PATIENT-REPORTED OUTCOME MEASUREMENT GROUP, OXFORD

AN OVERVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE WITH ANXIETY AND DEPRESSION

Report to the Department of Health, 2009

patient-reported

OUTCOME

measures

AN OVERVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE WITH ANXIETY AND DEPRESSION 2009

Ray Fitzpatrick
Elizabeth Gibbons
Anne Mackintosh

Patient-reported Outcome Measurement Group
Department of Public Health
University of Oxford

Copies of this report can be obtained from:

Elizabeth Gibbons
Senior Research Officer
Patient-reported Outcome Measurement Group

tel: +44 (0)1865 289402
e-mail: elizabeth.gibbons@dphpc.ox.ac.uk

Alternatively, it can be downloaded free of charge from the PROM Group website:

http://phi.uhce.ox.ac.uk/

EXECUTIVE SUMMARY

The purpose of this review is to scope the main issues regarding choice of patient-reported outcome measures (PROMs) for use in relation to depression and anxiety in a general population setting and to provide an overview of evidence for more commonly used or recommended instruments currently in the NHS. The context of the review is the potential for further exploration of the use of PROMs in the NHS to assess quality and outcomes of services. The review does not therefore examine related but distinct issues regarding PROMs used to screen for mental health. Nor does the review consider issues of patient experience or satisfaction with care.

The search strategy for this review was based on a decision to confine attention to those PROMs that have received significant recent attention in the NHS either through policy or professional recommendations or frequency of use.

Nine specific measures are reviewed and two broader classes of measure; preference measures and recovery measures

A number of strategic considerations were identified that needed to be addressed prior to selecting PROMs for further testing for use in the NHS. A first general issue that needs to be considered is whether to develop an approach to outcome assessment in relation to depression and anxiety separately. A second general issue that needs to be considered in determining a strategy for use of PROMs in depression and anxiety is the extent to which the focus should be on symptoms and symptom reduction or on broader aspects of function and health-related quality of life. A third consideration is whether a PROMs strategy for depression and anxiety should build on currently established uses of such PROMs in primary care. A fourth issue is to consider whether longitudinal change within individuals is a likely feature of future applications and whether stronger evidence of responsiveness is required from instruments than is currently available.

Evidence for most instruments as measures of outcome in populations was more limited than is often the case with PROMs for physical health. Research has tended to focus on uses such as screening. No compelling evidence for a condition-specific PROM was found although PHQ-9 has supportive evidence and is likely to be broadly acceptable. There are logistic advantages to using an instrument more broadly applicable to depression and anxiety. The CORE-OM seems particularly appropriate although it needs to be asked whether it has been tested against the full spectrum of presentations and interventions for depression and anxiety in primary care. If a preference measure is needed, there is sufficient evidence to prefer EQ-5D. Whilst recovery measures address distinct issues such as hope and recovery that are thought of major importance to users, more evidence is needed before any specific PROM can be high-lighted.

The evidence is that health professionals are still not strongly convinced of the value of PROMs. Moreover there is evidence that response rates in follow-up of patients via PROMs are moderate. In addition to debate about most promising PROMs, consideration has to be given to the logistics of administering PROMs for mental health.

AN OVERVIEW OF PATIENT-REPORTED OUTCOME MEASURES FOR PEOPLE WITH ANXIETY AND DEPRESSION

Purpose of review

The purpose of this review is to scope the main issues regarding choice of patient-reported outcome measures (PROMs) for use in relation to depression and anxiety in a general population setting and to provide an overview of evidence for more commonly used or recommended instruments currently in the NHS. The context of the review is the potential for further exploration of the use of PROMs in the NHS to assess quality and outcomes of services. The review does not therefore examine related but distinct issues regarding PROMs used to screen for mental health. Nor does the review consider issues of patient experience or satisfaction. A full systematic review of all PROMs in the context of depression and anxiety would be unwieldy given the very large number of instruments and substantial evidence in relation to some measures. The review is more a more limited scoping review of more promising or plausible measures.

The philosophy behind the review is that in principle a measure or small number of measures might eventually be identified that could be recommended for more general and standard use in relation to services for depression and anxiety in NHS, following the model of the process that resulted in PROMs being identified to monitor quality and outcomes for specific elective surgical procedures.

Introduction

The World Health Organisation recognizes depression as a major health problem which impacts on patient functioning, work productivity and healthcare utilisation (Kroenke et al., 2008). The number of people who experience depression each year in the UK is estimated to be 2.6 million with associated direct costs of treating depression as well as cost to the economy in lost output, tax receipts and benefit payments. It is not clear though about the direction of association between the financial implications of depression and anxiety, for example, whether debt and financial burden is one of the causes or effects (APMS 2007).

There is now recognition of the morbidity associated with depression and anxiety and whilst there remains a huge focus on the management of depression in primary and secondary care and for people with LTCs, mental health of the whole community is priority. Often, anxiety and depression are under diagnosed, are relapsing conditions and often patients either do not seek treatment or fail to adhere to treatment. Risk factors include increasing age, female, education, family history of depression, SF-12 scores, employment difficulties and experiences of discrimination (King et al., 2006, 2008).

UK policy context

The promotion of good mental health is the focus of recent policy consultation- *New Horizons; Towards a shared vision for mental health, Consultation (2009).* This

details a strategy to promote good mental health and well-being by improving services for people who have mental health problems. These strategies build on the NSF for Mental Health (1999) as well as supporting the delivery of the NHS Next Stage Review (Darzi 2008) and have several overarching aims but specifically:

- To improve mental health and well-being of the population
- To improve quality and accessibility of services for people with poor mental health.

Early recognition of mental ill-health is essential to be able to prevent the long-term consequences of mental health problems by enabling early interventions (DH. 2009). The *New Horizons* strategy is focused on child, adult and older people's mental health; tackling diversity and ensuring equality and social justice. Social exclusion is both a cause and consequence of mental health problems; stigma and discrimination have negative effects on mental health. High quality care is featured within the consultation with reference to clinical effectiveness of interventions and care pathways which are evidence based. Quality of life of service users is considered to be the centre of service design and delivery.

It has been ten years since the publication of the National Service Framework for Mental Health (1999) and other associated policies which set out standards and guidance for the promotion of mental health. The ten year programme of reform set out in the NSF standards were to promote mental health; increase access to services ; and to ensure that primary and secondary service models were effective (Appleby. 2007). The NHS plan (2000) identified mental health as one of the key clinical priorities with particular emphasis on strengthening community services for people with mental illness. Following this the Department of Health (DH) commissioned the National Institute for Clinical Excellence (NICE) to develop evidence-based guidelines for the main mental disorders including two which were specific to depression (2004) and anxiety (2004; 2007). The depression guidance has currently been updated and published and includes the management of depression in primary care and for patients with long-term conditions (LTC) (NICE 2009).

The NICE (2004, 2009) guidance includes recommendations for depression screening to be undertaken in primary and secondary care for high-risk groups- those with previous history of depression, significant physical illness causing disability and other mental health problems such as dementia. This should be conducted in 2 stages. Firstly, screening should include the use of two questions concerning mood and interest referred to as the Whooley questions[1]:

- During the last month, have you been bothered by feeling down, depressed or hopeless?
- During the last month, have you often been bothered by having little interest or pleasure in doing things?

Secondly, a positive response to these questions suggests a person may be at risk of or actually experiencing depression and should be referred to an appropriately trained

[1] The PHQ-2 was referred to in previous NICE guidance (2004)

healthcare professional for further comprehensive mental health assessment including mental health state and associated functioning, interpersonal and social difficulties. The guidance suggests that healthcare professionals may consider using a validated measure of symptoms, functions and or disability to inform and evaluate treatment (NICE 2009) but that assessment should include severity, duration and course.

It has been reported that GPs are more likely to detect severe depression in their patients unassisted but that those with mild depression and less symptoms are less likely to be identified (NICE. 2009). Furthermore, a meta-analysis of the accuracy of unassisted diagnoses of depression by GPs suggests that there are a high number of false positive cases identified. It is suggested that diagnosis could be improved by re-assessment of those who might be depressed (Mitchell et al., 2009).

Using diagnostic criteria for depression is problematic as outlined by NICE (2009). Thresholds for identifying depressive episodes using DSM-IV or ICD-10 differ namely by the number of symptoms present. It is suggested that it may be possible that more people are identified as depressed using ICD-10 (NICE. 2009). NICE guidance considers the DSM-IV to be of more value and it is the reference frame for the evidence base of recommendations.

The management of depression

NICE (2004,2009) based on the evidence, made recommendations for a Stepped Care approach to the management of depression to ensure that patient receive the most effective interventions for the severity of their symptoms. Psychological therapies were recommended for the management of depression, anxiety and related disorders. These recommendations are supported by good evidence of increased compliance and tolerability of psychological therapies compared to other treatments.

Despite these recommendations, at the time, there was a lack of therapists to provide such therapies and increased funding was made available. Several other reports and policies led to the development of guidance and initiatives' to increase access to such treatments. A report by Lord Layard (London School of Economics, 2006) supported a policy to promote access to psychological therapies to decrease the psychiatric burden of depression and anxiety and also the economic burden to the person suffering with the condition. In addition he argued that there would be economic benefits by reducing the loss of output to society and by reducing healthcare costs. Furthermore, the publication of Our health, Our Care, Our Say by the DH (2005) provided guidance for further development of community health and social care services and more personalised care for people with health problems.

The *Increased Access to Psychological Therapies (IAPT)* programme was the outcome of these reports and guidance and initiated by the DH in 2005 with the principal aim of supporting Primary Care Trusts (PCTs) in implementing NICE guidance. Additional funding was provided to support the implementation process. An outcome framework for the collection of service level performance indicators and individual patient health outcomes to monitor progress and for benchmarking was developed (DH, 2008). This will be discussed further in this document.

The measurement of outcomes in mental health

The measurement of outcomes is not new to the mental health sector. It has become central to policy for patient monitoring and service development (Evans et al., 2006). Furthermore, the collection of Mental Health Minimum Dataset (MHMDS) has been compulsory for NHS providers since April 2003 (NSF 1999). The MDS was developed to provide information for the planning and monitoring of services. This can be achieved at patient level with the monitoring of outcomes of care and nationally as part of monitoring services.

The DH Mental Health Information Strategy (2001) introduced mental health electronic records and set out a minimum dataset for the evaluation of services with clinical data collected by mental health professional using the Health of the Nations Outcome Scale (HoNOS). The HoNOS system was developed to measure the health and social functioning of people with severe mental illness and was recommended in the NSF for Mental Health (1999). It is a clinical rating system and assessments are made by trained mental health care professionals. Outcome measurement has traditionally been driven by such clinician-assessed methods but alongside this is the drive nationally to measure outcomes from the patient's perspective.

Methods of this report

The search strategy for this review was based on a decision to confine attention to those PROMs that have received significant recent attention in the NHS either through policy or professional recommendations or frequency of use.

The primary source of evidence was the bibliographic database compiled by the PROM group with funding from the Department of Health and hosted by the University of Oxford [2] and also from 2005 the property of the NHS Information Centre for Health & Social Care. These records were searched using keywords 'mental health' and abstract and title searches using 'depress*' and 'anxiety'. Searches also used Pubmed, instrument websites and other recent reviews. Additional searches were performed of specifically identified instruments, where necessary working backward from cited references..

PROMs for depression and anxiety

Since the development and publication of the NSF there has been an exponential growth in the publication of patient-reported outcome measures for mental health and associated conditions. A review in 2000, by the Patient-reported Outcome Measurement Group, Oxford identified 85 instruments with some evidence of measurement performance which were specific to mental health.

Patient-reported outcome measures (PROMs) offer enormous potential to improve the quality and outcomes of mental health services. Lord Darzi's Interim Report on the future of the NHS recommends that patient-reported outcome measures (PROMs)

[2] Available online at http://phi.uhce.ox.ac.uk

should have a greater role in the NHS (Darzi 2007). Guidance has now been issued regarding the routine collection of PROMs for selected elective procedures (DH, 2008).

However, the focus of patient-reported outcome measurement for people with depression has often been related to the presence and severity of depressive symptoms and not the greater impact of the disease on physical, social and role functioning. A number of instruments have been developed to assess depressive symptoms from the patient's perspective and are recommended in several of the above mentioned policies and guidance's. These will be discussed in further detail in the next section.

At present, the climate seems to be very much focused on the measurement of both clinical performance indicators and patient-reported health and experiences. The policies and guidance underpinning these changes have been introduced in the section above. In terms of specific depression questionnaires, many are recommended, particularly for screening.

The most recent policy driven approach to outcome measurement in the NHS is the Quality and Outcomes Framework (QOF) (2004 and subsequent revisions). Whilst QOF encompasses a broader outcome framework, there are domains specific to mental health. The QOF is an annual financial reward and incentive scheme for GP surgeries in England. Several indicators in five domains are specified for the measurement of achievement. Included in the clinical domain are indicators for mental health, and specifically for depression. The QOF has specified depression indictors which provide financial incentives. DEP 2 specifies using a validated tool to assess the severity of depression both at case finding and DEP 3 for further assessment (QOF 2009/10). The indicators are underpinned by the evidence from NICE recommendations (NICE 2004). Three specific depression severity measures are suggested in the QOF 2009/10: The PHQ-9, BDI-II and the HADS. Specific score thresholds for considering interventions are suggested for each instrument:

PHQ-9: 12
BDI-II: 20
HADS: 10

The Care Services Improvement Partnership (CSIP) and National Institute for Mental Health in England (NIMHE) provide further guidance for measurement of the indicators outlined in the QOF and support the use of the instruments mentioned above. In addition to this, the IAPT outcomes framework outlines specific measures for different stages of the patient journey.

It was decided to review the evidence for the chosen instruments in the QOF framework and also identify any promising instruments from the IAPT outcome framework.

The most frequently proposed measures are the Patient Health Questionnaires (2 & 9), Beck Depression Inventory (BDI) and the Hospital Anxiety and Depression Scale (HADS). The Whooley questions are suggested in the NICE (2009) guidance for

depression and will also be considered. *These questions are broadly similar to the PHQ-2.*

In addition, the following measures are included as optional measures suggested in the CSIP guidance and the IAPT MDS:

- CORE-OM
- Beck Anxiety Inventory (BAI)
- Generalised Anxiety Disorder-7 (GAD-7)
- Work & Social Adjustment Scale (WSAS)
- Preference-based measurement-EQ-5D/SF-6D

What follows is an overview of these measures with reference to measurement performance in terms of psychometric criteria and operational characteristics.

a). PHQ-9

The PHQ was originally developed from the self-administered version of the Primary Care Evaluation of Mental Disorders (PRIME-MD). The PRIME-MD uses diagnostic criteria from the Diagnostic and Statistical Manual of Mental Disorders (DSM). The original PRIME-MD assessed 18 mental disorders and the PHQ incorporated 8 symptoms to simplify the differential diagnosis. These include depressed mood, anhedonia, appetite change, sleep disturbance, psychomotor agitation or retardation, loss of energy, feelings of worthlessness or guilt, diminished concentration and suicidal thoughts or attempts (Dale et al., 2007). The response categories for the PHQ were increased to capture not only the presence of symptoms but also the severity to aid treatment decisions and monitor change over time. A ninth item included a question about the difficultly people have related to role and social functioning specific to the symptoms present. Scores range from 0 to 27 with a three point scale for the 9 items. Mild depression is considered with scores of 5 to 9; moderate for scores between 10 and 14 and severe, 20 to 27 (Spitzer et al., 1999).
A 2, 8 and15 item versions are also available (Kroenke et al., 2002). It takes less than three minutes to complete and less time to administer was reported for the PHQ-9 than the original PRIME-MD (Spitzer et al., 1999).

The threshold diagnosis for depression includes the presence of three symptoms (depressed mood and loss of interest) therefore, if patients did not endorse these symptoms they were considered negative cases (Spitzer et al., 1999).

Measurement and operational characteristics of the PHQ-9 were comparable to the original PRIME-MD with higher sensitivity for major depressive disorders. High correlation has been reported between Mental Health Professionals diagnosis and the PHQ-9 levels of severity (Spitzer et al., 1999).

The PHQ-9 has principally been evaluated in terms of its diagnostic ability. High sensitivity and specificity have been reported from major studies by the developers: accuracy=85%, sensitivity=75% and specificity=90% (Spitzer et al., 1999). Kroenke et al., (2001) reported higher sensitivity (88%) and similar specificity (88%) using a cut-off score of $\geq$10. Higher sensitivity is reported more recently (Gilbody et al.,

2007). Further studies have reported good evidence of its discriminative ability with clinical interviews as the criterion (Dale et al., 2007).

Substantial association has been reported between increasing PHQ-9 scores and the likelihood of major depression (Kroenke et al., 2002). Patients completing the PHQ-9 reported more symptoms than those documented by physicians from a medical chart review indicting that two thirds of patients with depression may be under diagnosed (Ani et al., 2008).

High internal consistency (0.89) and reproducibility is reported (Kroenke et al., 2002). Comparable performance of the PHQ-9 and HADS is reported for internal structure (Cameron et al., 2008).

Convergent validity is supported with high correlations between the PHQ-9 and HADS. Moderate correlation is reported between PHQ scores and utility measures-SF-6D and HUI (Revicki et al., 2008).

Cameron et al., (2008) evaluated the responsiveness of both the PHQ and the HADS in a group of patients who had been referred to a mental health worker for management. Statistically significant differences in scores pre and post intervention were reported for both measures (expressed as effect sizes 0.99 for PHQ and 1.0 for the HADS.

It has also been used to identify depression in older people as part of an assessment system (Ell et al., 2005) with greater than 70% compliance.

The PHQ-9 has been used to evaluate new multidisciplinary services for the management of depression in a community practice in the US (Dietrich et al., 2003). The PHQ-9 was reported as a feasible instrument to use by staff and approximately it took two hours to train staff to use the PHQ and administer by telephone. .

Electronic administration of the questionnaire has been evaluated and high completion rates and acceptability reported from patients and staff (Klien et al., 2006).
High correlation of responses is reported between scores from patients who self-completed in the clinic and those obtained by telephone (Kroenke et al., 2002).

Physicians' views of using the PHQ-9 were positive and patients found the questionnaire helpful (Spitzer et al., 1999). It took less than 3 minutes for physicians to review the questionnaire once completed by the patient.

A significant study carried out in 38 GP practices in the UK by Kendrick et al., (2009) reports the rates of GP anti-depressant prescribing and referrals to specialist services in relation to patient scores on the PHQ, HADS and BDI based on surgery preference Twice as many patients were assessed using either the PHQ or HADS than they anticipated. Fewer practices had used the BDI. The distribution of scores for the instruments was normal although the PHQ patient scores were slightly positively skewed. In this study, of 1658 patients who completed the PHQ, 83% were categorised as having moderate to severe depression compared to 55% of the 584 patients assessed with the HADS. Of these patients who were classified as having

moderate to severe depression, 80% received prescriptions for anti-depressants and 20% were referred for specialist services. This is inline with current treatment clinical guideline recommendations (NICE 2004).

Furthermore, in the Kendrick study, reported a significant association between anti-depressant prescriptions or referrals with patients with moderate to severe depression as classified by PHQ, HADS or BDI. This study identified that older people and those people with co morbidity were less likely to be offered treatment for depression based on scores from the questionnaires.

An 8 item version was used in a study by Kroenke et al., (2008). The item related to suicidal or self injurious thoughts were omitted to facilitate completion by telephone administration. Identical scoring thresholds for depression severity have been established (Kroenke and Spitzer. 2002). 100% sensitivity and 95% specificity was reported using a cut-off score of $\geq$10 for major depressive disorder with a likelihood ratio of 15 indicating a PHQ-8 score of 10-14 is 15 times more likely to occur in persons with a depressive disorder (Kroenke et al., 2008).

b). Beck Depression Inventory (BDI-II)

The BDI was developed as an interviewer-administered measure of the intensity of depression in psychiatric patients (i.e. those already diagnosed with depression) (Beck et al., 1961, but has come to be used widely also for detecting depression in general populations and been revised and modified several times and is commonly used in self completed questionnaire format. The scale comprises 21 items reflecting particular aspects of depression (symptoms and attitudes). For each, the respondent selects one of four statements, rated in severity from 0 to 3, and the total is calculated; score-ranges indicate whether depression is absent/minimal, mild, moderate, or severe.

Principal factor analysis of the BDI-II with a college student sample revealed two dimensions: Somatic-Affective and Cognitive, later re-labeled Cognitive (8 items) and Non-Cognitive (13 items) (Beck et al., 1996a; Steer et al., 1999a). Steer et al. (1999a) recommend the two subscales be scored separately, and suggest that the Cognitive subscale may be particularly useful for measuring depression in patients with somatic symptoms attributable to medical conditions.

Internal consistency and test-retest reliability of the BDI-II in psychiatric and non-psychiatric samples were widely supported in two major reviews (Beck et al., 1998; Richter et al., 1998). However, high internal consistency of the BDI could indicate item redundancy, whilst stability of scores may be the result of insensitivity to fluctuations in depression (Boyle, 1985). Others (Richter et al., 1998; Demyttenaere & De Fruyt, 2003) have found evidence of BDI instability. They suggest this could be partly due to items being formulated unidirectionally - i.e. responses are consistently ordered from least to most pathological, possibly resulting in a carryover effect with repeated measurement. Randomising the order of the answer options within each item could increase reliability of the BDI (Dahlstrom et al., 1990).

Content validity of the BDI is strongly supported by Richter et al., (1998) but challenged by Demyttenaere & De Fruyt (2003), who suggest that the measure is not specific to depression. Owing to the diversity of items, patients with very different profiles and treatment needs may be assigned the same (total) score; however, they contend, this weakness is shared by most depression measures. Headey et al., (1993) also find that some items relate to anxiety rather than depression.

Construct validity is supported by strong correlations with well-researched measures of depression, both self-report (Steer et al., 1997) and clinician rated – in particular, the Hamilton Rating Scale for Depression with which the BDI is often paired (Beck et al., 1988; Richter et al., 1998).

Findings in respect of discriminant validity are mixed. Although the BDI has been shown to distinguish depressed and non-depressed patients, its ability to differentiate subtypes of depression is not confirmed (Beck et al., 1988; Richter et al., 1998). It has also been criticised for failing to discriminate depression and anxiety (Richter et al., 1998; Enns et al., 1998), despite claims to the contrary (Steer et al., 1986). It is suggested that the BDI is unsuitable for use in patients with more than moderate depression (Richter et al., 1998; Kearns et al., 1982).

Despite the very widespread use of the BDI for nearly 50 years, which would appear to support feasibility and acceptability of the measure, a number of critics note interview burden and difficulty of completion as disadvantages (Kearns et al., 1982; Richter et al., 1998). The response format (choice among four statements) makes it unsuitable for telephone administration (Arnau et al., 2001; Williams, 2002). There is also a considerable financial cost involved. There would also seem to be limited evidence of assessment of the instrument's responsiveness.

c). Hospital Anxiety and Depression Scale/HADS

The HADS was developed as a screening instrument for use in hospital outpatient departments; it has subsequently been validated for use with primary care patients and the general population (Snaith, 2003; Bjelland et al., 2002). Its aim is to detect the presence and severity of depression and anxiety in non-psychiatric settings. The measure comprises 14 items divided equally between the two mood states (anxiety and depression), with 4-point verbal rating scales for each item. Respondents are asked to rate items according to how they have felt during the previous week; cut-off points indicate whether the respondent is 'within the normal range', or mildly, moderately, or severely disordered. The subscales are intended to be considered separately, not summated (Snaith, 1990), and can be self- or interviewer administered.

Depression subscale items focus on the anhedonic state (loss of pleasure), regarded by Zigmond & Snaith (1983) as the cardinal symptom of depression, and the most likely to respond to antidepressant drug therapy. Care was taken to avoid items that could be attributed to physical illness, such as headache or dizziness. The HADS is not intended for severe mood disorder, hence does not attempt to cover the full range of symptoms for clinical depression.

The two-factor structure, internal consistency, test-retest reliability, and construct validity of the HADS were supported in two major reviews (Herrmann et al., 1997; Bjelland et al., 2002).

However, content validity of the HADS is challenged in a recent review by Martin (2005) who concludes that HADS can no longer be regarded as a reliable and valid measure of two distinct dimensions, anxiety and depression, as it is essentially a tri-dimensional measure. It therefore accords with Clark and Watson's tripartite theory of anxiety and depression (Clark & Watson, 1991) but was not devised with this model in mind, and is not a sufficient as a model of the three constructs (Dunbar et al., 2000). To devise a scoring method to take account of this would nullify one of the main advantages of the measure, namely, its ease of scoring and interpretation; Martin (2005) therefore concludes that the HADS should now be superseded.

Construct validity has been demonstrated by moderate to strong correlations with comparator measures, both clinician-rated and self-report (Snaith & Taylor, 1985; Aylard et al., 1987).

Parker et al. (2001) suggest poor validity of the HADS may be due to the fact that anhedonia (the main component of the instrument) does not differentiate depressed from non-depressed subjects in the medically ill, despite being one of the DSM-IV criteria for diagnosis of MDD. As anhedonia is present in other mental disorders (Silverstone, 1991), it is argued this symptom alone should not form the basis of a diagnostic measure.

Good results for sensitivity and specificity of the HADS using the recommended cutoff of 8+ to define caseness have been reported (Herrmann et al., 1997; Bjelland et al., 2002). In a study aimed at establishing (UK) general population norms, Crawford et al. (2001) found the recommended cutoff score of 8 for 'caseness' too low and suggest 10/11 may be more appropriate. This reflected findings of an earlier study (Dowell & Biran, 1990) which also found a high proportion of false positives with a cutoff at 8; however, these findings were challenged by Snaith (1990) as the two scales were summated, contrary to the developers' recommendations.

Cameron et al., (2008) evaluated the responsiveness of both the PHQ and the HADS in a group of patients who had been referred to a mental health worker for management. Statistically significant differences in scores pre and post intervention were reported for both measures (expressed as effect sizes 0.99 for PHQ and 1.0 for the HADS.

Incorporation of the scoring system in the question sheet (though it can easily be concealed, if wished) enhances feasibility of the HADS. The developer's state that the scale is very acceptable to patients and that there appears to be no difficulties of comprehension; this is supported in a major review (Herrmann et al., 1997) which found response rates of 95%-100%. However, one reviewer suggests the measure has a difficult reading level (Williams et al., 2002).

d). CORE-OM

The CORE-OM 10 is recommended in the IAPT framework MDS to be used at intake as part of the generic assessment tool and at review sessions. This ten itemed measure is part of a collection of the Clinical Outcomes in Routine Evaluation system (CORE). The system was developed for quality evaluation, audit and outcome benchmarking for psychological therapy services and as a generic measure of emotional problems. It comprises three components ('hub') which can be complimented by population or condition-specific 'spokes'. The 'hub' includes the following data collection measures:

-CORE administration checklist for the monitoring of data administration and collection;
-CORE Assessment and End of Therapy Forms which are practitioner-completed forms to compliment patient-reported forms;
- CORE Outcome Measure (CORE-OM) which is a client-completed form to assess distress; subjective well-being; problems and symptoms; and life and social functioning.

Items about risk to self and others are also included. It is intended to be a measure of global distress and used as a screening tool and outcome measure.

The development of the CORE was underpinned by the evidence-based practice framework and the Department of Health Strategic review of Psychotherapy (1996). Computerised data entry was developed with anonymised data collated in a national database (Barkham et al., 1998). The CORE battery acts as a screen and aids deciding what other outcomes to measure. The use of referential and problem-specific measures compliments the CORE battery. Under the outcome framework referred to by Barkham et al., (1998) referential measures suggested include BDI and HADS. Problem-specific measures referred to be those developed for specific mental health conditions for example schizophrenia, bipolar disorders etc.

Barkham et al., (2001) surveyed the views of service providers and commissioners about the desirable aspects of a core outcome system and measures which were in use at the time. The five most frequently used measures at that time were the Beck Depression Inventory, Symptom Checklist-90 or Brief Symptom Inventory, Inventory of Interpersonal Problems (IIP), Hospital Anxiety and Depression Scale (HADS), and the General Health Questionnaire (GHQ). Items were pooled for these measures and others to generate a list for empirical evaluation. The item pool tapped intrapersonal and interpersonal experiences and positive and negative aspects of experience. The final version of the CORE-OM includes 34 items in four domains- Subjective well-being (4). Problems (12), Functioning (12) and Risk (6). Subsequent factor analysis though suggests 2 scales, one general factor of Psychological distress and another for the Risk items (Lyne et al., 2006). Ten and 5 itemed versions are available. The CORE-OM 10 is recommended in the IAPT MDS framework.

Two parallel short-forms were also developed. The mean of all 34 items provides an index of global distress. Mean item scores for each dimension can also be used if needed. Each item is scored on a 5 point Likert scale (0=not at all, to 4=most of the

time). The minimum score is 0 and maximum 136. Strategies for missing data are outlined.

While the CORE-OM has been developed for use across psychological therapy services, nine items within the instrument include symptoms suggested for the assessment of depression (NICE 2004) and 7 items for anxiety (NICE 2004). Furthermore, Leach et al., (2006) provide analyses of score transformation from the CORE-OM to predict values on the BDI (II) including cut-off points for diagnoses of mild, moderate and severe depressive symptoms. The utility of this method may be to reduce burden of completion of a battery of patient-reported measures if not needed.

Normative data is available with reports of statistically significant difference in scores from a clinical and non-clinical population (Barkham et al., 1998). Clinically significant change scores are presented for males and females based on these data. For males a cut-off score of 1.19 and females, 1.29 is suggested as clinically significant. A cut-off score of 2.5 is reported as 'severe'. Domain cut-off values are also reported. Descriptors Mild, Moderate and Severe and boundaries are suggested (Mullin et al., 2006). Further cut-off values are reported from clinical and non-clinical populations with a score of 10 representing maximum sensitivity and specificity (Connell et al., 2007).

The developers suggest that a primary measure of service quality and performance is the percentage of patients achieving reliable change or improvement rather than recovery. Most data reported for the CORE-OM is from large clinical and non-clinical populations in the UK and both primary and secondary care settings providing psychological therapy to a wide range of clinical conditions including depression and anxiety.

High internal consistency is reported for domains and total items for several studies (Evans et al., 2002; Barkham et al., 2005; Lyne et al., 2006; Connell et al., 2007). Item-total correlations all exceed 0.5 (Lyne et al., 2006). Acceptable Test-retest reliability is reported (Evans et al., 2002).

Construct validity is supported in several studies with strong correlations between measures of symptoms and depression (BDI) (Leach et al., 2007). The problems domain has been reported to be more strongly correlated with symptom measures (BDI, BAI, BSI) (Barkham et al., 2001) and GHQ, Symptom Checklist-90 and the IIP-32 (Evans et al., 2002). Moderate correlation is reported between the patient-reported CORE-OM and the clinician rated Health of the Nation Outcome Scale (HoNOS) (Leach et al., 2005).

CORE-OM is able to detect change using pre-defined cut-off scores (Barkham et al., 2001). High sensitivity and specificity is reported using a cut-off point of 13 (Gilbody et al., 2007).

Responsiveness is reported with large effect-sizes for each domain (excluding the risk items) post computerized CBT for patients with depression and anxiety (Learmonth and Rai, 2008).

Normal and symmetrical distributions of scores have been reported for patients in primary and secondary settings (Barkham et al., 2005; Connell et al., 2007). Risk scores have been shown to be significantly higher in secondary care patients to those in primary care (Barkham et al., 2005).

It is acceptable to patients and clients (Barkham et al., 2001). High completion rates are reported for both primary and secondary care settings (Barkham et al., 2005; Lyne et al., 2006; Connell et al., 2007). However, lower response rates are reported for the postal version (Connell et al., 2007)

e). Work & Social Adjustment Scale (WSAS)

The WSAS is a self-report measure of functional impairment which has been applied both in people with mental disorders, including depression (Mundt et al., 2001), obsessive-compulsive disorder (Greist et al., 2002), phobic disorders (Mataix-Cols et al., 2005), and bipolar disorder (Fagiolini et al., 2005), and in those with primarily physical illnesses (McLaughlin et al., 2005; Hommel et al., 2008). It comprises five items: Work, Home management, Social leisure activities, Private leisure activities, and Family and relationships. Each item is rated on a scale of 0 (no impairment) to 8 (very severe impairment); selected items or the whole scale may be used. An item score of 3-5 is suggestive of moderate impairment; an overall score of 10 or above is associated with significant functional impairment, with scores above 20 indicating severe psychopathology (Mundt et al., 2001). However, Mataix-Cols et al., (2005) suggest a higher cutoff of 15 for those with phobic disorders.

The findings of both the principal evaluative studies identified (Mundt et al., 2001; Mataix-Cols et al., 2005) were similar. Factor analysis yielded a single disability factor. Internal consistency and test-retest reliability were reported as generally good. The WSAS discriminated different levels of disorder severity, and showed sensitivity to treatment-related change. Convergence with assessor ratings was variable in those with phobic disorders; it is suggested that the WSAS is best rated by both patient and assessor (Mataix-Cols et al., 2005).

There is some limited evidence of responsiveness. Mataix-Cols et al., (2005) found the WSAS scores to be sensitive to change over time in patients participating in RCTs of self-exposure therapy.

The simplicity of the WSAS has been commended (Mundt et al., 2001); it appears to be readily understood and accepted by patients, and is easily scored. The WSAS has been applied using innovative technology in the form of interactive voice response (IVR) systems for telephone administration, which are economical of staff time. Conventional ‘paper and pencil’ self-report versions and interview administration can also be used. Results appear to be unbiased by mode of administration.

f). General Anxiety Disorder-7 (GAD-7)

Items for the GAD-7 were selected from the DSM-IV symptom criteria for Generalised Anxiety Disorder. Responses are obtains regarding the frequency and degree of bother the patients experienced the symptoms in the past 2 weeks on a 3

point scale. Scores range from 0 to 21. Cut-off points are suggested for identifying cases of mild, moderate and severe anxiety (5, 10 and 15 respectively).

The GAD-7 has high internal consistency; is reproducible within patients and correlation of scores between patients self-report and mental health professionals administration is high (Spitzer et al., 2006).

High sensitivity and specificity is reported when using a cut-off score of 10. Increasing GAD scores are significantly associated with decreasing function measured by the SF-20. Higher correlation is reported with the GAD and the MCS than the PCS of the SF-20 as would be expected. Construct validity is supported with high correlation of scores from the GAD-7 and other measures of anxiety (BAI, Symptom Checklist-90) (Spitzer et al., 2006).

Comparable sensitivity and specificity has been reported for the GAD-7 (using a cut-off score of 8) and GAD-2 (using a cut-off score of 3). The GAD-2 includes the first 2 items form the GAD-7 representing core anxiety items and scores range from 0 to 6. Sensitivity and specificity for the GAD-7 is higher for GAD but reasonable performance has been reported for other anxiety related disorders (Kroenke et al., 2007).

g). Beck Anxiety Inventory (BAI)

The BAI was developed to meet the perceived need for a measure capable of differentiating anxiety and depression in psychiatric populations. The final version of the measure comprises 21 items in two dimensions, namely, somatic symptoms, and subjective anxiety and panic (cognitive) symptoms. Scores differentiate low and moderate anxiety, and 'cause for concern'.

High internal consistency reliability and moderate test-retest reliability (partly due to wide variation in the time interval used) were reported in a review by De Ayala and colleagues (2005). Reliability appeared to be dependent on the diagnostic classification of the sample, reliability estimates being lower in non-psychiatric populations (e.g. University students). Construct validity was also supported by moderate to strong correlations with similar measures (De Ayala et al., 2005).

Content and discriminant validity of the BAI have been questioned (Enns et al., 1998; Wetherell & Gatz, 2005). Enns & colleagues (1998) found that the BAI did not convincingly distinguish anxiety from depression in a clinically depressed sample, whilst Cox et al. (1996) suggest that it measures panic attack symptoms only, rather than anxiety in general. Wetherell & Gatz (2005) suggest that, in non-psychiatric persons, the somatic items from the BAI may tap into symptoms of physical illness, leading to overestimates of anxiety symptoms in medically ill samples. This is particularly problematic when assessing anxiety symptoms in older adults, who are more likely to have somatic symptoms arising from medical causes; it is suggested that an instrument focusing on cognitive rather than physiological aspects of anxiety is needed with this group (Wetherell & Gatz, 2005).

Population norms (US) have been reported and it was found that younger age-groups (aged 18-44 yrs) scored significantly higher on the BAI, which is consistent other anxiety measures (Gillis et al., 1995).

h) Preference based measures

The allocation of healthcare resources relies on the use of cost effectiveness analysis using QALYs which are constructed from generic health status measures such as the EQ-5D or the SF-6D system.

The EQ-5D assesses health across five domains: anxiety/depression (AD), mobility (M), pain/discomfort (PD), self-care (SC), and usual activities (UA). Each domain has one item and a three-point categorical response scale; health 'today' is assessed. Weights based upon societal valuations of health states are used to calculate an index score of –0.59 to 1.00, where –0.59 is a state worse than death and 1.00 is maximum well-being. A score profile can be reported. The EQ thermometer is a single 20 cm vertical visual analogue scales with a range of 0 to 100, where 0 is the worst and 100 the best imaginable health. The SF-6D utilizes application of a preference-based algorithm to SF-12 data to generate an index. Mann et al., (2008) published utility values derived from clinically diagnosed cases of depression from the UK based on reference case measures (SCID and PHQ), EQ-5D and SF-6D.

Moderate correlation of scores has been reported between SF-6D dimensions and the EQ-5D items. Patients responded differently to the related dimensions on both instruments with more patients reporting themselves as having no problems or limitations on the EQ-5D than the SF-6D (Mann et al., 2008). Data was drawn from a RCT evaluating the impact of a collaborative care intervention in the UK (n=114 patients). Significant improvement in utility values were observed at three month follow-up in this study for both SF-6D and EQ-5D following treatment for major depression indicating responsiveness. Both instruments were responsive to symptomatic improvement and demonstrated health gains at three months follow-up. The EQ-5D demonstrated larger health gains than SF-6D. Furthermore, an increase in severity of depression, measured by the PHQ was significantly related to lower utility values for both instruments and similar in magnitude of disutility for SF-6D and EQ-5D.

Brazier (2008) report a study to examine the impact of mental health condition on the SF-6D index and found that statistically significant decrements were estimated for generalized anxiety disorder and depressive episodes disorders. These exceeded the clinical minimally important difference of the SF-6D scale (0.04). Scores were significantly different for varying levels of anxiety measured using HAMA- with lower scores for higher anxiety levels (Revicki et al., 2008). SF-6D sensitive to varying anxiety severity.

Significantly lower index scores have been reported for patients with major depressive disorder and those with anxiety disorder compared to population norms suggesting discriminative properties (Supina et al., 2007). Brazier (2008) examined the differences in health state valuations made by the public and by patients valuing their own health using the EQ-5D. Patients gave higher weightings to the mental health

domains than physical components compared to the general public. Brazier suggests that generic measures may be adequate in depression and anxiety but not for psychotic or complex medical conditions. Scores for patients with depression were significantly lower in a European study suggesting discriminative validity (mean 0.44 Vs. 0.85 population norms) Garcia-Cebrian et al., (2008).

i). Center for Epidemiologic Studies Depression Scale (CES-D)

Primarily the CES-D was developed for use in epidemiological studies of the general population to measure the current level of depressive symptomatology focusing on the affective component, depressed mood (Radloff 1977 USA). The 20 item scale is constructed with four domains supported by factor analysis: Depressed affective; Positive affect; Somatic and retarded activity; Interpersonal with a four point Likert scale for the presence of symptoms. The range of scores is 0 to 60 with higher scores indicating more symptoms. A cut-off score of 16 is considered to represent case-ness and scores classified into mild, moderate and severe depression.

It appears acceptable to patients. The distribution of scores was as would be expected for general populations and for patients. High internal consistency and reproducibility have been reported by the developers. However, the time frame for responses is current symptoms, so the cyclic nature of depressive symptoms may bias results for test-retest reliability. Scores discriminate patients diagnosed as depressed by clinicians and those compared to population norms. Construct validity is reported with moderate correlations with other self-reported measures of depressive symptoms. The CES-D has been evaluated in different populations including older people, general population and patients with physical health conditions.

The instrument does not have suicide ideation items but two screeners (2 questions in each screener) are been developed to be used with the CES-D with a reference period of one week with a 4 point Likert scale. Scores greater than 5 are considered to represent a high risk of suicide (Garrison et al., 1991; Lewinshon et al., 1996)).

j). General Health Questionnaire (GHQ)

The GHQ was developed as a screening instrument to detect people likely to be at risk of developing mental health problems including depression, anxiety, somatic symptoms and social withdrawal (Goldberg 1978 UK). Different versions are available with 60, 30, 28 and 12 itemed questionnaires. The most commonly applied version is the 28 item. A four point Likert scale is used to obtain responses with a total and sub-domain scores computable. The GHQ 28 scores range from 0 to 84. Domains include: Somatic symptoms; Anxiety and insomnia; Social dysfunction; and Severe depression. Thresholds are suggested for each method of scoring and only relevant for screening purposes.

Although extensively examined as a screening instrument the GHQ has not been as well evaluated as an outcome measure, ie by assessing responsiveness.

k). Mental Health Recovery measures

The concept of recovery for people with mental health problems does not necessarily equate with the process of recovery for physical ill health. It is more about maintaining a meaningful life and staying in control of one's life and not necessarily the reduction of symptoms. Recovery is not a new concept to mental health and referred to in previous government policy statements specifically *The Journey to Recovery-The Government's Vision for Mental Healthcare (2001).* Since then, several instruments and research tools have been developed to specifically measure patient's perceptions of recovery. To note, these measures are not for specific mental health problems but globally for all people with mental health problems.

Of particular interest is reference to measures of Recovery such as DREEM, Mental Health Recovery Star, Ohio Consumer Assessment I&II and Wellness Recovery Action Plan. .

The Recovery Enhancing Environmental Measure (REE)[3] is a self-report instrument that gathers information/data about mental health recovery from people who receive mental health services. The REE asks people where they are in their process of recovery and what markers of recovery they are currently experiencing. They rate the importance of several elements such as hope, sense of meaning and wellness to their personal recovery and rate the performance of their mental health service on three activities associated with each of these elements. They also rate the service on factors in the system that promote resilience

The delivery of Services is now orientated to mental health recovery. It has been reported to help users of services focus on what their recovery needs were. It is considered to be a superior measurement system as it focuses on the individual's perception of their own recovery and not those perceived by healthcare professionals. This data is aimed at helping mental health services identify what they can do to shape services and support recovery.

It was developed as a research tool to help services assess their recovery-commitment either from the user perspective or the perspective of the supports that the services provide. It looks at personal recovery and at mental health services and organisational climate at the same time in the attempt to bridge the gap between service configuration and perceived care. It is conceived that DREEM data can differentiate services that perform well from those that don't.

It is a self-report instrument which can also be administered by interview. Completion has been reported to take between 20 minutes for some patients and up to one and a half hours for others. It can also be completed by others on behalf of the patients and also in a group setting. Postal administration had produced very poor response rates and the developers suggest this only be applicable for people with a high level of literacy and that an honorarium may increase responses. Staff directly involved in patients care or in the configuration of services should not use the DREEM with patients as the instrument in parts assesses the behaviour of staff and evaluation of the

[3] Also referred to as DREEM

service from the patient's perspective. It is also considered to require ethical consent to complete the DREEM.

Items are within the following dimensions: Stages of recovery, Elements of recovery with sub-scales within each dimension. Each scale is scored using a 5 point Likert scale of agreement. All items are stated positively. There are also open ended questions.

Despite the attraction of a comprehensive measurement model, there is insufficient data to support its application at present and the complexity of completion and scoring a challenge.

Discussion and conclusions

The focus of this review is upon PROMs used as measures on a population basis and as potential evidence of quality and outcomes of services, specifically for depression and anxiety. There are a large number of potential measures of mental health potentially relevant to depression and anxiety, some of which have substantial bodies of evidence in relation to their performance. However much of the evidence concerns related but distinct issues such as the role of questionnaires to screen for mental health problems or monitoring of outcomes in the context of individual patient care. Less consideration has been given to PROMs in the field of mental health as measures of quality and outcomes. Some general issues therefore need to be considered prior to high-lighting of specific PROMs that could be further explored as measures of quality and outcomes.

A first general issue that needs to be considered is whether to develop an approach to outcome assessment in relation to depression and anxiety separately. They are distinct symptoms and often relate to distinct syndromes and causes with different treatments indicated for the two disorders. There are condition-specific PROMs that could be used to support separate outcome assessment of anxiety and depression. However it is often argued that there is considerable overlap between the two symptoms and that, especially in a primary care context a single approach to overall management is more realistic and feasible. A more unified approach to these common psychiatric disorders would be more effectively supported by a unified approach to outcome measurement, ie using those PROMs that are considered equally appropriate to anxiety and depression.

A second general issue that needs to be considered in determining a strategy for use of PROMs in depression and anxiety is the extent to which the focus should be on symptoms and symptom reduction or on broader aspects of function and health-related quality of life. Symptoms play a major role in the presentation, detection and management of depression and anxiety in primary care. It is understandable that the majority of PROMs available in this area focus on symptoms. However it can be argued that patients' and professionals' goals are as much concerned with improving broader aspects of psychological and social function and that outcome measurement should take account of broader dimensions than symptom severity. PROMs-based outcome strategies do not have to be exclusively focused on symptoms or broader aspects of function. Some instruments attempt to capture both aspects. It is also

possible to employ a battery combining measures that together focus on the spectrum from symptoms to function.

A specific issue within the debate about priorities for mental health outcome measurement focuses on the concept of recovery. Recent critiques of mental health services have focused on the importance to patients of experiences such as hope and empowerment and the neglect of such goals by services. PROMs have begun to appear to address these even broader aspects of outcome than either symptoms or social function. An argument in favour of their consideration is that measures of recovery are said more directly to address issues of concern to service users, whereas most PROMs in mental health have tended to reflect professional concerns. However it is a complex judgement whether such dimensions of outcome should be core to a PROMs based strategy given that less evidence has accumulated of their measurement properties and feasibility.

A third consideration is whether a PROMs strategy for depression and anxiety should build on currently established uses of such PROMs in primary care. Although there are limited levels of enthusiasm for the use of PROMs in relation to the management of mental health problems in primary care, the use of specific PROMs in the context of QOF has at least been professionally accepted and the approved PROMs are widely used. The judgement is whether it would on grounds of pragmatism it would be more effective to build around those measures that have been professionally accepted and with which more healthcare professionals in primary care are likely to be relatively familiar.

A fourth and slightly more technical consideration needs to be recognised. A key requirement of PROMs as they have developed for use in other areas of healthcare outside of mental health has been that they should be responsive ie be shown to be sensitive to changes over time, and particularly in terms that matter to patients. It is much harder to be enthusiastic about PROMs lacking evidence of responsiveness if they typically are used to assess outcomes of services by administration before and after interventions with particular emphasis on change scores in PROMs between time points. Responsiveness would appear to be a less salient requirement of self complete questionnaires developed for use in mental health, probably because of the greater emphasis on using them as screening instruments to detect the presence or absence of mental disorders.

If PROMs for depression and anxiety are used in the form of cross sectional survey measures of population mental health, for example by annual survey, this could be achieved by use of PROMs which focus on binary estimation of the presence or absence of mental disorder. Several instruments have supportive evidence of reliability and validity in this context. Relative lack of evidence of sensitivity to change within individuals tracked over time (a major consideration for example for PROMs currently used in NHS for elective surgical procedures) may not be such a great concern if the emphasis of PROMs use is estimation of changes over time in the prevalence of depression and anxiety.

If a more condition-specific approach to PROMs for depression and anxiety is adopted, then it is clear that there are several candidate instruments. All three

instruments associated with QOF, PHQ-9, BDI-II and HADS, have supportive evidence of their measurement properties in relation to depression. There is no compelling evidence to favour one instrument over the others especially since most of the evidence focuses on their cross-sectional usefulness in screening or assessing the severity of depression rather than as an outcome measure. It is interesting that in the one direct head to head comparison of PHQ-9 and HADS, the instruments displayed very similar responsiveness (Cameron et al., 2008). Much of the debate has focused on the evidence that PHQ-9 and HADS have different thresholds for caseness, with more respondents identified as having problems of depression with PHQ-9 (Cameron et al., 2008; Kendrick et al., 2009). Of possible significance is the evidence from Kendrick et al.'s study (2009) that PHQ-9 was more frequently used by practices than the other instruments.

Similarly there is no compelling evidence to identify a preferred PROM for anxiety, comparing BAI and GAD-7. Even less evidence is available for use in large scale longitudinal follow-up of outcomes for this form of measure.

It may be that a more general and combined approach to depression and anxiety should be preferred reflecting the degree of overlap in the underlying features of the two conditions and their management in the primary care context. Two different types of measure have been considered; HADS has sub-scales for the two conditions and so could be applied as a single measure and CORE-OM is intended to be used across common mental health disorders in primary care. In this comparison, the evidence supporting CORE-OM is greater, with good evidence of measurement properties, especially responsiveness but also some positive evidence of acceptability and completion rates. It has the desirable feature of including content that goes beyond symptoms to include social function. It is of note that a recent review of mental health measures by the Centre for Health Economics (Jacobs 2009) concluded that CORE-OM seemed the most promising measure for use to assess productivity in relation to mental health services. One problem that needs to be considered is whether CORE-OM is applicable to the full range of primary care interventions; much of the evidence to date appears to have been generated in the context of counseling and psychotherapy.

If it is decided that broader aspects of social function should be included in a PROMs strategy for mental health, the WSAS is clearly promising with some supportive evidence and has an advantage of brevity in terms of likely response rates. However there is not an overwhelming body of supportive evidence and it is not clear whether how much additional information it would provide if used in a battery that included CORE-OM.

The main argument for use of instruments such as EQ-5D is that it provides a distinct form of evidence in terms of preferences. Given the much more widespread use of EQ-5D across a wide range of other healthcare interventions outside of mental health and given the evidence of satisfactory measurement properties of EQ-5D, it is to be preferred if a preference measure is required.

Finally thought has to be given to a number of recovery measures that have begun to emerge that address dimensions of outcome considered of particular importance to

users of services. It is quite clear that constructs such as hope and empowerment are not at all addressed in any of the instruments considered under other headings. Unfortunately there is insufficient evidence accumulated to date either about validation or feasibility and acceptability to high-light or recommend specific measures. Such evidence does need to be produced. In particular, if used in the context of large scale population-level monitoring of quality and outcomes evidence of responsiveness and acceptability would be particularly valuable.

Several recent initiatives indicate increased attention to PROMs as a resource in mental health. Recently, a compendium of outcome tools has being developed by the National Institute for Mental Health; many of the instruments identified and assessed by the project are PROMs. A project run by the Mental Health Research Network is currently evaluating users views of PROMs for mental health. Nevertheless the evidence is that health professionals are still not strongly convinced of the value of PROMs. Moreover there is evidence that response rates in follow-up of patients via PROMs are moderate. In addition to debate about most promising PROMs, consideration has to be given to the logistics of administering PROMs for mental health but also to ensuring that meaningful evidence is produced that can be used by all stakeholders concerned, healthcare professionals, users, service providers and commissioners. 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.

APPENDIX A

The IAPT MDS Outcome Measures intend to collect data from four domains:

- o Health and well being
- o Inclusion and employment
- o Choice and Access
- o Patient Experience

Table 1

IAPT Minimum Data Set (MDS) The MDS comprise of the following tools at different stages of assessment through the stepped approach.
-Generic Assessment Tools
PHQ9, GAD7 and Disorder Specific Measure *(if applicable)*
CORE-10, W&SAS and Inclusion & Employment Questionnaire
Patient Experience Questionnaire Part 1 (PEQ1)
-Routine Outcome Measuring Tools - Sessional
PHQ and GAD
Disorder specific measure
-Routine Outcome Measuring Tools – Review Sessions
PHQ9, GAD7and Disorder Specific Measure *(if applicable)*
CORE-10, W&SAS and Inclusion &Employment Questionnaire
Patient Experience Questionnaire Part 2 (PEQ2)
The SF-6D and/or EQ-5D are also referred to as optional measures

Table 1: **IAPT MDS recommended outcome measures**

Instrument	*Domains (no. items)*	*Response options*	*Score*	*Feasibility*
Patient Health Questionnaire (PHQ)	2, 8 and 15 item versions available Somatic symptoms including: depressed mood, anhedonia, appetite change, sleep disturbance, psychomotor agitation or retardation, loss of energy, feelings of worthlessness or guilt, diminished concentration and suicidal thoughts or attempts	Three point scale	Scores range from 0 to 27 with a three point scale for the 9 items. Mild depression is considered with scores of 5 to 9; moderate for scores between 10 and 14 and severe, 20 to 27 (	Three minutes completion time for PHQ-9 Free of charge
General Anxiety Disorder (GAD)-7	Symptoms (7)	3 point scale	.Scores range from 0 to 21. Cut-off points are suggested for identifying cases of mild, moderate and severe anxiety (5, 10 and 15 respectively).	Free
CORE-OM	34 items in four domains- Subjective well-being (4) Problems (12) Functioning (12) Risk (6) *Nine items within the instrument include symptoms suggested for the assessment of depression (NICE 2004a) and 7 items for anxiety* IAPT MDS refers to the CORE-OM 10	5 point Likert	Scores range from 0=not at all, to 4=most of the time. The minimum score is 0 and maximum 136 Males a cut-off score of 1.19 and females, 1.29 are suggested as clinically significant. A cut-off score of 2.5 is reported as 'severe'. Domain cut-off values are also reported. Descriptors Mild, Moderate and Severe and boundaries are suggested	Copyright is asserted but automatic license granted to anyone wishing to use the measure in an unaltered state without the intention of profiting financially from its use. Consent from clients is requested. Software packages are available to support and maximise benefit of using the system.

Table 2: IAPT MDS optional measures

Instrument	*Domains (no. items)*	*Response options*	*Score*	*Administration Completion time*
Beck Anxiety Inventory **(Beck et al., 1988)**	21 items, 2 factors: 14 somatic, 7 subjective anxiety/panic (*A*): 1. numbness or tingling 2. feeling hot 3. wobbliness in legs 4. unable to relax *A* 5. fear of the worst happening *A* 6. dizzy 7. heart pounding or racing 8. unsteady 9. terrified or afraid *A* 10. nervous *A* 11. feeling of choking 12. hands trembling 13. shaky 14. fear of losing control *A* 15. difficulty in breathing 16. fear of dying *A* 17. scared *A* 18. indigestion 19. faint 20. face flushed 21. hot/cold sweats	Respondents select a verbal rating (4 options, from 'not at all' to 'severely') and circle the corresponding numerical value, based on feelings	Items scored 0 (best) to 3 (worst) 0-21 = low anxiety 22-35 = moderate anxiety >36 = cause for concern Range: 0-63	Self- or interviewer-administered 5-10 minutes
Beck Depression Inventory-II **(Beck et al., 1996a)**	21 items in 2 subscales – Cognitive (*C*) and Non-cognitive *or* Somatic-Affective 1. sadness 2. pessimism *C* 3. past failure *C* 4. loss of pleasure (anhedonia) 5. guilty feelings *C* 6. punishment feelings *C* 7. self-dislike *C*	Respondents select a verbal rating (4 options) and circle the corresponding numerical value, based on feelings over past two weeks	0-13 minimal depression 14-19 mild depression 20-28 moderate depression 29-63 severe depression	Self- or interviewer-administered 5-10 minutes

	8. self-criticalness *C* 9. suicidal thoughts or wishes *C* 10. crying 11. agitation 12. loss of interest 13. indecisiveness 14. worthlessness *C* 15. loss of energy 16. change in sleeping 17. irritability 18. change in appetite 19. concentration difficulty 20. tiredness or fatigue 21. loss of interest in sex			
Beck Depression Inventory-FastScreen **(Beck et al., 1997b)**	7 items: -sadness -pessimism -past failure -loss of pleasure (anhedonia) -self-dislike -self-criticalness -suicidal thoughts or wishes *(i.e. items 1-4, 7-9 of BDI-II)*	As above	<4 not depressed >/=4 depressed Maximum score 21	<5 minutes
Hospital Anxiety and Depression Scale/HADS **(Zigmond & Snaith 1983)**	14 items divided equally between two subscales: anxiety, depression	Respondents select one of four verbal ratings (assigned a numerical value by scorer) based on feelings over past week	Items scored 0 (best) to 3 (worst) 0-7 = normal 8-10 = possible case >10 = probable case	Self- or interviewer-administered 2-6 minutes; scoring 1 minute

Other measures				
Instrument	***Domains (no. items)***	***Response options***	***Score***	***Administration Completion time***
Center for Epidemiologic Studies Depression Scale (CES-D)	20 items/4 domains Depressed affective Positive affect Somatic and retarded activity Interpersonal	Four point Likert scale for the presence of current symptoms.	The range of scores is 0 to 60 with higher scores indicating more symptoms. A cut-off score of 16 is considered to represent case-ness and scores classified into mild, moderate and severe depression.	
General Health Questionnaire (GHQ)	GHQ-28 (28 items/four domains) Somatic symptoms (7) Anxiety and insomnia (7) Social dysfunction (7) Severe depression (7)	Four point Likert scale with total and sub-domain scores computable.	The GHQ 28 scores range from 0 to 84 using Likert scaling or 0 to 28 using GHQ methods. Thresholds for identifying cases are suggested for each scoring method: Likert- scores greater than 24 GHQ scores- 4	

PHQ-2

A score of 3 or more indicates a positive screen and patients should be referred for a fuller assessment of the severity using the PHQ-9

Over the last 2 weeks, how often have you been bothered by any of the following problems? *(Use √ to indicate your answer)*	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Little interest or pleasure in doing things…….	**0**	**1**	**2**	**3**
2. Feeling down, depressed or hopeless……	**0**	**1**	**2**	**3**

PHQ-9

Over the last 2 weeks, how often have you been bothered by any of the following problems? *(Use √ to indicate your answer)*	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Little interest or pleasure in doing things…….	**0**	**1**	**2**	**3**
2. Feeling down, depressed or hopeless……	**0**	**1**	**2**	**3**
3. Trouble falling or staying asleep, or sleeping too much…….	**0**	**1**	**2**	**3**
4. Feeling tired or having little energy……..	**0**	**1**	**2**	**3**
5. Poor appetite or overeating…...	**0**	**1**	**2**	**3**
6. Feeling bad about yourself- or that you are a failure or have let yourself or your family down……	**0**	**1**	**2**	**3**
7. Trouble concentrating on things, such as reading the newspaper or watching television.	**0**	**1**	**2**	**3**
8. Moving or speaking so slowly that other people could have noticed? Or the opposite- being so fidgety or restless that you have been moving around a lot more than usual……	**0**	**1**	**2**	**3**
9. Thoughts that you would be better off dead or hurting yourself in some way……..	**0**	**1**	**2**	**3**

(**Total score**----------= ---------- + ---------- + -----)

If you checked off any problems, how difficult have these problems made it for you to do your work, take care of things at home, or get along with other people?

Not difficult at all	Somewhat difficult	Very difficult	Extremely difficult
☐	☐	☐	☐

Specific instructions are available for using the PHQ-9 diagnosis and score for treatment selection.

GAD-7

Over the last 2 weeks, how often have you been bothered by the following problems? *(Use √ to indicate your answer)*	**Not at all**	**Several days**	**More than half the days**	**Nearly every day**
1. Feeling nervous, anxious or on edge	**0**	**1**	**2**	**3**
2. Not being able to stop or control worrying	**0**	**1**	**2**	**3**
3. Worrying too much about different things	**0**	**1**	**2**	**3**
4. Trouble relaxing	**0**	**1**	**2**	**3**
5. Being so restless that it is hard to sit still	**0**	**1**	**2**	**3**
6. Becoming easily annoyed or irritable	**0**	**1**	**2**	**3**
7. Feeling afraid as if something awful might happen	**0**	**1**	**2**	**3**

GAD-7 Total score for the seven items ranges from 0 to 21. Scores of 5, 10 and 15 represent cut points for mild, moderate, and severe anxiety, respectively. The GAD-7 was primarily developed and a screening and severity measure for generalised anxiety disorder it has been applied in panic disorder, social anxiety disorder and post-traumatic stress disorder. The developers suggest that when screening for any anxiety disorder, a cut point for further evaluation is a score of 10.

CORE-OM 10

IMPORTANT - PLEASE READ THIS FIRST
This form has 10 statements about how you have been OVER THE LAST WEEK.
Please read each statement and think how often you felt that way last week.
Then tick the box which is closest to this.
Please use a dark pen (not pencil) and tick clearly within the boxes.

	Not at all	Only occasionally	Sometimes	Often Most or all of the time
1 I have felt tense, anxious or nervous				
2 I have felt I have someone to turn to for support when needed				
3 I have felt able to cope when things go wrong				
4 Talking to people has felt too much for me				
5 I have felt panic or terror				
6 I have made plans to end my life				
7 I have had difficulty getting to sleep or staying asleep				
8 I have felt despairing or hopeless				
9 I have felt unhappy				
10 Unwanted images or memories have been distressing me				

Total (Clinical Score*)
*** Procedure:** Add together the item scores, then divide by the number of questions completed to get the mean score, then multiply by 10 to get the Clinical Score.
Quick method for the CORE-10 (if all items completed): Add together the item scores to get the Clinical Score.

References

Ambrosini PJ, Metz C, Bianchi MD, Rabinovich H, Undie A. Concurrent validity and psychometric properties of the Beck Depression Inventory in outpatient adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* 1991; 30(1):51-57.

Ani C, Bazargan M, Hindman D, Bell D, Farooq MA, Akhanjee L *et al*. Depression symptomatology and diagnosis: discordance between patients and physicians in primary care settings. *BMC.Fam.Pract.* 2008;9:1.

Andrews B, Hejdenberg J, Wilding J. Student anxiety and depression: comparison of questionnaire and interview assessments. *Journal of Affective Disorders* 2006; 95(1-3):29-34.

Arnau RC, Meagher MW, Norris MP, Bramson R. Psychometric evaluation of the Beck Depression Inventory-II with primary care medical patients. *Health Psychology* 2001; 20(2):112-119.

Aylard PR, Gooding JH, McKenna PJ, Snaith RP. A validation study of three anxiety and depression self-assessment scales. *Journal of Psychosomatic Research* 1987; 31(2):261-268.

Barkham M, Gilbert N, Connell J, Marshall C, Twigg E. Suitability and utility of the CORE-OM and CORE-A for assessing severity of presenting problems in psychological therapy services based in primary and secondary care settings.*Br J Psychiatry.* 2005 Mar;186:239-46

Barkham M, Margison F, Leach C, Lucock M, Mellor-Clark J, Evans C, Benson L, Connell J, Audin K, McGrath G.Service profiling and outcomes benchmarking using the CORE-OM: toward practice-based evidence in the psychological therapies. Clinical Outcomes in Routine Evaluation-Outcome Measures. *J Consult Clin Psychol*. 2001 Apr;69(2):184-96.

Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Archives of General Psychiatry* 1961; 4:561-571.

Beck AT, Beck RW. Screening depressed patients in family practice. A rapid technic. *Postgraduate Medicine* 1972; 52(6):81-85.

Beck AT. Cognitive therapy and the emotional disorders. New York, USA: International Universities Press, 1976.

Beck AT, Rush AJ, Shaw BF, Emery G. Cognitive therapy of depression. New York, USA: Guilford Press, 1979.

Beck AT, Steer RA. Internal consistencies of the original and revised Beck Depression Inventory. *Journal of Clinical Psychology* 1984; 40(6):1365-1367.

Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. *Journal of Consulting and Clinical Psychology* 1988a; 56(6):893-897.

Beck AT, Steer RA, Garbin MG. Psychometric properties of the Beck Depression Inventory: 25 years of evaluation. *Clinical Psychology Review* 1988b; 8:77-100.

Beck AT, Steer RA. Relationship between the Beck Anxiety Inventory and the Hamilton Anxiety Rating Scale with anxious outpatients. *Journal of Anxiety Disorders* 1991; 5(3):213-223.

Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. San Antonio, TX, USA: Psychological Corporation, 1996.

Beck AT, Guth D, Steer RA, Ball R. Screening for major depression disorders in medical inpatients with the Beck Depression Inventory for Primary Care. *Behaviour Research and Therapy* 1997; 35(8):785-791.

Beck JS, Beck AT, Jolly J, Steer RA. Beck Youth Inventories - Second Edition for Children and Adolescents (BYI-II). 2005.

Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *Journal of Psychosomatic Research* 2002; 52(2):69-77.

Boyle GJ. Self-report measures of depression: some psychometric considerations. *British Journal of Clinical Psychology* 1985; 24(Pt 1):45-49.

Callahan EJ, Bertakis KD, Azari R, Helms LJ, Robbins J, Miller J. Depression in primary care: patient factors that influence recognition. *Family Medicine* 1997; 29(3):172-176.

Cameron M, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *British Journal of General Practice* 2008; 58(546):32-36

Cannon DS, Tiffany ST, Coon H, Scholand MB, McMahon WM, Leppert MF. The PHQ-9 as a brief assessment of lifetime major depression. *Psychol Assess* 2007 Jun;19(2):247-51

Clark DA, Steer RA, Beck AT. Common and specific dimensions of self-reported anxiety and depression: implications for the cognitive and tripartite models. *Journal of Abnormal Psychology* 1994; 103(4):645-654.

Clark LA, Watson D. Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology* 1991; 100(3):316-336.

Cox BJ, Cohen E, Direnfeld DM, Swinson RP. Does the Beck Anxiety Inventory measure anything beyond panic attack symptoms? *Behaviour Research and Therapy* 1996; 34(11-12):949-961.

Coyne JC. Self-reported distress: analog or Ersatz depression? *Psychological Bulletin* 1994; 116(1):29-45.

Connell J, Barkham M, Stiles WB, Twigg E, Singleton N, Evans O, Miles JN. Distribution of CORE-OM scores in a general population, clinical cut-off points and comparison with the CIS-R. *Br J Psychiatry*. 2007 Jan;190:69-74.

Crawford JR, Henry JD, Crombie C, Taylor EP. Normative data for the HADS from a large non-clinical sample. *British Journal of Clinical Psychology* 2001; 40(Pt.4):429-434.

Dahlstrom WG, Brooks JD, Peterson CD. The Beck Depression Inventory: item order and the impact of response sets. *Journal of Personality Assessment* 1990; 55(1-2):224-233.

Darzi A. Our NHS Our Future: NHS Next Stage Review Interim Report. Department of Health, London, October 2007.

Darzi A. Our NHS Our Future: High Quality Care for All. NHS Next Stage Review Final Report. Department of Health, London, June 2008.

Department of Health. Our Health, Our Care, Our Say: A New Direction for Community Services. White Paper, London 2006.

Department of Health. New Horizons: A new vision for mental health and well-being London 2009

Department of Health: National Service Framework: Mental Health White Paper, London 1999

Department of Health: Mental health information strategy March 2001

Department of Health The journey to recovery: the Government's vision for mental health care 2001

De Ayala RJ. Assessing the reliability of the Beck Anxiety Inventory scores. *Educational and Psychological Measurement* 2005.

Demyttenaere K, De Fruyt J. Getting what you ask for: on the selectivity of depression rating scales. *Psychotherapy & Psychosomatics* 2003; 72(2):61-70

Dietrich AJ, Oxman TE, Burns MR, Winchell CW, Chin T. Application of a depression management office system in community practice: a demonstration. *J.Am.Board Fam.Pract.* 2003;16:107-14.

Dowell AC, Biran LA. Problems in using the Hospital Anxiety and Depression Scale for screening patients in general practice. *British Journal of General Practice* 1990; 40(330):27-28.

Dowrick C, Leydon GM, McBride A, Howe A, Burgess H, Clarke P, Maisey S, Kendrick T Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ*. 2009 Mar 19;338:b663.

Dunbar M, Ford G, Hunt K, Der G. A confirmatory factor analysis of the Hospital Anxiety and Depression scale: comparing empirically and theoretically derived structures. *British Journal of Clinical Psychology* 2000; 39(Pt.1):79-94.

Dutton GR, Grothe KB, Jones GN, Whitehead D, Kendra K, Brantley PJ. Use of the Beck Depression Inventory-II with African American primary care patients. *General Hospital Psychiatry* 2004; 26(6):437-442.

Ell K, Unutzer J, Aranda M, Sanchez K, Lee PJ. Routine PHQ-9 depression screening in home health care: depression, prevalence, clinical and treatment characteristics and screening implementation. *Home.Health Care Serv.Q.* 2005;24:1-19.

Enns MW, Cox BJ, Parker JD, Guertin JE. Confirmatory factor analysis of the Beck Anxiety and Depression Inventories in patients with major depression. *Journal of Affective Disorders* 1998; 47(1-3):195-200.

Evans C, Connell J, Barkham M, Margison F, McGrath G, Mellor-Clark J, Audin K. Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *Br J Psychiatry*. 2002 Jan;180:51-60.

Fagiolini A, Kupfer DJ, Masalehdan A, Scott JA, Houck PR, Frank E. Functional impairment in the remission phase of bipolar disorder. *Bipolar Disord.* 2005 Jun;7(3):281-5.

Gary FA, Yarandi HN. Depression among southern rural African American women: a factor analysis of the Beck Depression Inventory-II. *Nursing Research* 2004; 53(4):251-259.

Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *Br.J.Gen.Pract.* 2007;57:650-2.

Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J.Gen.Intern.Med.* 2007;22:1596-602.

Goldberg DP. Manual of the General Health Questionnaire. Windsor, UK: NFER Publishing, 1978.

Headey BW, Kelley J, Wearing AJ. Dimensions of mental health: life satisfaction, positive affect, anxiety and depression. *Social Indicators Research* 1993; 29(1):63-82.

Herrmann C. International experiences with the Hospital Anxiety and Depression Scale: a review of validation data and clinical results. *Journal of Psychosomatic Research* 1997; 42(1):17-41.

Improving access to psychological therapies (IAPT) programme: Computerised cognitive behavioural therapy (cCBT) implementation guidance 2007

Kashani JH, Sherman DD, Parker DR, Reid JC. Utility of the Beck Depression Inventory with clinic-referred adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry* 1990; 29(2):278-282.

Kendrick T, Dowrick C, McBride A, Howe A, Clarke P, Maisey S, Moore M, Smith PW. Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. *BMJ.* 2009 Mar 19;338:b750

Kearns NP, Cruickshank CA, McGuigan KJ, Riley SA, Shaw SP, Snaith RP. A comparison of depression rating scales. *British Journal of Psychiatry* 1982; 141:45-49.

King DA, Heisel MJ, Lyness JM. Assessment and psychological treatment of depression in older adults with terminal or life-threatening illness. *Clinical Psychology Science and Practice* 2005;12:339-53.

Klein EW, Hunt JS, Leblanc BH. Depression Screening Interfaced With an Electronic Health Record: A Feasibility Study in a Primary Care Clinic Using Optical Mark Reader Technology. *Prim.Care Companion.J.Clin.Psychiatry* 2006;8:324-8.

Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J.Gen.Intern.Med.* 2001;16:606-13.

Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom.Med.* 2002;64:258-66.

Kroenke K, Bair M, Damush T, Hoke S, Nicholas G, Kempf C *et al*. Stepped Care for Affective Disorders and Musculoskeletal Pain (SCAMP) study: design and practical implications of an intervention for comorbid pain and depression. *Gen.Hosp.Psychiatry* 2007;29:506-17.

Kroenke K, Strine TW, Spitzer RL, Williams JB, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. *J.Affect.Disord.* 2008.

Kumar G, Steer RA, Beck AT. Factor structure of the Beck Anxiety Inventory with adolescent psychiatric inpatients. *Anxiety, Stress, and Coping* 1993; 6:125-131.

Kumar G, Steer RA, Teitelman KB, Villacis L. Effectiveness of Beck Depression Inventory-II subscales in screening for major depressive disorders in adolescent psychiatric inpatients. *Psychological Assessment* 2002; 9(2):164-170.

Lord Laylard (2006) A New Deal for Depression and Anxiety Disorders Report The Centre for Economic Performance's Mental Health Policy Group

Leach C, Lucock M, Barkham M, Stiles WB, Noble R, Iveson S Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *Br J Clin Psychol.* 2006 Jun;45(Pt 2):153-66.

Learmonth D, Rai S. Taking computerized CBT beyond primary care. *Br J Clin Psycho*l. 2008 Mar;47(Pt 1):111-8.

Lyne KJ, Barrett P, Evans C, Barkham M. Dimensions of variation on the CORE-OM. *Br J Clin Psychol.* 2006 Jun;45(Pt 2):185-203.

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.* 2008.

Martin CR. What does the Hospital Anxiety and Depression Scale (HADS) really measure in liaison psychiatry settings? *Current Psychiatry Reviews* 2005; 1(1):69-73.

Mataix-Cols D, Cowley AJ, Hankins M, Schneider A, Bachofen M, Kenwright M, Gega L, Cameron R, Marks IM. Reliability and validity of the work and social adjustment scale in phobic disorders. Compr Psychiatry. 2005 May-Jun;46(3):223-8.

McLaughlin TJ, Aupont O, Bambauer KZ, Stone P, Mullan MG, Colagiovanni J, Polishuk E, Johnstone M, Locke SE. Improving psychologic adjustment to chronic illness in cardiac patients. The role of depression and anxiety. *J Gen Intern Med.* 2005 Dec;20(12):1084-90.

Mitchell AJ, McGlinchey JB, Young D, Chelminski I, Zimmerman M. Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric out-patients: data from the MIDAS project. *Psychol.Med.* 2008;1-10.

Moorey S, Greer S, Watson M, Gorman C, Rowden L, Tunmore R et al. The factor structure and factor stability of the Hospital Anxiety and Depression Scale in patients with cancer. *British Journal of Psychiatry* 1991; 158:255-259.

NICE guidance 2004 Depression: management of depression in primary and secondary care

NICE guidance 2009 Depression: management of depression in primary and secondary care

Norris MP, Arnau RC, Bramson R, Meagher MW. The efficacy of somatic symptoms in assessing depression in older primary care patients. *Clinical Gerontologist* 2003; 27(1-2):43-57.

Outcomes compendium: helping you select the right tools for best mental health care practice in your field. National Institute for Mental Health in England, Barts and the London School of Medicine, and Department of Health

Parker G, Hilton T, Hadzi-Pavlovic D, Bains J. Screening for depression in the medically ill: the suggested utility of a cognitive-based approach. *Australian and New Zealand Journal of Psychiatry* 2001; 35(4):474-480.

Revicki DA, Brandenburg N, Matza L, Hornbrook MC, Feeny D. Health-related quality of life and utilities in primary-care patients with generalized anxiety disorder. *Qual.Life Res.* 2008.

Revicki DA, Matza LS, Flood E, Lloyd A. Bipolar disorder and health-related quality of life: review of burden of disease and clinical trials. *PharmacoEconomics* 2005;23:583-94.

Richter P, Werne J, Heerlein A, Kraus A, Sauer H. On the validity of the Beck Depression Inventory. A review. *Psychopathology* 1998; 31(3):160-168.

Rogers WH, Adler DA, Bungay KM, Wilson IB. Depression screening instruments made good severity measures in a cross-sectional analysis. *Journal of Clinical Epidemiology* 2005; 58(4):370-377.

Scheinthal SM, Steer RA, Giffin L, Beck AT. Evaluating geriatric medical outpatients with the Beck Depression Inventory-Fastscreen for medical patients. *Aging and Mental Health* 2001; 5(2):143-148.

Silverstone PH. Is anhedonia a good measure of depression? *Acta Psychiatrica Scandinavica* 1991; 83(4):249-250.

Snaith P. Measuring anxiety and depression. *Practitioner* 1993a; 237(1528):554-559.

Snaith P. What do depression rating scales measure? *British Journal of Psychiatry* 1993b; 163:293-298.

Snaith RP, Taylor CM. Rating scales for depression and anxiety: a current perspective. *British Journal of Clinical Pharmacology* 1985; 19(Suppl.1):17S-20S.

Snaith RP. The concepts of mild depression. *British Journal of Psychiatry* 1987; 150:387-393.

Snaith RP. The Hospital Anxiety and Depression Scale - comment. *British Journal of General Practice* 1990; 40(336):305.

Snaith RP. The Hospital Anxiety and Depression Scale. *Health and Quality of Life Outcomes* 2003; 1(29).

Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999;282:1737-44.

Spitzer RL, Williams JB, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD patient health questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *Am.J.Obstet.Gynecol.* 2000;183:759-69.

Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA: Journal of the American Medical Association* 1999;282:1737-44.

Spitzer RL, Williams JBW, Kroenke K, Hornyak R, McMurray J. Validity and utility of the PRIME-MD Patient Health Questionnaire in assessment of 3000 obstetric-gynecologic patients: the PRIME-MD Patient Health Questionnaire Obstetrics-Gynecology Study. *American Journal of Obstetrics and Gynecology* 2000;183:759-69.

Steer RA, Beck AT, Riskind JH, Brown G. Differentiation of depressive disorders from generalized anxiety by the Beck Depression Inventory. *Journal of Clinical Psychology* 1986; 42(3):475-478.

Steer RA, Beck AT, Riskind JH, Brown G. Relationships between the Beck Depression Inventory and the Hamilton Psychiatric Rating Scale for Depression in depressed outpatients. *Journal of Psychopathology and Behavioral Assessment* 1987; 9(3):327-339.

Steer RA, Rissmiller DJ, Ranieri WF, Beck AT. Structure of the computer-assisted Beck Anxiety Inventory with psychiatric inpatients. *Journal of Personality Assessment* 1993; 60(3):532-542.

Steer RA, Kumar G, Ranieri WF, Beck AT. Use of the Beck Anxiety Inventory with adolescent psychiatric outpatients. *Psychological Reports* 1995a; 76(2):459-465.

Steer RA, Clark DA, Beck AT, Ranieri WF. Common and specific dimensions of self-reported anxiety and depression: a replication. *Journal of Abnormal Psychology* 1995b; 104(3):542-545.

Steer RA, Ball R, Ranieri WF, Beck AT. Further evidence for the construct validity of the Beck Depression Inventory-II with psychiatric outpatients. *Psychological Reports* 1997; 80(2):443-446.

Steer RA, Clark DA, Beck AT, Ranieri WF. Common and specific dimensions of self-reported anxiety and depression: the BDI-II versus the BDI-IA. *Behaviour Research and Therapy* 1998a; 37(2):183-190.

Steer RA, Kumar G, Ranieri WF, Beck AT. Use of the Beck Depression Inventory-II with adolescent psychiatric outpatients. *Journal of Psychopathology and Behavioral Assessment* 1998b; 20(2):127-137.

Steer RA, Ball R, Ranieri WF, Beck AT. Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. *Journal of Clinical Psychology* 1999a; 55(1):117-128.

Steer RA, Cavalieri TA, Leonard DM, Beck AT. Use of the Beck Depression Inventory for Primary Care to screen for major depression disorders. *General Hospital Psychiatry* 1999b; 21(2):106-111.

Volk RJ, Pace TM, Parchman ML. Screening for depression in primary care patients: dimensionality of the short form of the Beck Depression Inventory. *Psychological Assessment* 1993; 5(2):173-181.

Wetherell JL, Gatz M. The Beck Anxiety Inventory in older adults with generalized anxiety disorder. *Journal of Psychopathology and Behavioral Assessment* 2005; 27(1):17-24.

Williams JWJ, Pignone M, Ramirez G, Perez-Stellato C. Identifying depression in primary care: a literature synthesis of case-finding instruments. *General Hospital Psychiatry* 2002; 24(4):225-237.

Winter LB, Steer RA, Jones-Hicks L, Beck AT. Screening for major depression disorders in adolescent medical outpatients with the Beck Depression Inventory for Primary Care. *Journal of Adolescent Health* 1999; 24(6):389-394.

Zigmond AS, Snaith RP. The Hospital Anxiety And Depression Scale. *Acta Psychiatrica Scandinavica* 1983; 67:361-370.