

## Distribution of CORE–OM scores in a general population, clinical cut-off points and comparison with the CIS–R

JANICE CONNELL, MICHAEL BARKHAM, WILLIAM B. STILES, ELSPETH TWIGG, NICOLA SINGLETON, OLGA EVANS and JEREMY N. V. MILES

**Background** Although measures of psychopathology are designed for use in clinical populations, their meaning derives from comparison with normal populations.

**Aims** To compare the distribution of scores on the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE–OM) from a general population sample with the distribution in an aggregated clinical sample to derive recommended cut-off points for determining clinical significance.

**Method** The CORE–OM general population sample was based on a weighted subsample of participants in the psychiatric morbidity follow-up survey who completed valid CORE–OM forms following their interview (effective  $n=535$ ).

**Results** Comparison of the CORE–OM general population sample with a clinical sample aggregated from previous studies ( $n=10\,761$ ) yielded a cut-off score of 9.9 on the 0–40 scale of the CORE–OM. The CORE–OM was highly correlated ( $r=0.77$ ) with the Clinical Interview Schedule – Revised, supporting convergent validity.

**Conclusions** We recommend rounding the CORE–OM cut-off score to 10. However, cut-off scores must be used thoughtfully and adjusted to fit context and purpose.

**Declaration of interest** M.B. was funded by the Mental Health Foundation to develop the CORE–OM.

Although measures of psychological distress and psychopathology are designed for use in clinical populations, their meaning derives from comparison with a normal population. In the UK, the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE–OM) is one of the most widely used outcome measures for psychological therapies (Barkham *et al*, 2001; Evans *et al*, 2002) and has been used in primary and secondary care settings (e.g. Barkham *et al*, 2005). To assess the distribution of CORE–OM scores in a general population, we used data from the follow-up (Singleton & Lewis, 2003) to the psychiatric morbidity survey carried out in the UK in 2000 (Singleton *et al*, 2001), in which a sample completed the Clinical Interview Schedule – Revised (CIS–R; Lewis *et al*, 1992) and the CORE–OM.

The study aimed, first, to assess the internal consistency, normative values and acceptability of the CORE–OM in a general population; second, to examine the convergent validity of the CORE–OM with the CIS–R; and third, to determine appropriate cut-off values on the CORE–OM. Cut-off values contribute to both research and clinical practice by indicating a respondent's membership in the normal or clinical population. This is useful both for initial screening and for assessing whether an intervention has brought about clinically significant change.

### METHOD

#### Participants

A general population sample was obtained from the follow-up to the psychiatric morbidity survey in which 8580 adults aged 18–74 years were interviewed between March and September 2000 (Singleton *et al*, 2001). This sample had been randomly selected from people living in private households in Great Britain and stratified by National Health

Service (NHS) region and socio-economic conditions.

From the original survey sample, 3536 respondents were selected for re-interview approximately 18 months later. This follow-up sample was designed to include all people from the initial sample who scored 12 or more on the CIS–R (indicating the presence of mental disorder), all people who scored 6–11 on the CIS–R (indicating no disorder but who reported some symptoms of common mental disorder) and a random sample of 20% of respondents who scored 0–5 on the CIS–R (indicating no disorder). This differential sampling was compensated for in the analysis by weighting procedures, described below. A more detailed description of the follow-up survey and sampling methods is given by Singleton & Lewis (2003).

The follow-up interview included a second administration of the CIS–R. Of the 2406 respondents to the follow-up survey, the 2048 interviewed during the last 2 months of the survey were randomly allocated to complete one of three self-report paper measures of psychological well-being. Of these individuals, 682 were allocated to complete the CORE–OM and the remainder were allocated to complete other measures.

Of the 682 interviewees allocated to the CORE–OM, 558 returned questionnaires (511 immediately after the follow-up interview, 47 later by mail). Of those who completed the interview, only 5 refused to complete the CORE–OM and 9 were judged incapable of completing it; 32 agreed to return the form by mail but failed to do so. In 78 cases interviewers indicated that the CORE–OM had been completed at the time of the interview but the forms were missing – possibly because interviewers failed to return them, or because they were lost in the mail or other misadventure. Of the 558 returned forms, 5 were considered invalid because of missing data on more than three items. The resulting general population sample thus included 553 respondents with a valid CORE–OM.

#### Characteristics of the general population sample

The general population sample included 238 men (43.0%) and 315 women (57.0%), with a mean age of 44.3 years (*s.d.*=14.3); 527 (95.3%) were White; 288 (52.1%) were either married or cohabiting, 137 (24.8%) were single, 92 (16.6%) were

divorced or separated and 36 (6.5%) were widowed. As their highest qualification 94 (17.0%) had a university degree, 54 (9.8%) had a specialist qualification, 73 (13.2%) had A-levels, 195 (35.3%) had a General Certificate of Secondary Education (GCSE) or equivalent and 136 (24.6%) had no qualification; 356 (64.4%) were employed, 11 (2.0%) were unemployed and 186 (33.6%) were economically inactive.

#### Non-distressed subsample

A 'non-distressed' subsample was derived from the general population sample as follows. Beginning with the 300 respondents who completed a valid CORE-OM and scored 0–5 (indicating no mental disorder) on the follow-up CIS-R, additional screening was undertaken using responses to questions and measures in the follow-up survey (Singleton & Lewis, 2003). Respondents were excluded from the subsample if they had visited a general practitioner in the past year or had been an in-patient or out-patient in the previous 3 months for either a mental or physical disorder, were receiving psychotropic medication, were undertaking counselling or had had suicidal thoughts in the past year. Respondents were also omitted if they scored below 50 on the mental health score of the 12-item Short Form Health Survey (SF-12; Ware *et al*, 1996) and hence were regarded to be of below-average mental health. The resulting asymptomatic or non-distressed sample population comprised 85 respondents: 41 men (48%) and 44 women (52%) with a mean age of 43.8 years (s.d.=14.2).

#### Weighting and data analysis

As described by Singleton and colleagues (Singleton *et al*, 2002; Singleton & Lewis, 2003) data from survey participants who completed one of the three paper measures were weighted in several steps to take account of design factors and non-response in both the original psychiatric morbidity sample and the subsequent follow-up sample. Respondents' scores were weighted to adjust for the follow-up survey's differential selection of people. As noted earlier, by design only 20% of those scoring 0–5 on the CIS-R were selected, in comparison with 100% of those scoring 6 or higher. To compensate, respondents with a score of 0–5 (in the original survey) were given a weighting of 5, and those scoring 6 or higher a weighting of 1.

Non-response was adjusted by applying corrections for underrepresented demographic groups (age, gender, marital status, household size) and geographical groups (regional, urban/rural): that is, respondents representing undersampled groups or characteristics were given proportionally higher weights. The final weight for each participant was the product of the weights applied in each step. This weight was then scaled back to the actual size of the sample allocated to the three paper measures (i.e.  $n=2048$ ). Analyses on weighted data were done using Stata version 8 for Windows, applying the survey data commands designed for use with weighted data from complex sample surveys.

These weighting procedures yielded an effective general population sample of 660 who were allocated to complete the CORE-OM. An effective sample of 543 returned CORE-OM forms, of which an effective 535 were valid. This effective general population sample consisted of 268 men (50.2%) and 266 women (49.8%) with a mean age of 43.4 years (s.d.=15.3). The effective size of the non-distressed sample was 118, including 60 men (50.8%) and 58 women (49.2%) with a mean age of 44.5 years (s.d.=14.8). All effective sample sizes have been rounded to the nearest whole number. Effective sample sizes differ from the actual numbers of valid forms because the weights were scaled back to the number of respondents allocated to all three paper measures ( $n=2048$ ), rather than to the number of valid CORE-OM forms.

#### Clinical samples used for comparison

For comparison with the general population sample we used clinical data from four previously documented samples drawn from the following services:

- (a) primary care counselling ( $n=6610$ ; Evans *et al*, 2003; Barkham *et al*, 2005);
- (b) clinical psychology and psychotherapy services in secondary care settings ( $n=2311$ ; Barkham *et al*, 2005);
- (c) generic secondary care departments in out-patients and community settings, principally clinical psychology, psychiatry, counselling, psychotherapy, nursing and art therapy, NHS secondary care psychology and psychotherapy services ( $n=2710$ ; Barkham *et al*, 2001);

- (d) psychotherapy, psychology, primary care and student counselling ( $n=890$ ; Evans *et al*, 2002).

There was some overlap between samples (c) and (d), which accounted for 13.5% of the joint sample, and the individuals involved were counted only once. This resulted in a total clinical sample of 10 761 persons. Of these, 3419 were men (32%) and 7326 were women (68%); gender information was missing for 16 (0.1%). Their mean age was 37.7 years (s.d.=12.5).

#### Measures

##### Clinical Interview Schedule – Revised

The CIS-R (Lewis *et al*, 1992) is a standardised interview for assessing common psychiatric disorders and is designed to be administered by non-clinicians. It comprises 14 sections covering areas of neurotic symptoms: somatic symptoms, fatigue, concentration and forgetfulness, sleep problems, irritability, worry about physical health, depression, depressive ideas, worry, anxiety, phobias, panic, compulsions and obsessions. Each section has a lead-in question relating to symptoms experienced over the previous month; the response to this question is not included in the scoring. A positive response to the initial question leads to four further questions (five for depressive symptoms) relating to the frequency, duration and severity of the symptom over the past 7 days. Each positive response scores 1; thus, for each section, scores range from 0 to 4 (or 0 to 5 for depressive ideas). The total score is the sum of all 14 sections, giving a possible range of 0–57. A score of 12 or above on the CIS-R indicates caseness (Lewis *et al*, 1992; Singleton & Lewis, 2003), a score of 6–11 indicates some symptoms of mental disorder and a score of 0–5 indicates little evidence of mental disorder (Singleton & Lewis, 2003).

##### Clinical Outcomes in Routine Evaluation – Outcome Measure

The CORE-OM (Barkham *et al*, 2001, 2005; Evans *et al*, 2002) is a 34-item self-report measure designed to assess level of psychological distress and outcome of psychological therapies. The 34 items comprise four domains (with each domain comprising specific clusters): specific problems (depression, anxiety, physical problems, trauma), functioning (general day-to-day functioning, close relationships, social

relationships); subjective well-being (feelings about self and optimism about the future); and risk (risk to self, risk to others). Each domain contains equal numbers of high and low intensity/severity items to offset possible floor and ceiling effects. All items are scored on a five-point scale from 0 to 4 (anchored ‘all or most of the time’ ‘not at all’, ‘only occasionally’, ‘often’ and ‘sometimes’) and relate to the previous week. Clinical scores are calculated as the mean of all completed items on the form, which are then multiplied by 10, so that clinically meaningful differences are expressed in whole numbers. Thus, scores may range from 0 to 40 (see Leach *et al*, 2006). Forms with three or fewer items missing are considered reliable, with scores based on completed items. The internal consistency of the CORE-OM has been reported as  $\alpha=0.94$  and the 1-week test-retest reliability as Spearman’s  $\rho=0.90$  (Evans *et al*, 2002).

**RESULTS**

**Acceptability and internal consistency**

Aspects of acceptability of the CORE-OM in the general population were examined by means of completion rates and number of missing items, since a fundamental requirement of a measure is that respondents agree and are able to complete it. Fewer than 2% of those who completed the follow-up interview refused to complete the CORE-OM or were deemed incapable of completing it, although among the minority of respondents who promised to return their form by mail, 40% failed to do so.

All of the respondents who returned invalid CORE-OM forms (missing more than three items) failed to complete the 20 items on the reverse side of the form, which suggests that they neglected to turn over the page. The mean omission rate on all items across the respondent group as a whole was 1.4%. When those who did not complete the second page were disregarded, this was reduced to 0.4%. Of these, the most commonly missed items were item 12 ‘I have been happy with the things I have done’ (1.3%); item 4 ‘I have felt OK about myself’ (1.1%); item 20 ‘My problems have been impossible to put to one side’ (0.9%); and item 9 ‘I have thought of hurting myself’ (0.9%).

The internal consistency, calculated using Cronbach’s ( $\alpha$ ) coefficient (Cronbach, 1951) was 0.91 (effective  $n=535$ ) in the general population sample.

**Distributions of CORE-OM clinical scores**

The distributions of CORE-OM clinical scores in the three samples (Table 1) are shown in Fig. 1. The mean CORE-OM clinical score for the aggregated clinical sample (total  $n=10\,761$ ) was 18.3 (s.d.=7.1). The women’s scores (mean 18.6, s.d.=6.9) on average were slightly higher than the men’s (mean 17.9, s.d.=7.3) in the aggregate sample ( $P<0.001$ ; confidence interval for the difference 0.4–1.0). The negative correlation with age was small but statistically significant:  $r=-0.10$ ;  $P<0.001$ . The mean CORE-OM clinical score for the general population sample (effective  $n=535$ ) was 4.8 (s.d.=4.3). There was no statistically significant difference between men (mean 4.9, s.d.=4.1) and women (mean 4.8, s.d.=4.5), and no statistically significant association with age ( $r=0.02$ ;  $P=0.63$ ).

The mean CORE-OM clinical score for the non-distressed sample (effective  $n=118$ ) was 2.5 (s.d.=1.8). Women’s scores (mean 2.2, s.d.=1.4) on average were slightly lower than men’s (mean 2.9, s.d.=2.0) in the non-distressed sample ( $P=0.04$ ); the correlation with age was  $r=-0.16$  ( $P=0.08$ ).

As would be expected, the CORE-OM clinical scores for the general population and non-distressed samples were highly skewed (see Table 1 and Fig. 1), with 54.8% of the general population sample and 83.2% of the non-distressed sample scoring below 4 out of a maximum of 40. The clinical population scores were more normally distributed.

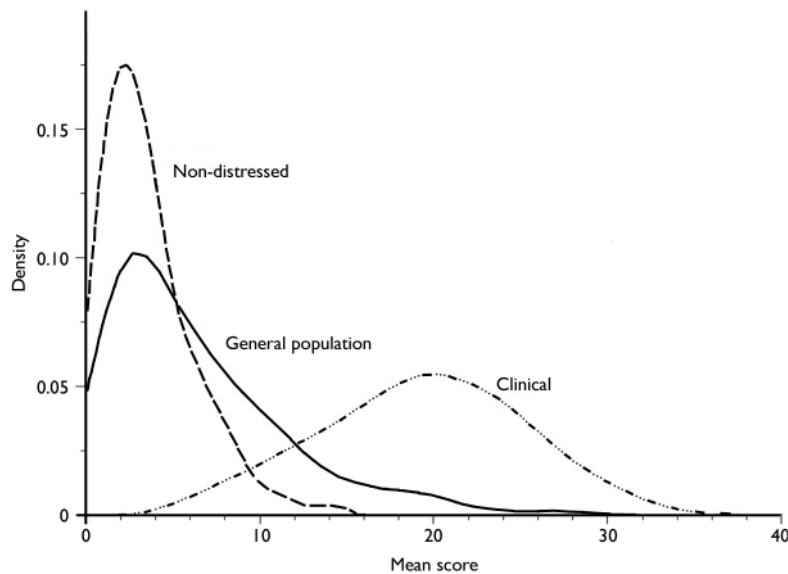
**Convergence of the CORE-OM with the CIS-R in the general population**

CORE-OM clinical scores were strongly correlated with the CIS-R total scores obtained in the follow-up interviews:  $r=0.77$ ,  $P<0.001$ , effective  $n=535$  in the general population sample. Table 2 presents mean CORE-OM clinical scores for

**Table 1** Distribution of CORE-OM clinical scores in clinical, general population and non-distressed samples

CORE-OM interval	Clinical sample			General population sample			Non-distressed sample		
	<i>n</i>	%	Cum %	Effective <i>n</i>	%	Cum %	Effective <i>n</i>	%	Cum %
0.0–1.9	65	0.6	0.6	144	26.9	26.9	52	44.1	44.1
2.0–3.9	170	1.6	2.2	149	27.9	54.8	47	39.8	83.9
4.0–5.9	292	2.7	4.9	84	15.7	70.5	14	11.8	95.8
6.0–7.9	399	3.7	8.6	56	10.5	81.0	4	3.4	99.2
8.0–9.9	482	4.5	13.1	37	6.9	87.9	1	0.8	100.0
10.0–11.9	686	6.4	19.5	34	6.4	94.3			
12.0–13.9	828	7.7	27.2	6	1.1	95.3			
14.0–15.9	995	9.2	36.4	10	1.8	97.1			
16.0–17.9	1118	10.4	46.8	5	0.9	98.0			
18.0–19.9	1095	10.2	57.0	6	1.1	99.1			
20.0–21.9	1191	11.1	68.0	2	0.4	99.5			
22.0–23.9	1074	10.0	78.0	1	0.2	99.7			
24.0–25.9	846	7.9	85.9	0	0.1	99.7			
26.0–27.9	628	5.8	91.7	1	0.2	100.0			
28.0–29.9	357	3.3	95.0						
30.0–31.9	294	2.7	97.8						
32.0–33.9	158	1.5	99.2						
34.0–35.9	58	0.5	99.8						
36.0–37.9	21	0.2	100.0						
38.0–40.0	4	0.0	100.0						
<b>Total</b>	<b>10761</b>	<b>100.0</b>		<b>535</b>	<b>100.0</b>		<b>118</b>	<b>100.0</b>	

CORE-OM, Clinical Outcomes in Routine Evaluation – Outcome Measure; Cum %, cumulative percentage.



**Fig. 1** Distributions of clinical scores on the Clinical Outcomes in Routine Evaluation – Outcome Measure for the clinical, general population and non-distressed samples, showing the recommended cut-off score (10) between the clinical and general populations.

**Table 2** Comparison of CORE-OM clinical score with follow-up CIS-R scores split into four levels of severity in the general population sample

CIS-R score	Effective <i>n</i>	CORE-OM clinical score		
		Mean	95% CI	s.d.
0–5	375	3.2	2.9–3.5	2.7
6–11	95	6.1	5.5–6.6	2.9
12–17	37	10.5	9.5–11.5	3.0
18+	28	15.1	12.8–17.3	5.8

CIS-R, Clinical Interview Schedule – Revised; CORE-OM, Clinical Outcomes in Routine Evaluation – Outcome Measure; CI, confidence interval.

four CIS-R levels of severity (see Singleton & Lewis, 2003).

### CORE-OM reliable change index and cut-off values

According to Jacobson & Truax (1991), achieving reliable and clinically significant improvement in psychological treatment requires the client to meet two criteria. First, pre-post improvement must be reliable, in the sense of being large enough not to be attributable to measurement error. Second, improvement must be clinically significant, which is most often understood as the person beginning treatment as part of the dysfunctional clinical population and entering the non-clinical population during or after treatment, assessed as a change in score from above

to below a clinical cut-off level on the criterion measure.

As a reliable change index (RCI), Jacobson & Truax (1991) suggested the pre-post difference that, when divided by the standard error of measurement, is equal to 1.96, calculated as  $RCI = 1.96sd \sqrt{2\sqrt{1-r}}$ . The RCI thus depends on the measure's standard deviation (sd) and reliability (*r*). It is likely to be smaller in a general population sample than in a clinical sample because of the reduced variability of scores. Using the general population internal consistency reliability (0.91) yielded RCIs of 3.6 in the general population sample and 5.9 in the clinical sample.

Following the logic and procedures of Jacobson and colleagues (see Jacobson & Truax, 1991), we calculated a clinical cut-off value between the clinical and normal

populations on the CORE-OM using the following expression:

$$\frac{\text{mean}_{\text{clin}}\text{sd}_{\text{norm}} + \text{mean}_{\text{norm}}\text{sd}_{\text{clin}}}{\text{sd}_{\text{norm}} + \text{sd}_{\text{clin}}}$$

The cut-off value between the clinical population and the general population was 9.9. Calculated separately, the cut-off score for men was 9.3 and the cut-off score for women was 10.2, reflecting the slightly higher mean for women in the clinical sample. We recommend rounding this to 10 for all respondents (see Fig. 1). As can be calculated from Table 1, the cut-off of 10 yields a sensitivity (true positive rate) of 87% and a specificity (true negative rate) of 88% for discriminating between members of the clinical and general populations. The cut-off value between the clinical population and the non-distressed population was 7.3.

## DISCUSSION

### Acceptability

The very low rate of explicit refusal to complete the forms when presented in a face-to-face situation and the low number of missing items on returned forms indicate that the CORE-OM is acceptable for use in a general population. The lower rate of completion among those agreeing to return the form by mail represents a problem that is not restricted to the CORE-OM. On the other hand, we cannot rule out the possibility that some of the non-returned forms reflected unacceptability of the measure.

### Internal consistency and convergent validity

The high internal consistency of the CORE-OM ( $\alpha=0.91$ ) confirms its robust structure in a general population, although this may also be an indication of redundant items. Its correlation of 0.77 with the CIS-R is consistent with its previously reported convergence with other measures of psychological distress and disturbance (Evans *et al*, 2002; Leach *et al*, 2005, 2006; Cahill *et al*, 2006).

### CORE-OM cut-off scores

Our recommended CORE-OM cut-off score of 10 between clinical and general populations (see Fig. 1) has the advantage of a straightforward interpretation, equivalent to a mean item score of 1.0. This cut-off score represents an advance over

previous cut-off scores, insofar as the weighted general population sample drawn from the Singleton & Lewis (2003) psychiatric morbidity survey follow-up was a more representative sample of British adults than were previous comparison samples. A cut-off score based on representative samples is essential for determining rates of reliable and clinically significant change (following Jacobson & Truax, 1991) – an important procedure in evaluating the effectiveness of contrasting psychological interventions (e.g. Stiles *et al*, 2006).

The cut-off score of 10 is somewhat lower than the previously reported separate cut-off scores of 11.9 for men and 12.9 for women (Evans *et al*, 2002), reflecting the relatively lower mean CORE-OM clinical score in the general population sample (4.8, with no gender difference), as compared with the university students and convenience sample used previously (6.9 for men and 8.1 for women; Evans *et al*, 2002). The latter, somewhat higher, means may reflect higher distress levels among students than in the general population (Stewart-Brown *et al*, 2000) and the inclusion of relatively psychologically aware people in the convenience sample. Using the earlier, higher cut-off scores left 20% of people referred to therapy services below the cut-off level (Evans *et al*, 2003; Barkham *et al*, 2005); revising this indicator of caseness downwards acknowledges that such people are being referred for clinically significant distress. Congruently, the customary cut-off between clinical and non-clinical populations on the Beck Depression Inventory (BDI; Beck *et al*, 1988) is also 10, and transformation tables between the BDI and the CORE-OM (Leach *et al*, 2006) suggest that a BDI score of 10 is equivalent to a CORE clinical score of 10.0 for men and 9.7 for women.

The cut-off score of 10 represents a distinction between a clinical population (those attending psychological therapy services) and a non-clinical (general) population, rather than between those with or without a diagnosis. The cut-off score for distinguishing a sample meeting criteria for a specific diagnosis (e.g. depression) might be higher.

#### Validity of cut-off scores: additional considerations

Psychological disturbance, as measured by the CORE-OM and CIS-R, is not a discrete phenomenon but a matter of degree.

JANICE CONNELL, BSC, MICHAEL BARKHAM, PhD, Psychological Therapies Research Centre, University of Leeds, Leeds, UK; WILLIAM B. STILES, PhD, Miami University, Oxford, Ohio, USA; ELSPETH TWIGG, BSc, Psychological Therapies Research Centre, University of Leeds; NICOLA SINGLETON, MSc, OLGA EVANS, PhD, Office for National Statistics, London; JEREMY N. V. MILES, PhD, Department of Health Sciences, University of York, York, UK

Correspondence: Janice Connell, Psychological Therapies Research Centre, 17 Blenheim Terrace, Leeds LS2 9JT, UK. Email: j.connell@leeds.ac.uk

(First received 30 September 2005, final revision 18 May 2006, accepted 4 July 2006)

Consequently, any cut-off point is to some degree arbitrary. In contrast, when detecting the presence or absence of discrete medical conditions such as prostate cancer, only the test is continuous, and cut-off scores are selected to optimise prediction. Even for discrete target conditions, optimal cutting scores may vary substantially and systematically depending on the base rates in the local population and on the value placed on alternative types of detection and error (Rorer *et al*, 1966a,b). Optimal cutting scores tend to fall as the base rate of the target (high-scoring) and the relative cost of false negatives (undetected members of the target group) increase. Thus, any recommended cut-off may require adjustment to fit circumstances. In this context, the CORE-OM cut-off score of 7.3 between the clinical and non-distressed populations and the previous recommended cut-off score of 11.9 or 12.9 (Evans *et al*, 2002) helpfully bracket our recommended cut-off score of 10.

Although CIS-R scores were used in the procedures for setting up the general population sample – sampling only 20% of respondents scoring 0–5 in the original survey – this was compensated for by the weighting procedures. Consequently, the validity of the general population CORE-OM cut-off scores did not depend on the CIS-R. On the other hand the CIS-R scores were used in defining the non-distressed sample (i.e. only those scoring 0–5 in the follow-up survey were included), so the validity of the cut-off between it and the clinical sample (7.3) rests partly on the validity of the CIS-R.

#### Limitations and caveats

In assessing the convergent validity between two measures, the order of presentation would ideally be counterbalanced. However, in the design of the psychiatric morbidity follow-up survey the CORE-OM was administered at the end of a 1–1.5 h interview which included the CIS-R. This

might also have adversely affected the response rate.

General population samples, because of their skewed distributions, tend to violate implicit assumptions of normality and distort calculation of the cut-off points (Martinovich *et al*, 1996). Because of the skew, the calculated cut-off scores between the clinical population and the general population (9.9) and the non-distressed group (7.3) were lower than the points where the distribution lines cross in Fig. 1, which would be optimal cutting scores if one assumed that clinical and general populations were discrete, with a 50% base rate and equal dis-utility of false negatives and false positives. The violation of all of these assumptions (normal distribution, discrete groups, equal occurrence rates of target and non-target groups, equal utilities of detection) under realistic clinical conditions underlines our caution against rigid application of a fixed cut-off.

#### ACKNOWLEDGEMENTS

We thank Howard Meltzer and colleagues at the Office for National Statistics for enabling this work to be carried out, and Rachael Harker for her invaluable help. Authors affiliated to the Psychological Therapies Research Centre were supported by the Priorities and Needs Research and Development Levy from Leeds Mental Health and Teaching National Health Service Trust.

#### REFERENCES

- Barkham, M., Margison, F., Leach, C., *et al* (2001) Service profiling and outcomes benchmarking using the CORE-OM: towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, **69**, 184–196.
- Barkham, M., Gilbert, N., Connell, J., *et al* (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.
- Beck, A. T., Steer, R. A. & Garbin, M. G. (1988) Psychometric properties of the Beck Depression

- Inventory: twenty-five years of evaluation. *Clinical Psychology Review*, **8**, 77–100.
- Cahill, J., Barkham, M., Stiles, W. B., et al (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.
- Cronbach, L. J. (1951)** Coefficient alpha and the internal structure of tests. *Psychometrika*, **16**, 297–334.
- Evans, C., Connell, J., Barkham, M., et al (2002)** Towards a standardised brief outcome measure: psychometric properties and utility of the CORE–OM. *British Journal of Psychiatry*, **180**, 51–60.
- Evans, C., Connell, J., Barkham, M., et al (2003)** Practice-based evidence: benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology and Psychotherapy*, **10**, 374–388.
- 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.
- Leach, C., Lucock, M., Barkham, M., et al (2005)** Assessing risk and emotional disturbance using the CORE–OM and HoNOS outcome measures at the interface between primary and secondary mental healthcare. *Psychiatric Bulletin*, **29**, 419–422.
- Leach, C., Lucock, M., Barkham, M., et al (2006)** Transforming between Beck Depression Inventory and CORE–OM scores in routine clinical practice. *British Journal of Clinical Psychology*, **45**, 153–166.
- Lewis, G., Pelosi, A. J., Araya, R. et al (1992)** Measuring psychiatric disorder in the community: a standardized assessment for use by lay interviewers. *Psychological Medicine*, **22**, 465–486.
- Martinovich, Z., Saunders, S. & Howard, K. I. (1996)** Some comments on 'Assessing clinical significance.' *Psychotherapy Research*, **6**, 124–132.
- Rorer, L. G., Hoffman, P. J., LaForge, G., et al (1966a)** Optimum cutting scores to discriminate groups of unequal size and variance. *Journal of Applied Psychology*, **50**, 153–164.
- Rorer, L. G., Hoffman, P. J. & Hsieh, K. (1966b)** Utilities as base-rate multipliers in the determination of optimum cutting scores for the discrimination of groups of unequal size and variance. *Journal of Applied Psychology*, **50**, 364–368.
- Singleton, N. & Lewis, G. (2003)** *Better or Worse: A Longitudinal Study of the Mental Health of Adults Living in Private Households*, 2000. London: TSO (The Stationery Office).
- Singleton, N., Bumpstead, R., O'Brien, M., et al (2001)** *Psychiatric Morbidity Among Adults Living in Private Households, 2000*. London: TSO (The Stationery Office).
- Singleton, N., Lee, A. & Meltzer, H. (2002)** *Psychiatric Morbidity Among Adults Living in Private Households, 2000: Technical Report*. London: Office for National Statistics.
- Stewart-Brown, S., Evans, J., Patterson, J., et al (2000)** The health of students in institutes of higher education: an important and neglected public health problem? *Journal of Public Health Medicine*, **22**, 492–498.
- Stiles, W. B., Barkham, M., Twigg, E., et al (2006)** Effectiveness of cognitive–behavioural, person-centred, and psychodynamic therapies as practiced in UK National Health Service settings. *Psychological Medicine*, **36**, 555–566.
- Ware, J. E., Kosinski, M. & Keller, S. D. (1996)** A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Medical Care*, **34**, 220–233.