figure 1: Modelling approach used in psychological therapy HTAs



Legend: Orange framed boxes with uppercase text describe the health states used in the diabetes TA models while the purple framed boxes with lower case (plain) text describe the evidence used. Italised text indicative of additional variables which would be informative for future economic evaluations in psychological therapies.

One of the two high-intensity studies compared the cost-effectiveness of mindfulness-based cognitive therapy (MBCT) compared to maintenance antidepressant medication in people with depression.(32) The second compared cognitive therapy plus antidepressants and clinical management with antidepressants and clinical management in people with partially remitted major depression.(33) The primary outcome measure in both studies was cost per relapse or recurrence avoided. Neither study extrapolated beyond the duration of the studies used for effectiveness (recurrence/relapse), and neither reported results in terms of cost per QALY.

Table 3: Summary of existing models used in psychological therapy TAs

Model method, clinical effect	Method used to model utilities
CG (CG90): Depression in adults, the treatment and management of depression in adults; 2010(30)	
Did not construct a new evaluation for psychological therapies but identified 2 publications (below) describing economic evaluations in this area.	
McCrone, 2003(31)	
Economic evaluation alongside an RCT, generating the cost per point reduction in BDI, cost per symptom free day, cost per QALY Patient outcomes measured using: Beck Depression Inventory (BDI), Beck Anxiety Inventory (BAI), Work and Social Adjustment Scale (WSA) Effectiveness: depression free days per intervention Source: RCT evidence	Utility: measure not reported; applied to days with and without symptoms of depression Source: published literature synthesis a variety of different measures AEs: not applicable
Kaltenthaler, 2008(34)	
Decision tree model Discrete health states: minimal, mild, moderate, severe depression , relapse vs. non-relapse Effectiveness: rates for depression severity Source: RCTs used for clinical effect	Utility: EQ-5D; mean values assigned to discrete HS based on a relationship between the BDI and the CORE-OM Source: published literature showing a relationship between clinical severity (CORE-OM) and EQ-5D scores AEs: not applicable

HS: health states; AE: Adverse Events; CG: Clinical Guideline; RCT: randomised controlled trial

In summary, the following evidence would be required to compare providers or the cost-effectiveness of interventions for psychological therapies:

- Screening for anxiety / depression
- Information on physical comorbidities
- Intervention
- Compliance to intervention/completion rates
- Condition specific PROM (such as BDI, BAI)
- Severity measure (using validated measure for mild, moderate or severe)
- Recurrence/relapse rates (with dates)
- Utility values

The majority of this evidence would need to be dated and linked through timings of collection. Although the models reviewed applied utility values to discrete health states, with more detailed information from a large dataset, the association between depression and HRQoL could potentially

be measured as a continuous relationship using a similar approach to that used in the diabetic vision model (see Appendix E).

4.3.2 Fields collected in the psychological therapy NCA

The National Audit of Psychological Therapies for Anxiety and Depression (NAPT) aims to promote access, appropriateness, acceptability and positive outcomes of treatment for people who have anxiety or depression. The audit is open to all NHS-funded services providing psychological therapies in the community in England and Wales for people with anxiety or depression. The data give a snapshot of information for a designated recall period (e.g. between 1st September and 30th November 2010) within the audit timescales (Appendix). The Therapist questionnaire covers areas directly relating to the therapist's training and qualifications and the type and hours of service provided. The therapist completed retrospective case record questionnaire provides data on individual patients (gender, age, date of referral, type of treatment offered and whether taken up/ completed, and several outcome measures for example the HADS). However, the majority of fields in these questionnaires are optional. There appears to be an 'outcome measure' (within the Retrospective audit section, Appendix), but no detail is provided on what measure is used. The service user's questionnaire (Talking treatment Survey) is optional and includes information on the type and volume of current treatment, and the patient's personal experience of the treatment. The questionnaire includes a five item measure (the ARM-5) derived from the full 28 item Agnew Relationship Measure (ARM), which was designed to assess the strength of the therapeutic relationship between clients and their therapists.(37)

4.3.3 Comparing fields in psychological therapy NCA with variables used in existing HTAs

The existing models used health states categorised by severity of condition and intervention specific relapse rates were used to compare individual therapies. Responses to the type of psychological therapy provided, sub-categorised by high (e.g. cognitive analytic therapy) or low (e.g. psycho education) intensity therapy, the number of sessions attended and completion of therapy, are mandatory in the retrospective service user questionnaire. These could potentially be used to model adherence and withdrawal rates. It is not clear if there are currently any fields which could be used to model relapse, which is a frequent occurrence in this chronic condition and a key parameter for any economic model in this area.

The retrospective service user questionnaire also includes information on initial and final outcome scores such as HADs, CORE-10, and BAI which would be useful to measure severity. While in theory

these could be used to identify response to treatment, this would depend on the timings of the data collection. There is evidence in the literature which could be used to link some of these variables to preference-based utility values (e.g. HADs to EQ-5D). However, the functions currently available have not been validated on external data and there are issues related to sample sizes and representativeness of the clinical severity in the populations used to obtain the functions. Consequently it is recommended that the service user questionnaire also includes a measure which could be used to generate utilities.

4.4 Recommendations for psychological therapies

The NCA collects information from patients with anxiety or depression receiving psychological therapies in the community, and there is a mandatory field (Part D, ICD-10 diagnosis) which could be used to differentiate between the subcomponents of this condition. The Service user questionnaire includes a measure to capture the strength of the therapeutic relationship between clients and their therapists,(37) and the retrospective service user questionnaire also includes information on measures such as the HADs, CORE-10 and BAI. However, there is no measure which could be used to generate preference-based scores directly. Potential recommendations (PR) and areas for future research (FR) are discussed below. All suggested future research areas are indicative and would require a discussion and detailed proposal if required.

It is recommended that the EQ-5D is collected in the Service user questionnaire alongside clinical measures, and that the ReQOL is considered once this becomes available (PR.1, PR.2). Due to the uncertain evidence in patients with anxiety, and the limited evidence in patients with depression, it is recommended that additional research is conducted to assess the appropriateness of the measure in patients receiving psychological therapies using data collected in the NCA (FR.1).

It is also recommended that therapists use a common set of condition specific measures to capture the severity of the condition and response to treatment (for example the PHQ-9 and HAD-7). To facilitate links and comparisons with other sources of data, the measures should be synchronised to match measures adopted by the DH for use in the NHS Outcomes Framework (PR.3).

The psychological therapies NCA data is currently being analysed under a separate research project (WP3), and the results of this research will inform additional recommendations for the fields in the audit (FR.2).

Table 4: Recommendations and associated future research for psychological therapies

PR.1	*Collect the EQ-5D in the service user questionnaire alongside clinical measures such as the PHQ-9 and GAD-7*
PR.2	*Collect the ReQOL in the service user questionnaire once it becomes available*
FR.1	*Assess the appropriateness of the EQ-5D and the ReQoL in patients receiving psychological therapies using the data from the NCA*
PR.3	*All therapists use a common set of measures (to be decided and ultimately synchronised with the measure adopted for use in NHS Outcome Framework)*
PR.4	*Include additional mandatory fields in the psychological therapies NCA*
FR.2	*Detailed analyses of fields collected in the current NCA is being undertaken under a separate research project within this programme of work (WP.3).*

5. SUMMARY

5.1 Summary of evidence used to inform the conclusions for WP1.1 and WP1.2

An existing review provided evidence from 21 primary studies relating to the EQ-5D in psychological therapies (Table 5). Construct validity (both known group and convergent) was reported to be good, though data from one study showed that the EQ-5D correlated better with depression-specific measures and subscales than with anxiety-specific ones in people with anxiety. This suggests that additional research is required to confirm the appropriateness of the EQ-5D in patients with anxiety. Responsiveness was more mixed, but generally good, though notably one study in the elderly showed poor responsiveness. Better responsiveness of the EQ-5D was observed at the lower end of the HRQoL spectrum (e.g. severe depression) when compared to the SF-6D, which was more sensitive to changes at the top end of the spectrum. Overall, the EQ-5D was considered appropriate for use in anxiety and depression, though further validation work is required in anxiety. Searches were conducted to identify other measures. In keeping with The Royal College of Psychiatry, the GAD-7 and PHQ-9 measures are recommended for use alongside the EQ-5D. ReQoL, a measure currently in development by EEPRU for use in psychotic and non-psychotic mental health conditions is due to be available in July 2015, and could be considered for use once available.

Table 5: Summary of evidence currently available for recommended measure(s)

Measure	N	Acceptability	Reliability	Construct		Responsiveness		Overall
				KGV	Convergent	Change over time	Ceiling Effect	
EQ-5D	21	NR	NR	Good	Good	Mixed	NR	Acceptable
PHQ-9		Recommended by the Royal College of Psychiatrists						
GAD-7		Recommended by the Royal College of Psychiatrists						
ReQOL		This measure is currently in development and will be available in 2015						

N= number of studies used to inform conclusions, KGV: known group validity; NR, the existing review did not review this psychometric property

5.2 Summary of evidence required for use in economic evaluations (WP1.3)

Although the audit for patients receiving psychological therapies does not collect PROMs, there is a service user questionnaire which could potentially be amended to include a PROM. Two measures (the Beck's depression index, and the Beck's anxiety index) are also collected in the audit and it is possible that this evidence could be used to predict preference-based utility data using existing published relationships. There also appears to be an 'outcome measure' within the retrospective audit but it is unclear what this measure is hence it is not possible to determine its usefulness. Relapse rates are high for this condition and compliance to therapy can be problematic. Together with severity of the condition, these are key variables within economic evaluations but it is not clear

if there are currently any mandatory fields within the audit relating to these. This audit is currently being used as a case-study in an associated project (WP3), and the results of this project will provide an indication of what can be achieved with the data collected.

APPENDIX: PSYCHOLOGICAL THERAPIES

The tables in this Appendix provide additional information for the reviews (WP1.1, 1.2 and 1.3) conducted for psychological therapies.

Table A1: Quality assessment for Peasgood et al (Psychological therapies)(1)

Quality assessment criteria	**Compliance with criteria**
AMSTAR	
Was an a priori design provided?	Yes
Was there duplicate study selection and data extraction?	No
Were the methods used to combine the findings of the studies appropriate?	Yes, narrative synthesis due to heterogeneity.
Was the scientific quality of the included studies assessed and documented?	Yes, using method described in Fitzsimmons et al., only for items relating to utility measures
Was the scientific quality of the included studies used appropriately in formulating conclusions?	Yes
Overall judgement of quality of review	Good but only 1 reviewer.
Quality of the searches	**Acceptable**
Strength of the evidence	
Were the conclusions robust and conclusive?	Yes for depression, mixed for anxiety
Quantity of the evidence	
Was there enough data to be confident that any additional data published subsequently would be very unlikely to change the conclusions drawn?	Yes
Adequacy of data reported	
Did the review provide sufficient data to allow integration of an update/assessment of the methods used?	Yes
Did the review assess EQ-5D in a way compatible with the aims of work package 1.1?	Yes, construct validity (known groups or convergent) or responsiveness (effect sizes, standardised response means, or correlation with change scores on symptom measures).

Table A2: Characteristics of primary studies included in Peasgood review for psychological therapies. Adapted from Peasgood et al.(1)

Author, year	Study design	Condition	Study information	Male/female	Mean age at baseline
Aydemir et al, 2009, Turkey(23)	RCT – no further information	Major depressive episode according to DSM-IV criteria	N=74	36.5%/63.5%	Mean age 39.6 years
Bosmans et al., 2008, The Netherlands(18)	RCT: 2 x intervention groups: Usual care no AD Usual care plus AD	Major or mild-major depression in primary care	N=89	27%/73%	Mean age 48
Caruso et al., 2010, Multinational(12)	Cross-sectional FINDER study	Clinically diagnosed episode of depression requiring pharmacological treatment	N=513	16.1%/72.9%	Mean age 49.2 years
Ergun, (no year), Turkey(24)	RCT – no further information	Major depressive disorder	N=74	N/R but is same study as Aydemir 2009	N/R but is same study as Aydemir 2009
Fernandez et al., 2005, multinational(13)	RCT of Escitalopram vs venlafaxine	DSM-IV criteria for severe major depressive disorder	N=293	Escitalopram 24.6%/75.4% Venlafaxine 28.8%/71.2%	Escitalopram mean age 48.4 years Venlafaxine mean age 46.5 years
Gunther et al., 2008, Germany(5)	N/R	Patients with a depressive episode according to ICD-10 classification	N/R	N/R	N/R
Konig et al., 2009, Germany(16)	Controlled trial. Intervention group = training (n=23 GP practices) vs control group = usual care (n=23 GP practices)	Patients with anxiety disorder	N=389	N/R	N/R
Konig et al., 2010, Germany(17)	N/R	Anxiety disorder	N=389	N/R	N/R
Lamers, 2006, Netherlands(19)	N/R	Diagnosis of major depressive disorder, dysthymic disorder, panic disorder, social phobia	N/R	N/R	N/R
Mann et al., 2009, UK(7)	RCT on collaborative care	Depression (MDD according to SCID)	N=114	23%/77%	Mean age 42.5 years
Mychaskiw et al., 2008, USA(25)	Controlled trial – treatment with pregablin, venlafaxine-XR or	Non-depressed patients with Generalised Anxiety Disorder.	N=374	N/R	N/R

Author, year	Study design	Condition	Study information	Male/female	Mean age at baseline
	placebo				
Petrou et al., 2009, UK(8)	RCT. Usual care vs community postnatal support visits	Postnatal women	N=623 (complete data for 493)	100% female	Aged 17 and over
Peveler et al., 2005, UK(10)	RCT. Patients receive TCA, SSRI or lofepramine	Patients with a new depressive episode	N=327	32.7%/67.3%	Mean age 42.5 years.
Reed, 2009, multinational(14)	N/R	Patients with clinical depression	N=3468 at baseline N=2854 data at both 3 and 6 month follow up	N/R	N/R
Saarni, 2007, Finland(6)	Population survey	Assessment of 12 month prevalence of depressive anxiety or alcohol disorders (DSM-IV)	N=5219	N/R	N/R
Sapin, 2004, France(15)	Population survey	New episode of major depressive disorder (MDD) according to DSM-IV	N/R	N/R	N/R
Serfaty et al., 2009, UK(11)	RCT, CBT vs TAU	Older people (aged 65 and over) with depression screened by 15-item geriatric depression scale or BDI-II score 14 or more	N=204	21.6%/79.4%	Mean age 74.1 years
Sobocki et al., 2007, Sweden(22)	Population survey	Diagnosis of depression	N=447, baseline data n=394	23%/67%	Mean age 47 years
Van Straten et al., 2008, Netherlands(20)	Controlled trial – web-based self-help	Depression, anxiety or work-related stress	N=213	N/R	N/R
Supina et al., 2007, Canada(21)	Population survey	Random population sample to identify Major Depressive Episode or Anxiety disorders	N=5410 sample size N=5,383 successful data	39.8%/61.2%	Mean age 40.8 years
Swan et al., 2004, UK(9)	Cross-sectional – patients attending Coping with Depression course	Primary diagnosis of chronic or recurrent depressive disorder; current depressive episode of at least moderate severity	N=76 entrants, 31 completed intervention	N/R	N/R

AD: antidepressant; RCT: randomised controlled trial; N/R: not reported; DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, 4th Edition; MDD: major depressive disorder; BDI: Beck depression inventory; CBT: cognitive behavioural therapy; TAU: treatment as usual; TCA: tricyclic antidepressant; SSRI: selective serotonin reuptake inhibitor; SCID: structured clinical interview for DSM-IV axis 1 disorders

Table A3: Method of assessing validity and responsiveness for individual studies in Peasgood review for psychological therapies. Adapted from Peasgood et al.(1)

Author, Year, Location	EQ-5D	Comparison measure	Psychometric properties assessed	Assessment of psychometric properties
Aydemir et al, 2009, Turkey(23)	EQ-5D UK, EQ-VAS	HAM-D, SF-36	Known groups validity, convergent validity	Correlation between EQ-5D and HAM-D EQ-5D for single/recurrent depression groups
Bosmans et al., 2008, The Netherlands(18)	EQ-5D UK	MADRS	Responsiveness	Mean difference in QALYs gained between the two intervention groups
Caruso et al., 2010, Multinational(12)	EQ-5D	HADS-D, HADS-A, SSI-28, VAS pain	Responsiveness	Regression analysis to explore predictors of EQ-5D
Ergun, 2007 Turkey (linked paper to Aydemir 2009)(24)	EQ-5D UK	HAM-D	Convergent validity, responsiveness	Correlation with HAM-D, Change in mean from baseline to 6 weeks follow-up
Fernandez et al., 2005, multinational(13)	EQ-5D UK	QLDS, MADR~~A~~S	Responsiveness	Change in mean from baseline to week 8 by treatment group
Gunther et al., 2008, Germany(5)	EQ-5D UK and German (based on TTO), EQ-5D VAS	WHOQoL-BREF, CGI-S, GAF, BRAMES	Convergent validity, responsiveness	Spearman rank correlations between EQ-5D and all other measures Change in mean for health severity groups, t statistic, ES & SRM
Konig et al., 2009, Germany(16)	EQ-5D UK, EQ-VAS	PHQ-D, BAI, BDI-II	Responsiveness	Differences between intervention and control group
Konig et al., 2010, Germany(17)	EQ-5D UK, EQ-VAS	WHO-QoL-BREF, BSQ, ACQ, BAI, BDI-II, MI (MIA and MIB)	Known group validity, convergent validity, responsiveness	Effect sizes between those with and without anxiety Correlations between EQ-5D and other measures Effect sizes, SRM for anxiety severity
Lamers, 2006, Netherlands(19)	EQ-5D	SF-36	Known group validity, responsiveness	Mean differences for anxiety severity groups Mean improvement in EQ-5D scores at 1.5 year follow-up, by severity groups
Mann et al., 2009, UK(7)	EQ-5D	SCID, PHQ-9, SF-36	Known group validity, convergent validity, responsiveness	Correlations between EQ-5D and other measures EQ-5D scores for depression severity groups

Author, Year, Location	EQ-5D	Comparison measure	Psychometric properties assessed	Assessment of psychometric properties
				Change from baseline scores for EQ-5D and SF-36 Remission rates and MDD rates at follow-up for EQ-5D, SF-36
Mychaskiw et al., 2008, USA(25)	EQ-5D UK	HAM-A	Known group validity, responsiveness	EQ-5D scores by anxiety severity group Functional remission at 8 weeks (SDS<5) Symptomatic remission (HAM-A score 7) at 8 weeks
Petrou et al., 2009, UK(8)	EQ-5D	SF-6D, EPDS, self-rated health status	Known group validity	Change in scores by SR health status group. Relative efficiency statistic: how well do 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
Peveler et al., 2005, UK(10)	EQ-5D	HAD-D, CIS-R, PROQSY, SF-36	Responsiveness	Improvement from baseline by intervention groups
Reed, 2009, multinational(14)	EQ-5D, EQ-VAS	SF36, HADS-D, HADS-A, SSI-28-item, Pain VAS	Known group validity	Regression analysis
Saarni, 2007, Finland(6)	EQ-5D UK, 15D measure with Finnish valuations.	M-CICI	Known group validity	Comparison of individuals with and without psychiatric diagnosis
Sapin, 2004, France(15)	EQ-5D	SF-36, QLDS Clinician/physician reported MADRS, CGI-S	Known group validity, convergent validity, responsiveness	Differences in EQ-5D by disease severity level/response group. Correlation between EQ-5D and other measures. Ability to detect differences in mean EQ-5D (and Anxiety/Depression health dimension responses) over time, and in responders vs. non responders
Serfaty et al., 2009, UK(11)	EQ-5D (no reference to scoring system)	BDI-II, SFQ	Responsiveness	Ability to detect differences in mean EQ-5D over time compared to changes in BDI-II
Sobocki et al., 2007, Sweden(22)	EQ-5D UK	CGI-S (1-7) (severity), CGI-I (Improv.)	Known group validity Responsiveness	Differences in EQ-5D scores by disease severity as assessed by CGI-S. EQ-5D score up to 6 months follow-up, and by severity
Van Straten et al., 2008, Netherlands(20)	EQ-5D	CES-D, MDI, HADS, SCL-A, MBI, work-related stress – 3 subscales	Responsiveness	EQ-5D for intervention groups pre and post intervention Effect size (Cohen's D) for all scales
Supina et al., 2007, Canada(21)	EQ-5D, EQ-VAS	MINI	Known group validity	EQ-5D scores for anxiety and depression groups

Author, Year, Location	EQ-5D	Comparison measure	Psychometric properties assessed	Assessment of psychometric properties
Swan et al., 2004, UK(9)	EQ-5D (no reference to scoring system)	BDI-II, BSI which generates the GSI	Responsiveness	EQ-5D at baseline, week 12 and week 26. Improvement in BDI and GSI (baseline to week 12, baseline to week 26).

EQ-5D: EuroQoL 5 dimensions; HAM-D: Hamilton depression scale; SF36: short-form 36; MADRS: Montgomery-Asberg depression rating scale; HADS-A: hospital anxiety and depression scale - anxiety; HADS-D: hospital anxiety and depression scale - depression; QALYs: quality of life years; SSI-28: somatic symptom inventory; QLDS: quality of life in depression scale; WHOQoL-BREF: WHO quality of life - brief; CGI-S: clinical global impression - severity; GAF: global assessment of functioning; TTO: time trade-off; BRAMES: Bech-Rafaelsen melancholia scale; ES: effect size; SRM: standardised response mean; BSQ: body shape questionnaire; ACQ: agoraphobic cognitions questionnaire; BAI: Beck anxiety inventory; BDI: Beck depression inventory; MI (A&B): mobility inventory, (avoidance alone and accompanied); SCID: structured clinical interview for DSM-IV axis 1 disorders; PHQ-9: patient health questionnaire; SDS: self-rating depression scale; EPDS: Edinburgh postnatal depression scale; CIS-R: clinical interview schedule - revised; PROQSY: a computerised version of the CIS; M-CICI: chronicity coping and impact instrument; SFQ: social functioning questionnaire; CES-D: Center for epidemiological studies - depression; MDI: major depression inventory; SCL-A: symptom checklist - anxiety; MBI: Maslach burnout inventory; MINI: mini international neuropsychiatric interview; BSI:brief symptom inventory; GSI: global severity index

Table A4: Convergent validity results for psychological therapies, adapted from Peasgood et al (1)

Author, year	Method of measuring convergence (e.g. Spearman rank correlation, statistical significance)	Convergent validity results
Aydemir 2009(23)	Correlation between EQ-5D and HAM-D	HAM-D correlated with EQ-5D at r=-0.77
Gunther 2008(5)	Spearman rank correlation between EQ-5D and: WHOQoL-BREF, CGI-S, GAF, BRAMES	Significant correlations between EQ-5D and all comparator measures: BRAMES: -0.576 WHO-BREF: 0.545 CGI: -0.539 GAF: 0.492
Konig 2010(17)	Correlations between EQ-5D and WHO-QoL-BREF, BSQ, ACQ, BAI, BDI-II, MI (MIA and MIB)	At baseline EQ-5D correlations with other measures: Physical health WHO-QoL: 0.7 Mental health WHO-QoL: 0.5 Overall WHO-QoL: 0.58 BAI: -0.58 BDI-II: -0.54 BSQ, ACQ, MIA, MIB: all 0.4 and below.
Mann 2009(7)	Correlations between EQ-5D and PHQ-9	EQ-5D correlations with PHQ-9: -0.451 at baseline, -0.638 at 3 month follow-up
Sapin 2004(15)	Correlations between EQ-5D and SF-36, QLDS Clinician/physician reported MADRS, CGI-S	EQ-5D correlations with SF-36 MHC: 0.49 baseline, 0.56 at day 28, 0.63 at day 56

EQ-5D: Euro-QoL 5 dimensions; HAM-D: Hamilton depression rating scale; WHOQoL-BREF: WHO quality of life – brief; CGI-S: clinical global impression scale – severity; GAF: global assessment of functioning; BRAMES: Bech-Rafaelsen melancholia scale; BSQ: body shape questionnaire; ACQ: agoraphobic cognitions questionnaire; BAI: Beck anxiety inventory; BDI: Beck depression inventory; MI (A&B): mobility inventory (avoidance alone & accompanied); PHQ-9: patient health questionnaire; SF-36: short-form 36; QLDS: quality of life in depression scale; MADRS: Montgomery-Asberg depression rating scale

Table A5: Known groups results for psychological therapies, adapted from Peasgood et al (1)

Author, year	Method of measuring known groups validity	Known groups validity results
Aydemir 2009(23)	Mean EQ-5D for sub-groups with single vs. recurrent occurrence of depression episodes	EQ-5D: Single: 0.45 (SD 0.29) Recurrent: 0.41 (SD 0.31) (no significant difference, no p-value).
Konig 2010(17)	Compared the mean BAI scores for groups who indicated they had a problem on the EQ5D anxiety/depression health dimension with those who had no problem	Most of the EQ-5D dimension response levels (especially anxiety and depression) were associated with significant differences in BAI scores
Lamers 2006(19)	Comparing SCL subgroups (no detail on categories reported, presumed split by severity of symptoms of anxiety, comparing mean EQ-5D for those with more severe anxiety symptoms with those with less severe)	Mean EQ-5D scores showed expected pattern, with a large drop in utility in the most severe quartile of the SCL-A list (compared to less severe). The standardised difference was smaller in EQ-5D than observed in SF-6D
Mann 2009(7)	EQ-5D scores for depression severity groups	EQ-5D: Mild 0.645 (SD 0.23) Moderate 0.656 (SD 0.21) Moderate severe 0.558 (SD 0.27) Severe 0.337 (SD 0.29)
Mychaski 2008(25)	EQ-5D scores for anxiety severity groups	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
Petrou 2009(8)	Relative efficiency statistic – EQ-5D vs SF-6D/EPDS	Both EQ-5D and SF-6D show monotonically decreasing scores in line with SR health status. SF-6D found to be more efficient by 29% to 423.6%. Also more efficient using EPDS profiles (between 129.8% and 161.7%).
Reed 2009(14)	Regression analysis	Regression analysis found EQ-5D score had significant negative relationship with clinical characteristics (number of previous depressive episodes; and duration of current episode). Also negatively related to somatic symptoms and VAS pain.
Saarni 2007(6)	Compared mean EQ-5D for sub-groups categorised by conditions	Unadjusted mean EQ-5D scores were: population (0.83); any psychiatric diagnosis (0.72). Controlling for socio-economic status, somatic comorbidity and psychiatric comorbidity: Depressive disorders reduced 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
Sobocki 2007(22)	Differences in EQ-5D by disease severity as assessed by CGI-S	Significant differences in EQ-5D by disease severity groups 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)
Supina 2007(21)	EQ-5D for mental health diagnosis group	EQ-5D: Anxiety only (n=601) 0.84 (0.83-0.85) MDE only (n=140) 0.83 (0.81-0.85) Anxiety and MDE (n=280) 0.70 (0.69-0.72) Neither 0.92 (n=4338) (0.91-0.92)

EQ-5D: Euro-QoL 5 dimensions; SD: standard deviation; BAI: Beck anxiety inventory; SCL: symptom checklist; SF-6D: short-form 6 dimensions; HADS: hospital anxiety scale; EPDS: Edinburgh postnatal depression scale; CI: confidence interval; GAD: generalised anxiety disorder; MDD: major depressive disorder; NS: not significant; CGI-S: clinical global impression scale – severity; MDE: major depressive episode

Table A6: Responsiveness results for psychological therapies, adapted from Peasgood et al.(1)

Author, year	Method of measuring responsiveness	Responsiveness results
Bosmans 2008(18)	Mean difference in QALYs gained between the two groups (usual care no antidepressants vs usual care plus anti-depressants)	Mean difference in QALYs gained between the two groups – 0.00045 (95%CI -0.093; 0.084) (not significant) Difference in improvement in MADRS score -0.81 (95% CI -5.6; 4.0) (not significant)
Ergun(24)	Change in mean from baseline to 6 weeks follow-up	EQ-5D increase from mean 0.44 to 0.91 at 6 weeks follow-up
Caruso 2010(12)	Mean differences for intervention group over time	EQ-5D: Baseline 0.40 (SD 0.01) 3 month 0.66 (SD 0.26) 6 month 0.73 (SD 0.23)
Fernandez 2005(13)	Mean differences for intervention groups over time	EQ-5D: Baseline to week 8 Escitalopram arm 0.52 to 0.78 (p<0.001) Venlaxafine arm 0.54 to 0.77 (p<0.001)
Gunther 2008(5)	Change in mean for health severity groups, t statistic, ES & SRM	3 groups based on those who think health is worse, same or better than at baseline. Also 3 group based on BRAMES score. EQ-5D (UK) show deterioration for those in worst health (-0.290) larger than the improvement for those in better health (0.155) For t statistic, ES & SRM: EQ-5D t stat, ES and SRM find greater responsiveness to deteriorating health (almost twice as large as clinical measures). 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).
Konig 2009(16)	Mean differences between intervention (training) and control (usual care) group.	No significant differences between control and intervention group. BAI, BDI also showed no significant differences.
Konig 2010(17)	Effect sizes, SRM for all measures by anxiety severity group.	EQ-5D, EQ-VAS, WHO-QoL, BSQ, ACQ all show significant differences between more anxiety, constant and same (t stat). Effect sizes for EQ-5D: -0.99* for more anxiety, 0.39 for less anxiety (*more than twice that for other measures). SRM =-0.54* for more anxiety, 0.46 for less anxiety (-*EQ-5D higher than other measures: BSQ -0.72, WHO-QoL 0.35).
Lamers 2006(19)	Mean differences for anxiety severity groups	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 two subgroups with highest severity.
Mann 2009(7)	Remission rates and MDD rates at follow-up for EQ-5D, SF-36 Change from baseline scores for EQ-5D and SF-36.	Change from baseline EQ-5D increase 0.147 (change in median scores 0.069) (significant, no p value reported) 62% assessed as in remission at follow-up according to SCID. Remission at follow-up: mean EQ-5D= 0.759 (SD 0.25), mean SF-6D= 0.707 (SD 0.12) MDD at follow-up: mean EQ-5D=0.506 (0.37), SF-6D 0.550. Mean improvement between baseline of overall study population and follow-up utility for those with remission was EQ-5D=0.243, SF-6D= 0.140. EQ-5D showed larger health gains at follow up for all patients, and for those in remission.
Mychaski 2008(25)	Functional remission at 8 weeks (SDS<5) Symptomatic remission (HAM-A score 7) at 8 weeks. Note, this is as categorised (responsiveness) by Peasgood et al.[Peasgood 2012] but could be considered as known group validity.	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: HAM-A score 7: remission EQ-5D=0.84, no remission EQ-5D=0.60 HAM-A score 10: remission EQ-5D=0.83, no remission EQ-5D=0.57
Peveler 2005(10)	Improvement from baseline by intervention groups.	EQ-5D of 3 intervention groups showed improvement of about 0.22 points, most of which occurred in the 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 0.194) No significant differences between groups.
Sapin 2004(15)	Improvement in EQ-5D score at 4 and 8 week follow-up	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) Percentage with extreme difficulties on anxiety & depression was 77.9% at baseline and 9.3% at day 56.
Serfaty 2009(11)	Improvement in EQ-5D score and BDI-II score from Baseline to 4 and 10 week follow-up.	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 Taking Control intervention 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

		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 CBT and TAU for BDI-II results CBT BDI-II: Baseline: 27.3 4 months: 18.4 10 months: 18.3 TAU BDI-II: Baseline: 27.7 4 months: 20.3 10 months: 20.8
Van Straten 2008(20)	EQ-5D for intervention groups pre and post intervention Effect size (Cohen's D) for all scales	Mean EQ-5D scores: Control pre: 0.61 post: 0.66 Intervention pre: 0.62 post 0.73 Intervention complete: pre 0.63 post 0.8 Effect sizes (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 (0.14-0.72), 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 significant.
Sobocki 2007(22)	EQ-5D at baseline, first follow-up, 6 months follow-up, last visit	Mean EQ-5D scores: Baseline: 0.47 First follow-up: 0.60 6 months: 0.66 Last follow-up: 0.69
Swan 2004(9)	EQ-5D at baseline, week 12 and week 26. Improvement in BDI and GSI (baseline to week 12, baseline to week 26).	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)

QALY: quality of life years; MADRS: Montgomery-Asberg depression rating scale; CI: confidence interval; EQ-5D: EuroQoL 5 dimensions; ES: effect size; SRM: standardised response mean; CGI: clinical global impression scale; VAS: visual analogue scale; BAI: Beck anxiety inventory; BDI: Beck depression inventory; WHOQoL: WHO quality of life; BSQ: body shape questionnaire; ACQ: agoraphobic cognitions questionnaire; SF-6D: short-form 6 dimensions; MDD: major depressive disorder; SDS: self-rating depression scale; HAM-A: Hamilton anxiety scale; TAU: treatment as usual; MBI, Maslach Burnout Inventory

Table A7: Fields collected in the psychological therapy NCA

NAPT BASELINE DATA FIELDS (2nd round questionnaires in appendix)

SERVICE CONTEXT

11 – 11c decision tree questions (checks eligibility), Number of people in the service who deliver therapy for anxiety and depression, Whole Time Equivalents (WTE) the above represent IAPT funding (fully, partly or none), Sector managing the service (NHS, voluntary sector, private), Level of service (primary care, secondary care, mixture of primary and secondary care), Therapeutic modalities offered by the service for anxiety and depression, Age range of patients that are generally seen (working age only, older people 65+ only, both working age and older people), Access to therapy in another language than English (through therapists and interpretation services)

THERAPIST QUESTIONNAIRE (Anonymous *mandatory fields

What is your service's NAPT code*, What is the name of the service you are completing this for*, Number and response rate per service, Hours per week working in the service, In an average week, the number hours spent in direct contact with patients, Therapist's core profession / occupation, Therapies currently delivered and how the therapist developed their expertise (no formal training*, working with supervision without training, short workshops up to 10 days, formal training completed*, currently undertaking formal training*), Formal training completed (Doctorate, MSc/MA, postgraduate diploma, postgraduate certificate, other diploma, other certificate, other), Currently registered as a clinical practitioner with a professional body (Yes/No), If yes, the professional bodies are specified

RETROSPECTIVE AUDIT OF PATIENTS ENDING THERAPY BETWEEN 1 SEPT AND 30 NOV 2010

Total number of patients ending therapy during the audit period, Patient's PCT/LHB, Gender, Age in years, Ethnic Group, Main / Primary diagnosis, Secondary diagnosis, Date referral received, Date first appt offered / made, Date of first appt attended, Purpose of first appointment, Date of first treatment session offered / made, Date of first treatment session attended, Date of last attended treatment session, Number of therapy sessions attended, Reason why therapy ended (completed treatment, dropped out/unscheduled discontinuation, declined treatment, not suitable for service, deceased, unknown), Type of therapy provided, Outcome measure(s) used (Yes/No), If yes, scores requested at start of treatment and end of treatment; or the last occasion that the scale was rated

SERVICE USER QUESTIONNAIRE

Number and response rate per service

ACCESS

I was referred for talking treatment at the right time (Yes/No), The waiting time for my talking treatment to start was reasonable (Yes/No), My appointment was scheduled on a day/time that was convenient to me (Yes/No), I was able to get to my appointment location without too much difficulty (Yes/No), I received enough information about my talking treatment before it began (Yes/No)

OUTCOMES

This talking treatment helps me to understand my difficulties (Yes/No), I am getting the right kind of help (Yes/No), I am receiving the right number of sessions of talking treatment (Yes/No), If I have similar difficulties in the future, I would take up this talking treatment again (Yes/No), This talking treatment helps me cope with my difficulties (Yes/No)

ARM-5

Individual item scores and total score

BACKGROUND INFORMATION

Age, Gender, Ethnic group, Talking treatment (CBT, MBCT, person centred/humanistic, solution-focused, psychodynamic, CAT, counselling, low intensity treatment, other therapy, not sure), Number of sessions in the current course of treatment, Waiting time for current talking treatment to start (1 month or less, 1-3 months, 4-6 months, 7-9 months, 10-12 months, more than 12 months)

Table A8: NAPT SECOND ROUND DATA FIELDS (Psychological therapies NCA)

REGISTRATION FORM

Baseline participation (Yes and the service is fundamentally the same, Yes but the service has changed, No), Sector managing the service (NHS, voluntary sector, private), Level of service (primary care, secondary care, mixture of primary and secondary care), IAPT programme (yes, no), Stepped care (yes, no, don't know), Home visits (yes, no), Follow-up appointments (yes, no, only when clinically indicated), Number of people in the service who deliver therapy for anxiety and depression, Whole Time Equivalents (WTE) the above represent, Therapeutic modalities offered by the service for anxiety and depression Age range of patients (working age only, older people 65+ only, both working age and older people), Number of patients seen in 3 month period (less than 10, 11-50, 51-100, 101-500, 501+), Access to therapy in another language than English (yes -through therapists, yes – through interpretation services, no), Structured collection of service user feedback in last year (yes, no)

RETROSPECTIVE CASE RECORD AUDIT
(completed by therapist for all service users who ended therapy 1 July-31 October 2012)

PART A: Unique Identifiers (*mandatory fields)

NAPT Service Code*, Patient code*, Therapist initials

PART B: Patient information

Year of birth, Gender,* Ethnic group*

PART C: Referral (*mandatory fields)

Date referral received, Referral source*

PART D: Reason for therapy (*mandatory fields)

Diagnosis (list of conditions with ICS-10 codes for primary and secondary diagnosis)*, Problem for which psychological therapy offered* (Same as primary diagnosis, Same as secondary diagnosis, Depression, Mixed anxiety and depression, Social phobias, Specific (isolated phobias), Panic disorder (with or without agrophobia), Obsessive compulsive disorder, Generalized anxiety disorder, Post-traumatic stress disorder Body dysmorphic disorder,, Other anxiety disorder)

PART E: Appointment dates and Attendance (*mandatory fields)

Date of first appointment attended by service user, What was the reason for this first appointment* (Assessment only, Treatment only, Assessment and treatment, Not Known), Date of first treatment appointment attended by service user, Date of last treatment appointment attended, How many therapy sessions did the patient attend, What was the reason for therapy ending*(Completed Treatment, Deceased, Declined Treatment, Dropped out of Treatment (unscheduled discontinuation), Not Suitable for the Service, Referral to another Service, Not Known)

PART F: Type of therapy (*mandatory fields)

High intensity therapy provided to service user:* (individual/group: Cognitive Behavioural Therapy (CBT), Person-Centred (or other Humanistic Therapy), Solution-Focused Therapy, Psychodynamic/Psychoanalytic Therapy, Behavioural Activation, Interpersonal Therapy, Cognitive Analytic Therapy (CAT), Systemic/Family Therapy, Arts Psychotherapies, Mindfulness-Based Cognitive Therapy, Dialectical Behavioural Therapy, Counselling, Eye Movement Desensitisation & Reprocessing Therapy (EMDR), Problem-Solving Therapy, Couples Therapy

Low Intensity therapy provided to service user:* (individual/group: Computerised Cognitive-behavioural therapy –facilitated, Guided/Facilitated Self-Help, Psycho-Education, Pure Self-Help (e.g. books on prescriptions, unfacilitated cCBT via DVD, etc), Signposting/Referral Facilitation Schemes, Structured Exercise, Support and Advice in Adherence of Antidepressant/GP-Prescribed Medication, Other

PART G: Outcomes (*mandatory fields)

Enter outcomes scores you have for this service user (both first and last score): HADs (Anxiety, Depression Subscales), PHQ-9, GAD-7, W&SAS, CORE-10, CORE-OM*, CES-D, HoNOS, BAI, BDI-II, Other Standardised Measures

THERAPIST SURVEY (completed by someone who provides psychological therapy as part of their role in a service registered for NAPT)

Service NAPT code*, name of service completing audit for*, Number and response rate per service, Hours per week working in the service, In an average week, the number hours spent in direct contact with patients, Qualified member of staff or in training, Therapies currently delivered and how the therapist developed their expertise (no formal training, working with supervision without training, post-qualification CPD e.g. short workshops, currently undertaking formal training, formal training completed,), Currently registered as a clinical practitioner with a professional body (Yes/No), If yes, the

professional bodies are specified, Rating of formal supervision, Annual appraisal (Yes, No, prefer not to say), Degree to which organisation supports CPD, Supervision of other psychological therapists (Yes, No, prefer not to say), Received training in supervising other psychological therapists (Yes, No, prefer not to say)

TALKING TREATMENT SERVICE USER QUESTIONNAIRE - optional

ACCESS (all questions use 5 point Likert scale)

Please indicate if self-referral or referred by someone else, I was referred for talking treatment at the right time, The waiting time for my talking treatment to start was reasonable, My appointment was scheduled on a day/time that was convenient to me, I was able to get to my appointment location without too much difficulty, I received enough information about my talking treatment before it began

CHOICE

I was offered choice about the venue where my talking treatment would take place, I was offered choice about the time of day my talking treatment would take place, I was offered choice about the gender of my therapist, I was offered my talking treatment in another language or with an interpreter, I was offered choice about the type of talking treatment I would receive, 4 possible responses including: This was not important to me – I had no strong preference either way, This was important to me and I was given enough choice, This was important to me, but I was not given enough choice, unsure

EXPERIENCE OF THERAPY (all questions use 5 point Likert scale)

This talking treatment helped me to understand my difficulties, I am getting the right kind of help, I am receiving the right number of sessions of talking treatment, If I have similar difficulties in the future, I would take up this talking treatment again, This talking treatment helps me cope with my difficulties, I feel that my needs were taken seriously, understood and appropriately considered, I have experienced lasting bad effects from the treatment, I am asked by the therapist to give feedback on how helpful I am finding the treatment, I understand where my information is kept, who can see it and when it might be shared

BACKGROUND INFORMATION

age, gender, ethnic origin, whether or not have military background, sexuality, disability, therapy information (group vs. individual, type of therapy, number of sessions to date, length of wait, self-referral referred by someone else)

Table A9: Reports relating to other measures used in psychological therapies, with an emphasis on anxiety

Source, date	Population	Method used to reach recommendation			Measures recommended	Implementation issues for large scale use?
		General methods	Psychometric properties considered?	Measures considered		
Oxford PROMS group, 2009(38)	Anxiety and depression	Selected measures to review that have received significant recent attention in the NHS either through policy or professional recommendations or frequency of use. Aim to identify PROMS for use as evidence of quality and outcomes of services.	Considered, but not defined in methods section	PHQ 2 & 9 BDI-II HADS Whooley questions CORE-OM BAI GAD-7 WSAS EQ-5D SF-6D	Several caveats and options were discussed in the review, and different options recommended for different situations. Many PROMS in mental health have not been developed with responsiveness (a key requirement in assessment of services) in mind, but rather as screening/diagnostic tools. **For depression:** QOF, PHQ-9, BDI-II and HADS, though not much evidence about responsiveness **Anxiety:** BAI and GAD-7. Even less evidence of responsiveness. **For both:** HADS and CORE-OM – better evidence for CORE-OM, also includes social function. **To include social function**: WSAS, but little research available. **For preference measure:** EQ-5D preferred over SF-6D. **For recovery:** more evidence required. **Concluded:** Health professionals not strongly convinced of the value of PROMS in mental health. Debate about usefulness of PROMS in mental health is still in earliest stages.	Responsiveness evidence is low in quantity.
EMA research guideline,	Generalised anxiety	Expert panel and stakeholder consultation: Efficacy working	Unclear	Unclear	A rating scale, choice justified according to validity and reliability.	

2005(39)	disorder	party of CHMP			Hamilton anxiety rating scale (HAM-A) is mentioned as being widely used, but not an optimal scale, with the total scale as a primary endpoint, and the HAM-A psychic anxiety factor as a secondary endpoint.	
EMA research guideline, 2005(40)	Panic disorder	Expert panel and stakeholder consultation: Efficacy working party of CHMP	Unclear	Unclear	A rating scale, choice justified according to validity and reliability. Should include frequency and severity of panic attacks, severity of agoraphobic avoidance and anticipatory anxiety. Scales given as examples (i.e. not exhaustive list): Panic disorder severity scale (PDSS) Panic and Agoraphobia Scale (PAS) An improvement in scale should be supported by a relevant decrease in frequency and severity of attacks.	
EMA research guideline, 2006(41)	Social anxiety disorder	Expert panel and stakeholder consultation: Efficacy working party of committee for medicinal products for human use (CHMP)	Unclear, but mentioned for one measure.	Unclear	A rating scale, choice justified according to validity and reliability. Scales given as examples (i.e. not exhaustive list): Liebowitz Social Anxiety scale (LSAS) Brief Social Phobia Scale (BSPS) LSAS most commonly used. Paediatric version available. Self-rated scales include the Social phobia and anxiety inventory (SPAI); the social phobia inventory (SPIN), which has good psychometric properties; the Sheehan	LSAS and BSPS are clinician-administered

					disability scale (SDS).	
Royal College of Psychiatrists guideline(42)	Adult psychiatry (section on anxiety)	Methods not described.	Unclear	Unclear	Condition-specific scales given as examples (i.e. not exhaustive list): PHQ-9 (Patient Health Questionnaire – 9) A 9-item depression scale assessing symptoms and functional impairment to make a tentative diagnosis of depression, and deriving a severity score to help select and monitor treatment.[spitzer 1999] GAD -7 (Generalised Anxiety Disorder Assessment – 7) A self-administered, 7-item patient questionnaire used as screening tool and severity measure for generalised anxiety disorder.[Spitzer 2006] Clinical Outcomes in Routine Evaluation — Outcome Measure (CORE-OM) A client self-report questionnaire designed to be administered before and after therapy. The client is asked to respond to 34 questions about how they have been feeling over the past week, using a 5-point scale ranging from 'not at all' to 'most or all of the time'. The 34 items cover four dimensions: subjective well-being; problems/symptoms; life functioning; and risk/harm.[Barkham 2001] Generic HRQoL measures Both SF-6D and EQ-5D are recommended.	Several methods not relevant to WP1.2 were recommended: Patient-identified goals - can only be assessed in an ongoing treatment context Health of the Nation Outcome Scale – routinely collected as part of the minimum mental health data set, but is a clinician-administered measure.
Royal College	Older adults	Authored by a working group	Unclear	Unclear	Recommends	

of Psychiatrists guideline(43)					Patient Health Questionnaire – Generalised Anxiety Disorder severity index (GAD-7) Hospital anxiety and depression scale (HADS) – notes that as it is designed for use with patients with physical illnesses, and so may miss somatic symptoms, which it is not designed to assess.	

EMA: European Medicines Agency; HRQoL: Health related quality of life; CHMP: committee for medicinal products for human use; SF-6D: short form 6 dimensions

Reference List

(1) Peasgood T, razier J, apaioannou D. A systematic review of the validity and responsiveness of EQ-5D and SF-6D for depression and anxiety . HEDS Discussion Paper 2012;No.12.15(http://eprints.whiterose.ac.uk/74659).

(2) Fitzpatrick R, Gibbons E, Mackintosh A. An overview of patient-reported outcome measures for people with anxiety and depression. Report to the department of health, 2009. Patientreported Outcome Measurement Group 2009.

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

(4) Walters SJ. Quality of Life Outcomes in Clinical Trials and Health-Care Evaluation: A Practical Guide to analysis and interpretation. Wiley Blackwell; 2009.

(5) Gunther OH, Roick C, Angermeyer MC, Konig HH. The responsiveness of EQ-5D utility scores in patients with depression: A comparison with instruments measuring quality of life, psychopathology and social functioning. J Affect Disord 2008 Jan;105(1-3):81-91.

(6) Saarni SI, Suvisaari J, Sintonen H, Pirkola S, Koskinen S, Aromaa A, et al. Impact of psychiatric disorders on health-related quality of life: general population survey. Br J Psychiatry 2007 Apr;190:326-32.

(7) 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. Soc Psychiatry Psychiatr Epidemiol 2009 Jul;44(7):569-78.

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

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

(10) Peveler R, Kendrick T, Buxton M, Longworth L, Baldwin D, Moore M, et al. A randomised controlled trial to compare the cost-effectiveness of tricyclic antidepressants, selective serotonin reuptake inhibitors and lofepramine. Health Technol Assess 2005 May;9(16):1-134, iii.

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

(12) Caruso R, Rossi A, Barraco A, Quail D, Grassi L. The Factors Influencing Depression Endpoints Research (FINDER) study: final results of Italian patients with depression. Ann Gen Psychiatry 2010;9:33.

(13) Fernandez JL, Montgomery S, Francois C. Evaluation of the cost effectiveness of escitalopram versus venlafaxine XR in major depressive disorder. Pharmacoeconomics 2005;23(2):155-67.

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

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

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

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

(18) Bosmans JE, Hermens ML, de Bruijne MC, van Hout HP, Terluin B, Bouter LM, et al. Cost-effectiveness of usual general practitioner care with or without antidepressant medication for patients with minor or mild-major depression. J Affect Disord 2008 Nov;111(1):106-12.

(19) Lamers LM, Bouwmans CA, van SA, Donker MC, Hakkaart L. Comparison of EQ-5D and SF-6D utilities in mental health patients. Health Econ 2006 Nov;15(11):1229-36.

(20) van SA, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008;10(1):e7.

(21) Supina AL, Johnson JA, Patten SB, Williams JV, Maxwell CJ. The usefulness of the EQ-5D in differentiating among persons with major depressive episode and anxiety. Qual Life Res 2007 Jun;16(5):749-54.

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

(23) Aydemir R, Ergun H, Soygur H, Kesebir S, Tulunay C. Quality of life in major depressive disorder: A cross sectional study. Turkish Journal of Psychiatry 2009;20:205-12.

(24) Ergun H, Aydemir O, Kesebir S, Soygur H, Tulunay FC. SF-36 and EQ-5D quality of life instruments in major depressive disorder patients: Comparisons of two different treatment options. Value Health 2007;10(6):A303.

(25) Mychaskiw M, Hoffman D, Dodge W. EQ-5D index scores by remission status in patients with generalized anxiety disorder. Int J Neuropsychopharmacol 2008;11:279.

(26) Brazier J, Ratcliffe J, Tsuchiya A, Solomon J. Measuring and valuing health for economic evalation. Oxford: Oxford University Press; 2007.

(27) EEPRU Programme. EEPRU. www eepru org uk 2014Available from: URL: www.eepru.org.uk

(28) Kaltenthaler E, Shackley P, Stevens K, Beverley C, Parry G, Chilcott J. Computerised cognitive behaviour therapy for depression and anxiety TA51. 2002.

(29) National Institute for Health and Care Excellence. Computerised cognitive behaviour therapy for depression and anxiety. Review of Technology Appraisal 51 TA97. 2006.

(30) National Institute for Health and Care Excellence. Depression in adults. The treatment and management of depression in adults CG90. 2009.

(31) McCrone P, Chisholm D, Knapp M, Hughes R, Comi G, Dalakas MC, et al. Cost-utility analysis of intravenous immunoglobulin and prednisolone for chronic inflammatory demyelinating polyradiculoneuropathy. Eur J Neurol 2003 Nov;10(6):687-94.

(32) Kuyken W, Byford S, Taylor RS, Watkins E, Holden E, White K, et al. Mindfulness-based cognitive therapy to prevent relapse in recurrent depression. J Consult Clin Psychol 2008 Dec;76(6):966-78.

(33) Scott J, Palmer S, Paykel E, Teasdale J, Hayhurst H. Use of cognitive therapy for relapse prevention in chronic depression. Cost-effectiveness study. Br J Psychiatry 2003 Mar;182:221-7.

(34) Kaltenthaler E, Parry G, Beverley C, Ferriter M. Computerised cognitive-behavioural therapy for depression: systematic review. Br J Psychiatry 2008 Sep;193(3):181-4.

(35) Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. Behav Res Ther 1997 Aug;35(8):785-91.

(36) Beck AT, Steer RA, Brown G. BDI-Fast Screen for Medical Patients: Manual. San Antonio, Texas: The Psychological Corporation; 2000.

(37) Cahill J, Stiles WB, Barkham M, Hardy GE, Stone G, Agnew-Davies R, et al. Two short forms of the Agnew Relationship Measure: the ARM-5 and ARM-12. Psychother Res 2012;22(3):241-55.

(38) Patient Reported Outcomes Measurement Group O. An overview of patient-reported outcome measures for people with anxiety and depression. 2009.

(39) European Medicines Agency. Guideline on the Clinical Investigation of Medicinal Products Indicated for Generalised Anxiety Disorder. 2005.

(40) European Medicines Agency. Guideline on Clinical Investigation of Medicinal Products Indicated for the Treatment of Panic Disorder. 2005.

(41) European Medicines Agency. Guideline on clinical investigation of medicinal products indicated for the treatment of Social Anxiety Disorder (SAD). European Medicines Agency; 2006.

(42) Royal College of Psychiatrists. Outcomes Measures Recommended for use in Adult Psychiatry. 2011. Report No.: OP78.

(43) Royal College of Psychiatrists Faculty of the Psychiatry of Old Age. Individual Patient Outcome Measures Recommended for Use in Older People's Mental Health. 2012. Report No.: OP86.