

RESEARCH ARTICLE

The CORE-10: A short measure of psychological distress for routine use in the psychological therapies

MICHAEL BARKHAM^{1*}, BRIDGETTE BEWICK², TRACY MULLIN², SIMON GILBODY³, JANICE CONNELL¹, JANE CAHILL², JOHN MELLOR-CLARK⁴, DAVID RICHARDS⁵, GISELA UNSWORTH⁶, & CHRIS EVANS⁷

¹Centre for Psychological Services Research, University of Sheffield, ²University of Leeds, ³University of York, ⁴CORE IMS, Rugby, ⁵University of Exeter, ⁶Kingston Hospital, and ⁷Nottingham NHS Trust & University of Nottingham, Nottingham, UK

Abstract

Background: There is a need for a generic, short, and easy-to-use assessment measure for common presentations of psychological distress in UK primary care mental health settings. This paper sets out the development of the CORE-10 in response to this need. *Method:* Items were drawn from the CORE-OM and 10 items were selected according to a combination of usefulness, coverage of item clusters, and statistical procedures. Three CORE-OM datasets were employed in the development phase: (1) a primary care sample, (2) a sample from an MRC platform trial of enhanced collaborative care of depression in primary care, and (3) a general population sample derived from the Office of National Statistics Psychiatric Morbidity Follow-up survey. A fourth dataset comprising a sample from an occupational health setting was used to evaluate the CORE-10 in its standalone format. *Results:* The internal reliability (alpha) of the CORE-10 was .90 and the score for the CORE-10 correlated with the CORE-OM at .94 in a clinical sample and .92 in a non-clinical sample. The clinical cut-off score for general psychological distress was 11.0 with a reliable change index (90% CI) of 6. For depression, the cut-off score for the CORE-10 was 13 and yielded sensitivity and specificity values of .92 (CI = .83–1.0) and 0.72 (CI = .60–.83) respectively. *Conclusion:* The CORE-10 is an acceptable and feasible instrument that has good psychometric properties and is practical to use with people presenting with common mental health problems in primary care settings.

Keywords: CORE-10; CORE-OM; outcome measures; routine practice; practice-based evidence

Background

There is a strong tradition of outcome measurement within healthcare generally, and outcome measurement systems are central in building a practice-based evidence for the psychological therapies (see Barkham, Hardy, & Mellor-Clark, 2010). Within this tradition, one measure that has been widely adopted by practitioners and services is the Clinical Outcomes in Routine Evaluation-Outcome Measure (CORE-OM; Barkham, Mellor-Clark, Connell, & Cahill, 2006; Evans et al., 2002). This is a 34-item generic measure of psychological distress tapping four domains: *subjective well-being* (4 items), *symptoms* (12

items comprising item clusters of depression, anxiety, physical, and trauma), *functioning* (12 items comprising item clusters of general functioning, social relationships, and close relationships), and *risk* (six items comprising item clusters of risk to self and risk to others). Although the CORE-OM has proved acceptable in routine practice settings, mostly used at the beginning and end of therapy, restrictions on time and resources result in a demand for very short measures where session-by-session change monitoring is required. In this context, there is a need for a quick and easy-to-use shortened form of the CORE-OM to assess the needs and change/outcomes of people presenting with a range of psychological difficulties.

*Corresponding author. Email: m.barkham@sheffield.ac.uk

Shortening measures raises the question of focus: although the CORE-OM is a generic measure, it correlates very highly and consistently with measures of depression, in particular the Beck Depression Inventory (BDI-I: Beck, Ward, Mendelson, Mock, & Erbaugh, 1961; BDI-II: Beck, Steer, & Brown, 1986; for details of mapping between CORE-OM and BDI, see Cahill et al., 2006; Evans et al., 2002; Leach et al., 2006). In addition, the CORE-OM also closely maps to the Patient Health Questionnaire (PHQ-9), an increasingly adopted measure of depression (Gilbody, Richards, & Barkham, 2007). Our aspiration in shortening the CORE-OM was to enable measurement of depression while retaining coverage of general psychological distress within a measure that could be used easily by practitioners and clients at screening as well as used for ongoing review during the course of therapy.

In this article, we report the development of this short version to these guiding principles, outline its key psychometric properties, and describe its application in practice settings. We have blended technical details with explanations of their meaning to help practitioners understand the strengths of the measure from both research and practice perspectives. The paper deviates somewhat from traditional headings as so much is condensed into one paper. The layout we have adopted is to summarise the data used, describe the strategy that selected 10 items from the 34 in the CORE-OM, and then to present psychometric properties of the CORE-10 both based on the items when embedded in the CORE-OM and from a dataset where the CORE-10 was used alone. Finally the discussion section covers issues about when and how it may be best to use the CORE-10. We believe that a grounded understanding of the design of the CORE-10 will lead to more informed use by practitioners, greater confidence in their arguing for the appropriate adoption of the measure, and to better practice-based evidence in the field of counselling and the psychological therapies. We have particularly underlined that all screening measures have a trade off between sensitivity (detecting correctly that someone has a problem) and specificity (minimising the number of false positives).

Datasets

We used three CORE-OM item datasets and one CORE-10 dataset to check the psychometric properties of the standalone instrument. All data were fully anonymised and approval given by NHS Ethics 05/Q1206/128.

CORE-OM dataset 1: Primary care – selection dataset

This dataset was used to select 10 items for the CORE-10. The data were drawn from 33 primary care services and comprised data from 6610 clients, 5831 of whom had completed a CORE-OM. Using SPSS 12.01 the data were split into three random groups with each sub-sample drawing from all 33 services. We used these three sub-samples respectively for the following purposes: (1) item selection ($n=1618$), (2) replication of selection ($n=1642$), and (3) establishing the psychometric properties of the items when embedded in the CORE-OM ($n=1835$). Table I shows the demographics and presenting problems of participants in each of the three sub-samples.

CORE-OM dataset 2: Medical Research Council Enhanced Care for Depression trial

This dataset was used to establish the screening properties of the selected 10 items. The data was obtained within a trial of depression that was able to resource a diagnostic interview with participants, thereby providing a benchmark against which the 10 items could be evaluated. Dataset 2 comprised 114 patients of whom 88 were female (77%). The mean age was 42.5 years ($SD=13.6$) ranging from 18–77 years. A total of 98 (86%) of the study population described their ethnic origin as White British. There was a range of educational attainment with 25% of the sample having no educational qualifications and 9% with a degree or higher qualification. For a full account of this sample, see Richards et al. (2008).

Dataset 3: Office of National Statistics dataset

This sample comprised 553 adults who returned valid CORE-OM forms in the follow-up to the psychiatric morbidity survey (Singleton, Bumpstead, & O'Brien, 2001). Because this dataset comprised data from clinical and non-clinical samples, this data was used in the calculation of cut-off scores. The data were weighted to be representative of a general population taking into account design factors, non-response and sampling procedures from the original psychiatric morbidity sample (see Connell et al., 2007). This resulted in an effective general population sample of 535 cases consisting of 268 males (50.2%) and 266 females (49.8%), with a mean age of 43.4 years ($SD=15.3$).

Table I. Descriptive statistics for demographics for the three CORE-OM sub-samples.

| Demographic variable | Sub-sample 1 ^a (N=1618) | | Sub-sample 2 ^a (N=1642) | | Sub-sample 3 ^a (N=1835) | |
|----------------------------|---------------------------------------|------|---------------------------------------|------|---------------------------------------|------|
| | n | % | n | % | n | % |
| Gender | | | | | | |
| Male | 457 | 28.2 | 478 | 29.2 | 515 | 28.1 |
| Female | 1115 | 71.4 | 1157 | 70.8 | 1317 | 71.8 |
| Ethnicity | | | | | | |
| Asian | 66 | 4.1 | 62 | 3.8 | 66 | 3.6 |
| Black African/Caribbean | 29 | 1.8 | 23 | 1.4 | 38 | 2.1 |
| White (European) | 1279 | 79.1 | 1295 | 78.9 | 1477 | 80.5 |
| Other | 24 | 1.5 | 16 | 1.0 | 25 | 1.4 |
| Presenting problems | | | | | | |
| Anxiety/stress | 1136 | 70.2 | 1138 | 69.3 | 1305 | 71.1 |
| Depression | 1042 | 64.4 | 1062 | 64.7 | 1198 | 65.3 |
| Interpersonal | 810 | 50.1 | 839 | 51.1 | 934 | 50.9 |
| Eating disorder | 47 | 2.9 | 46 | 2.8 | 38 | 2.1 |
| Personality problems | 65 | 4.0 | 73 | 4.5 | 82 | 4.5 |
| Psychosis | 11 | 0.7 | 11 | 0.7 | 13 | 0.7 |

^aTotal percentage may not equal 100% because demographics are not available for all participants.

Dataset 4: CORE-10 Occupational Health sample

This sample comprised a total of 77 participants who were referred to an occupational health service and completed the CORE-10 of whom 52 (68%) were women. Age was not recorded.

Rationale and structure of the CORE-10

The first objective was to select items to produce a shortened measure of acceptable brevity and format, with coverage of a range of problem intensity and good content coverage. The length chosen was 10 items to balance completion time against coverage but also to allow simplicity in scoring (see below). The choice of items was then driven by the following criteria: including two items each for depression and anxiety and one item each covering trauma and physical problems; choosing to have six high and four low intensity items to reduce floor and ceiling effects; aiming for coverage of all three sub-domains (social, general and close) of the functioning domain; opting to retain only one risk item tapping risk to self from the risk domain, and finally, ensuring that at least one item was retained from each of the four groups of high-low intensity and positive-negative keying (i.e. high intensity/positive; high intensity/negative; low intensity/positive; and low intensity/negative). All well-being items were dropped as they have always shown high correlation with the problem domain items, higher than the correlation between any other two domains (Evans

et al., 2002). This still left a considerable number of items in competition for retention.

Stage 1: Development of CORE-10

We carried out three distinct steps in identifying 10 items for the new measure. These steps are outlined here and in Table II in summary.

Step 1: Removal of items with lower response rates.

Rationale: Non-completion rates on the CORE-OM are low but it is even more important with a shorter measure that item non-completion be as low as possible so the first aim was to exclude items by non-completion rate.

Results: We omitted five items (from the domains of problems, and functioning) that had been identified in earlier reports as having lower completion rates by respondents: 19, 32, 30, 20, and 21 (Barkham, Gilbert, Connell, Marshall, & Twigg, 2005; Evans et al., 2002).

Step 2: Item coverage and mapping exercise.

Rationale: The map of the areas the items had to cover (see above) meant that an item being excluded for one reason might force selection of another if it was the only remaining item in a particular grouping. Table II presents the map of items and requirements and summarises item omissions and inclusions.

Table II. Implementation rationale for the CORE-10 item selection.

| Domain & cluster | No. of items | High intensity | | Low intensity | |
|-----------------------|--------------|------------------------|----------------------|-----------------------|---------------------------------------|
| | | Negatively keyed [5] | Positively keyed [1] | Negatively keyed [3] | Positively keyed [1] |
| Subjective well-being | 0 | x | x | x | x |
| Problem: Anxiety | 2 | [11 ³ , 15] | | 2, (20) ¹ | |
| Problem: Depression | 2 | [5 ³ , 23] | | 27, (30) ¹ | |
| Problem: Physical | 1 | | | [8 ³ , 18] | |
| Problem: Trauma | 1 | [13 ³ , 28] | | | |
| Functioning: Close | 1 | (1, 26) ² | | | 3, (19) ¹ |
| Functioning: General | 1 | | 7, (32) ¹ | | (12) ² , (21) ¹ |
| Functioning: Social | 1 | [33 ³ , 10] | | (25, 29) ² | |
| Risk: To self | 1 | [16, 34 ³] | | (9, 24) ² | |
| Risk: To others | 0 | x | x | x | x |

Items in bold indicated selected items arising from the three-stage process as follows:

x = Excluded domains

() = items omitted in stages 1 or 2; ¹ Items omitted due to most frequently omitted in completion; ² Items omitted due to mapping exercise
 [] = pairs of items considered in the regression analysis stage 3; ³ Items omitted due to regression analyses

Step 3: Regression analyses to optimise correlation with the full CORE-OM. Rationale: Elimination on content and on highest non-completion rates definitively selected in four items but left 12 candidate items in six pairs for the remaining six items needed. One pair was within the Low (negative) item group – *physical* (8 vs. 18) and the remaining five pairs within the High (negative) group of items: *high intensity anxiety* (11 vs. 15), *high intensity depression* (5 vs. 23), *trauma* (13 vs. 28), *social relationships* (10 vs. 33), and *risk to self* (16 vs. 34). We carried out correlation analyses on sub-sample 1, checked in sub-sample 2, in which each item in the pairs was correlated with its corresponding item cluster in the CORE-OM. This identified which item within each pair better predicted the score of the original item cluster on the CORE-OM. For each pair, we retained the item with the highest R value – that is, the item that best predicted the original cluster.

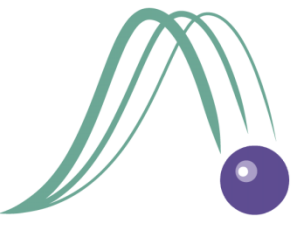
Results: We report here the R values calculated for samples 1 and 2 for the selected items from each pair only. These were as follows: *high intensity anxiety*, item 15 (R = .79, .81); *high intensity depression*, item 23 (R = .84, .83); *social functioning*, item 10 (R = .77, .75); *physical*, item 18 (.68, .66); *risk to self*, item 16 (R = .82, .82); and *trauma*, item 28 (R = .77, .75). These led to the final item selection shown in bold in Table II.

The procedures above produced the CORE-10 measure containing items 2, 3, 7, 10, 15, 16, 18, 23, 27, and 28 of the CORE-OM and fulfilling our design criteria. The CORE-10 is presented in Figure 1. The items appear in the order in which they appear in the CORE-OM.

Stage 2: Screening properties of CORE-10 items

Having selected the 10 items, we then evaluated their usefulness as a potential screening measure. We therefore used samples different from those used in Stage 1 but still samples of CORE-OM data. Accordingly we employed sub-sample 3 from dataset 1 (see Table I for details) as our clinical sample to determine the screening properties of the 10 selected items. In addition, we used dataset 2 comprising the MRC trial data (Richards et al., 2008), and also dataset 3 comprising a representative sample (N = 535) of the general population who participated in the Office of National Statistics (ONS) Psychiatric Morbidity Follow-up Survey (Singleton et al., 2001; for details, see Connell et al., 2007).

Screening properties. One established technique for assessing the ability of a measure to detect the presence or not of a specified presentation is to determine what is referred to as the Receiver Operator Characteristics (ROC) of the measure (Knottnerus, van Weel, & Muris, 2002). ROC curves represent an informative way of describing the inherent trade-offs between detecting the presence of a particular problem (sensitivity) and correctly identifying that the problem is not at diagnostic level (specificity). A measure could detect everyone presenting with a problem (high sensitivity) but do so by including almost everyone, including those without the problem (low specificity), thereby leading to many false positives. An alternative is to set a cutting point on a screening measure such that virtually no false positives (high specificity) are included but at the cost of reduced sensitivity – that



CORE - 10

Site ID

Client ID

Therapist ID numbers only (1) numbers only (2)

Sub codes

D D M M Y Y Y Y

Date form given

Age

Stage Completed

S Screening

R Referral

A Assessment

F First Therapy Session

P Pre-therapy (unspecified)

D During Therapy

L Last Therapy Session

X Follow up 1

Y Follow up 2

Male

Female

Stage

Episode

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.

| Over the last week | <i>Not at all</i> | <i>Only Occasionally</i> | <i>Sometimes</i> | <i>Often</i> | <i>Most or all the time</i> |
|---|----------------------------|----------------------------|----------------------------|----------------------------|-----------------------------|
| 1 I have felt tense, anxious or nervous | <input type="checkbox"/> 0 | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| 2 I have felt I have someone to turn to for support when needed | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 | <input type="checkbox"/> 0 |
| 3 I have felt able to cope when things go wrong | <input type="checkbox"/> 4 | <input type="checkbox"/> 3 | <input type="checkbox"/> 2 | <input type="checkbox"/> 1 | <input type="checkbox"/> 0 |
| 4 Talking to people has felt too much for me | <input type="checkbox"/> 0 | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| 5 I have felt panic or terror | <input type="checkbox"/> 0 | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| 6 I made plans to end my life | <input type="checkbox"/> 0 | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| 7 I have had difficulty getting to sleep or staying asleep | <input type="checkbox"/> 0 | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| 8 I have felt despairing or helpless | <input type="checkbox"/> 0 | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| 9 I have felt unhappy | <input type="checkbox"/> 0 | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |
| 10 Unwanted images or memories have been distressing me | <input type="checkbox"/> 0 | <input type="checkbox"/> 1 | <input type="checkbox"/> 2 | <input type="checkbox"/> 3 | <input type="checkbox"/> 4 |

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. If fewer than nine items completed, score should only be used very cautiously.
Quick method for the CORE-10 (if all items completed): Add together the item scores to get the Clinical Score.

THANK YOU FOR YOUR TIME IN COMPLETING THIS QUESTIONNAIRE

Figure 1. CORE-10.

Downloaded by [University of Sheffield] at 04:17 28 October 2013

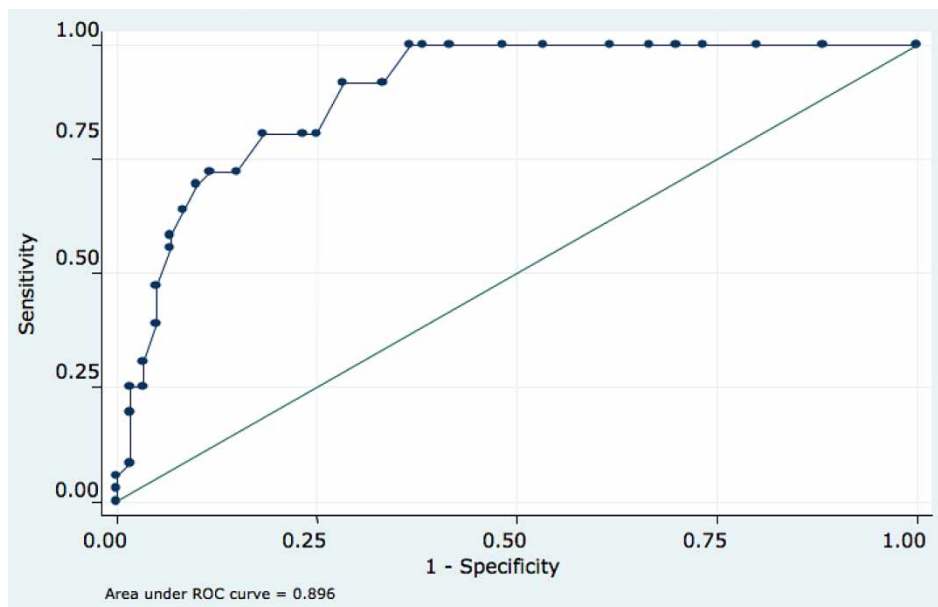


Figure 2. ROC curve of the CORE-10 against the SCID diagnosis of moderate depression.

is, more failures to detect the problem (false negatives). No measure is perfect and all will show a trade-off between sensitivity (decreasing) and specificity (increasing) as increasingly higher cut-off scores are chosen for use of the measure as a screening tool. As the purpose was to use the CORE-10 as a screening measure for quite common psychological distress, and as false positives should cause fairly low additional distress and be easily detected, we sought a higher sensitivity value at the expense of specificity. Designing a diagnostic measure would reverse these priorities – that is, would prioritise specificity over sensitivity – as does much screening for rare conditions where false positives will be distressing and costly to correct.

Receiver operating characteristic (ROC) analysis evaluates the screening accuracy of a continuous score on a screening test against a gold standard diagnosis. By expressing the relationship between sensitivity and specificity for each score, the ROC curve represents an index of the overall accuracy of a scale. The area under the curve (AUC, see Figure 2) is a measure of the information value inherent in a test to determine cases over the whole range of possible threshold values. An AUC of 0.5 indicates that the measure is useless to discriminate between cases and non-cases and an AUC 1.0 is obtained when the test gives perfect discrimination between cases and non-cases. Hence, the higher the value, the better the CORE-10 items are at capturing the phenomenon of depression as measured by the diagnostic standard.

Depression. We use dataset 2 (MRC ECD trial) because this dataset contained the CORE-10 items together with a gold standard clinical assessment of the presence or not of the DSM-IV moderate depression (diagnosed using the Structured Clinical Interview for DSM, SCID; Spitzer, Williams, Gibbon, & First, 1992). Sensitivity and specificity values are shown in Table III and Figure 2. The area under the curve (AUC) was 0.90 with a 95% confidence interval (CI) from 0.84–0.96. The 95% CI indicates how precisely the sample estimates the unknown population value, here fairly precisely. A cut-off score for depression of 13 on the CORE-10 yields sensitivity and specificity values of .92 (CI = .83–1.0) and 0.72 (CI = .60–.83) respectively. This means that using 13 as the cut-off will capture 92% of people presenting with depression (sensitivity) and exclude 72% of people without depression (specificity). However, differing requirements may be needed in different situations and Table 3 displays the sensitivity and specificity values of cutting scores from 13 to 21.

General psychological distress. As well as determining the screening properties of the CORE-10 for depression, we also considered its use to screen for general psychological distress. Lacking data with a gold standard measure of general psychological distress, we used a different approach to determine a sensible cutting point. The method used was Jacobson and Truax's (1991) formula 'c'. This sets a cutting point such that the sensitivity and specificity

Table III. Sensitivity and specificities for various cut points for depression for CORE-10 items.

| Cut point | Sensitivity (true positives) | Specificity (true negatives) |
|-----------|------------------------------|------------------------------|
| ≥ 13 | 0.92 | 0.72 |
| ≥ 14 | 0.81 | 0.75 |
| ≥ 15 | 0.81 | 0.77 |
| ≥ 16 | 0.81 | 0.82 |
| ≥ 17 | 0.72 | 0.85 |
| ≥ 18 | 0.72 | 0.88 |
| ≥ 19 | 0.69 | 0.90 |
| ≥ 20 | 0.64 | 0.92 |
| ≥ 21 | 0.58 | 0.93 |

are balanced when considering general psychological problems.

In order to determine the cut-off score to differentiate between membership of a clinical versus non-clinical population, we used the primary care clinical dataset that had not been used in the selection of the 10 items (i.e. sub-sample 3) together with the ONS dataset 3 as that comprised non-clinical/general population data. The cut-off value for the sample as a whole was 11.0 and the difference between men (10.6) and women (11.1) was less than one full point. A total of three clients out of 1835 scored a maximum of 40 representing 0.17% of the sample. In terms of categorising clients as being above or below the cut-off points on both the CORE-10 and CORE-OM, 84.5% in the clinical sample were above the cut-off on both measures, 10.3% were below the cut-off on both measures, and there was disagreement on 5.2% of cases. For this latter group, 2.4% of clients exceeded the cut-off score on the CORE-10 but scored below the cut-off on the CORE-OM while 2.8% of clients scored below the cut-off score on the CORE-10 but above the cut-off on the CORE-OM. Overall, the kappa agreement

(k) in terms of category assignment between the CORE-OM measure and CORE-10 items was .77.

Convergent validity: Correlation with CORE-OM. CORE-OM and CORE-10 scores were calculated for the clinical sample (sub sample 3) and the non-clinical sample (dataset 3). For the clinical sample the CORE-OM and CORE-10 scores correlated at $r = 0.94$ (CI: 0.93–0.95). For the non-clinical sample the two scores correlated at $r = 0.92$ (CI 0.91–0.93).

Convergent validity: Correlations with other outcome measures. We identified reports in which the CORE-OM had been correlated with other outcome measures to which we had access to the original data (Cahill et al., 2006; Connell et al., 2007; Evans et al., 2002; Gilbody et al., 2007). We recalculated the correlations using the CORE-10 items drawn from the CORE-OM. Table IV presents the results of the correlations compared with the original correlations using the CORE-OM.

Stage 3: Acceptability and psychometric properties of CORE-10 measure

Our final stage of work considered the acceptability and psychometric properties of the CORE-10 when administered as a measure in its own right. This is an important final stage as it uses the CORE-10 as a measure rather than as the items embedded within the CORE-OM. Our purpose here was to provide basic information on the acceptability and psychometric properties as a check on the measure.

Acceptability. A key indicator of acceptability is the readability of the measure. We determined this component by calculating the Flesch Ease of Reading

Table IV. Correlation coefficients between general, depression, and anxiety outcome measures with the CORE-OM and CORE-10 items.

| Measures | No. of items in measure | CORE-OM | CORE-10 items |
|--|-------------------------|---------|---------------|
| General | | | |
| Brief Symptom Inventory (BSI) ¹ | 53 | .81 | .75 |
| Symptom CheckList-90-R (SCL-90-R) ¹ | 90 | .88 | .81 |
| Clinical Interview Schedule-R (CIS-R) ³ | 56 | .77 | .74 |
| Depression | | | |
| Beck Depression Inventory (BDI) ¹ | 21 | .85 | .77 |
| Beck Depression Inventory-II (BDI-II) ¹ | 21 | .81 | .76 |
| Beck Depression Inventory-II (BDI-II) ² | 21 | .75 | .75 |
| Personal Health Questionnaire (PHQ-9) ⁴ | 9 | .63 | .56 |
| Anxiety | | | |
| Beck Anxiety Inventory (BAI) ¹ | 21 | .65 | .65 |

Note: ¹Evans et al. (2002); ²Cahill et al. (2006); ³Connell et al. (2007); ⁴Gilbody et al. (2007)

score that has a maximum score of 100 with a higher score indicative of easier reading. The score for the CORE-10 items was 79.9 (76.0 for the CORE-OM). In addition, we calculated the Flesch Kincaid Reading Grade, which establishes the reading age. The score was 4.0 (4.6 for the CORE-OM), which equates to a reading age around the fifth year of UK schooling.

Psychometrics. A total of 77 participants completed the CORE-10 of whom 52 (68%) were women. There were no missing items and scores ranged from 1 to 33. The mean CORE-10 score was 17.1 ($SD = 8.6$); median = 18 with quartiles at 10 and 24. The internal consistency of the CORE-10 was .90 (95% CI .86–.92) indicating that there is strong covariance across the ten items supporting the argument that it is not heavily contaminated by random variance (unreliability). For women the alpha was .89 (95% CI .84–.92) with a mean of 15.9 ($SD = 8.4$) while for men the alpha was .89 (95% CI .80–.94) and a mean of 19.6 ($SD = 8.2$). The overall alpha (.90) and SD (8.6) yielded a 95% reliable change index of 7.5 (rounded up to 8) and a 90% reliable change index of 6.3 (rounded down to 6). Box 1 presents a summary of the items and psychometrics for the CORE-10 measure.

Box 1. Summary information on the CORE-10

| | |
|--|---|
| <i>Items</i> | |
| Problems: | Depression (2 items) Anxiety (2 items) Physical problem Trauma |
| Functioning: | General functioning Social functioning Close relationships |
| Risk: | To self |
| <i>Acceptability</i> | |
| Flesch Ease of Reading score | 79.9 |
| Kincaid Grade score | 4.0 |
| <i>Convergent validity</i> | |
| CORE-10 and CORE-OM correlations: | |
| Clinical sample | .94 |
| Non-clinical sample | .92 |
| <i>Reliability</i> | |
| Reliability (alpha) | .90 |
| <i>Clinical and reliable change</i> | |
| General psychological distress cut-off score | 11 |
| Depression cut-off score | 13 |
| Reliable change index (90% CI) | 6 |

Discussion

This article has described the rationale, development, and the basic screening and psychometric properties of the CORE-10. The purpose of the CORE-10 is to assist busy practitioners in their decision-making when responding to people presenting with general psychological distress, particularly depression, in primary care settings. The results show that the psychometric properties of the CORE-10 are very satisfactory for screening so the CORE-10 can be a useful and very short tool for this purpose. The CORE-10 also shows good internal consistency, a fairly simple psychometric structure, and has broad coverage not only of depression and anxiety but of a wider spread of problems including general, social, and close relationships as well as risk to self. It has two positively keyed items that militate against a depressing mood induction or negative response set. The item breadth of the CORE-10 does result in the Reliable Change Index being greater than for those measures with either a tight focus (i.e. condition specific) or comprising many more items. However, as noted above, we recommend using an RCI based on a 90% false positive reliable change rate rather than a traditional 95% rate, which provides an RCI giving very similar rates of reliable change to those for the CORE-OM. The vast majority of the data reported here are based on analysis of the CORE-10 items embedded within the CORE-OM and we confidently predict that when large enough datasets of CORE-10 data are available, although all parameters will be statistically significantly slightly different from those found for items embedded in the CORE-OM, these differences will not be substantial nor impact on the use of the measure.

Scoring the CORE-10

One feature of 10 items is that simply summing the item scores, providing all items have been completed, yields a score that is consistent with the revised scoring method of the CORE-OM whereby the score range is 0–40 (see Barkham et al., 2006). If one item has been omitted, then the item scores would be summed, divided by the number of items completed (i.e. 9), and then multiplied by 10. In line with advice on all CORE instruments, prorating like this is not recommended if more than 10% of items have been omitted (i.e. if more than one item is omitted on the CORE-10). As an example, a CORE-10

form with nine items completed summing to 18 would result in 18 divided by 9 = 2, multiplied by 10 yielding a prorated score of 20. The purpose of the CORE-10 is to be used as a single score – there are no subscales. However, working with only 10 items makes identifying and using information from single items with clients easier.

Comparison with other measures

Our comparison of the CORE-10 items with the CORE-OM against a range of outcome measures showed slightly lower correlations, as would be expected with fewer items, but the reduced number of items did not appear to have an appreciable effect on the relationships. The correlations of the CORE-10 items with traditional global distress and depression measures all exceeded .70 while the correlations with an established anxiety measure and a short measure of depression – the Patient Health Questionnaire-9 (Kroenke, Spitzer, & Williams, 2001) – were lower. There has been an understandable growth in the promotion of shorter measures, exemplified by the adoption in the Improving Access to Psychological Therapies (IAPT) initiative of the PHQ-9 and a corresponding 7-item measure of generalised anxiety disorder, the GAD-7 (Spitzer, Kroenke, Williams, & Löwe, 2006). Unlike these unifocal, condition-specific measures, the CORE-10 draws on items that tap a broader range of activities, in particular, the area of functioning and relationships. Thus, where there is a single apparent diagnostic problem like depression, it captures the impacts of depression rather than just the symptoms.

Some supporting evidence regarding the difference between measuring the impacts rather than only the symptoms of depression and/or anxiety is suggested in a study of self-help computerised CBT for depression and anxiety (Cavanagh, Seccombe, & Lidbetter, 2011). Participants completed the CORE-10 together with the standard battery of IAPT measures: the PHQ-9 and GAD-7 (as reported above) and the Work and Social Adjustment Scale, which is a measure of functioning (WSAS; Mundt, Marks, Shear, & Greist, 2002). The authors reported pre-post effect sizes for these measures as follows: 0.9 (GAD-7), 0.8 (PHQ-9), 0.6 (CORE-10), and 0.4 (WSAS). That is, the focused measures on specific conditions showed the largest effect while both CORE-10 (comprising symptoms and functioning) and WSAS (functioning only), showed

change but less so. Similarly, the mean end-point scores for both the GAD-7 and PHQ-9 were below their respective clinical cut-off (i.e. indicating moving to recovery). By contrast, the CORE-10 mean score remained slightly above its cut-off score. One reading of these results is that the CORE-10, and likewise the CORE-OM, provide a more rounded view of the change achieved (or not) by someone and that the change captured by the symptom specific measures, if considered alone, may not reflect the state of other domains, especially social and functioning elements, of a person's life. Given the central importance of functioning in people's lives, further research is needed in this area.

Use of CORE-10 by practitioners and researchers

The CORE-10 has proved easy to use by individual practitioners using 'pen and paper', partly because multiplying by 10 to get a score is easier for most of us than multiplying by other numbers. That the items are all from the CORE-OM facilitates comparison with the longer and more comprehensive measure. In addition, as it is free to photocopy it provides minimal barriers to usage by individual practitioners. Though use of the paper version is easy and helpful, the CORE-10 supports much more sophisticated and extensive exploration if used within a well-designed computer system (e.g. CORE Net) enabling both individual but also service wide analysis of data (see Barkham et al., 2010). Change on the CORE-10 can be looked at against the clinical cutting point and the RCI for comparison with conventional 'initial-last' change measurements. Such a simple measure will enable practitioners to carry out simple hand plots of change while practitioner-researchers and pure researchers will be able to explore a variety of computer supported ways to explore change trajectories including a range of growth curve, multilevel modelling, and nearest neighbour methods.

The adoption of sessional use is rapidly becoming a norm via the IAPT initiative. The time frame of 'one week' makes the CORE-10 appropriate to be used weekly without confusion. By contrast the PHQ-9 and GAD-7 have a 14-day time frame that is likely to lead to some confusion if used weekly for clients and, in principal, will dampen measurement of change. Regardless of the measure, there needs to be further research into the impact on clients as well as the psychometrics of such repeated use (e.g. Durham et al., 2002). For example, Longwell and

Truax (2005) found that the weekly administration of the Beck Depression Inventory-II over a two-month period in a nonclinical population resulted in significantly lower estimates of depression. This same effect, however, was not observed with monthly or bi-monthly administrations of the BDI-II. CORE-IMS is collecting information about the CORE-10 being used in a repeated manner and we hope to report on this data in a future paper. Additionally, we are mindful of the need to establish the test-retest reliability of the CORE-10 within a clinical sample on a waiting list, akin to that carried out for the CORE-OM (Barkham, Mullin, Leach, Stiles, & Lucock, 2007).

In conclusion, the CORE-10 is ideally suited to help clinical decision-making in the assessment of clients presenting in a range of primary care settings. As such it has a specific purpose that differs from the CORE-OM, which, however, should remain the preferred measure for detailed evaluation at intake (i.e. assessment) and moderately wide time intervals across the course of delivering medium- and longer-term psychological therapies.

Acknowledgements

This research was initially supported by funding from the Artemis Trust and the NHS Priorities & Needs Levy via Leeds Mental Health & Teaching NHS Trust. We thank Howard Meltzer and colleagues at the Office for National Statistics for access to the ONS survey data and to Rachel Harker for her invaluable help.

References

- Barkham, M., Gilbert, N., Connell, J., Marshall, C., & Twigg, E. (2005). 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. *British Journal of Psychiatry*, *186*, 239–246.
- Barkham, M., Hardy, G.E., & Mellor-Clark, J. (Eds.) (2010). *Developing and delivering practice-based evidence: A guide for the psychological therapies*. Chichester: Wiley.
- Barkham, M., Mullin, T., Leach, C., Stiles, W.B., & Lucock, M. (2007). Stability of the CORE-OM and BDI-I: Psychometric properties and implications for routine practice. *Psychology & Psychotherapy: Theory, Research & Practice*, *80*, 269–278.
- Barkham, M., Mellor-Clark, J., Connell, J., & Cahill, J. (2006). A CORE approach to practice-based evidence: A brief history of the origins and applications of the CORE-OM and CORE System. *Counselling & Psychotherapy Research*, *6*, 3–15.
- Beck, A.T., Steer, R.A., & Brown, G.K. (1986). *Manual for the Beck Depression Inventory – Second Edition (BDI-II)*. San Antonio, TX: The Psychological Corporation.
- Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*, 561–571.
- Cahill, J., Barkham, M., Stiles, W.B., Twigg, E., Rees, A., Hardy, G.E., & Evans, C. (2006). Convergent validity of the CORE measures with measures of depression for clients in brief cognitive therapy for depression. *Journal of Counseling Psychology*, *53*, 253–259.
- Cavanagh, K., Secombe, N., & Lidbetter, N. (2011). The implementation of computerized cognitive behavioural therapies in a service user-led, third sector self help clinic. *Behavioural and Cognitive Psychotherapy*, *39*, 427–442.
- Connell, J., Barkham, M., Stiles, W.B., Twigg, E., Singleton, N., Evans, O., & Miles, J.N.V. (2007). Distribution of CORE-OM scores in a general population, clinical cut-off points, and comparison with the CIS-R. *British Journal of Psychiatry*, *190*, 69–74.
- Durham, C.J., McGrath, L.D., Burlingame, G.M., Schaalje, G.B., Lambert, M.J., & Davies, D.R. (2002). The effects of repeated administrations on self-report and parent-report scales. *Journal of Psychoeducational Assessment*, *20*, 240–257.
- Evans, C., Connell, J., Barkham, M., Margison, F., Mellor-Clark, J., McGrath, G., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, *180*, 51–60.
- Gilbody, S., Richards, D.A., & Barkham, M. (2007). Diagnosing depression in primary care using self-completed instruments: A UK validation of the PHQ-9 and CORE-OM. *British Journal of General Practice*, *57*, 650–652.
- Jacobson, N.S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*, 12–19.
- Knottnerus, J.A., van Weel, C., & Muris, J.W. (2002). Evaluation of diagnostic procedures. *BMJ*, *324*, 477–480.
- Kroenke, K., Spitzer, R.L., & Williams, J.B. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, *16*, 606–613.
- Leach, C., Lucock, M., Barkham, M., Stiles, W.B., Noble, R., & Iveson, S. (2006). Transforming between Beck Depression Inventory and CORE-OM scores in routine clinical practice. *British Journal of Clinical Psychology*, *45*, 153–166.
- Longwell, B.T., & Truax, P. (2005). The differential effects of weekly, monthly, and bimonthly administrations of the Beck Depression Inventory-II: Psychometric properties and clinical implications. *Behavior Therapy*, *36*, 265–275.
- Mundt, J.C., Marks, I.M., Shear, K., & Greist, J.M. (2002). The Work and Social Adjustment Scale: A simple measure of impairment in functioning. *British Journal of Psychiatry*, *180*, 461–464.
- Richards, D.A., Lovell, K., Gilbody, S., Gask, L., Torgerson, D., Barkham, M., Bower, P., Bland, M., Lankshear, A.J., Simpson, A., Fletcher, J., Escott, D., Hennesey, S., & Richardson, R. (2008). Collaborative care for depression: A randomized controlled trial. *Psychological Medicine*, *38*, 279–287.

- Singleton, N., Bumpstead, R., O'Brien, M., Lee, A., & Meltzer, H. (2001). *Psychiatric morbidity among adults living in private households, 2000*. London: The Stationary Office.
- Spitzer, R.L., Kroenke, K., Williams, J.B., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, 166, 1092–1097.
- Spitzer, R.L., Williams, J.B., Gibbon, M., & First, M.B. (1992). The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Archives of General Psychiatry*, 49, 624–629.

Biographies

Michael Barkham is professor of clinical psychology and director of the Centre for Psychological Services Research, University of Sheffield.

Bridgette Bewick is senior lecturer in health research, Leeds Institute of Health Sciences, University of Leeds.

Tracy Mullin was research officer at the Psychological Therapies Research Centre, University of Leeds.

Simon Gilbody is professor of psychological medicine and health services research at the University of York.

Janice Connell is research associate at the Centre for Psychological Services Research, University of Sheffield.

Jane Cahill is senior research fellow at the School of Healthcare, University of Leeds.

John Mellor-Clark is managing director of CORE IMS, Rugby.

Gisela Unsworth is head of the Psychological Well Being Service at Kingston Hospital NHS Trust, Kingston upon Thames.

David Richards is professor of mental health services research at the University of Exeter.

Chris Evans is a consultant in medical psychotherapy, Nottingham NHS Trust and Professor, University of Nottingham.