

ORIGINAL PAPER

Charlotte Hanlon · Girmay Medhin · Atalay Alem · Mesfin Araya · Abdulreshid Abdulahi · Markos Tesfaye · Dawit Wondimagegn · Hassan Taha · Birke Anbesse · Yonas Baheretibeb · Michael Dewey · Martin Prince

Measuring common mental disorders in women in Ethiopia

Reliability and construct validity of the comprehensive psychopathological rating scale

Received: 18 May 2007 / Accepted: 5 March 2008 / Published online: 24 April 2008

Abstract *Background* There is a dearth of methodological studies critically evaluating reliability, validity and feasibility of measures of common mental disorders (CMD) in low-income countries. *Methods* Test-retest and inter-rater reliability of categorisation of CMD caseness, according to locally agreed criteria using the Comprehensive Psychopathological Rating Scale (CPRS), was measured in 99 women from out-patient clinics (inter-rater) and 99 women from a primary healthcare centre (test-retest) in Ethiopia. The construct validity of CMD as measured with CPRS was assessed with exploratory factor analysis using maximum likelihood with varimax rotation. *Results* Test-retest reliability was fair ($\kappa = 0.29$). Subsequent assessment of inter-rater reliability found excellent agreement ($\kappa = 0.82$). The construct of CMD appeared unidimensional, combining depressive, anxiety and somatic symptoms. *Conclusions* Detection of socioculturally meaningful cases of CMD in Ethiopia can be reliably achieved with local psychiatrist assessment using CPRS, although thorough training is essential.

Key words Reproducibility of Results – Ethiopia – Africa South of the Sahara – Mental disorders – Factor analysis, statistical

Introduction

Common mental disorders (CMD), characterised by significant levels of depressive, anxiety and/or somatic symptoms, appear to have a unidimensional underlying construct in community samples [11, 14]. “Caseness” for CMD may, therefore, best be defined as a level of symptom burden crossing a threshold of clinical significance, regardless of the precise constellation of symptoms present [11]. Conceptualising CMD caseness in this way has particular value cross-culturally as it (1) avoids presupposing the presence of Western-based diagnostic syndromes of mental disorder such as those outlined in the Diagnostic and Statistical Manual of Mental Disorders, fourth edition (DSM-IV) or International Classification of Disease, tenth edition (ICD-10) [2, 29], (2) recognises the relevance of non-Western manifestations of mental distress and (3) allows significance thresholds to be developed that are salient in the cultural setting [7, 21].

Reference measurement of CMD for research purposes most often employs a standardised, semi-structured clinician interview, for example the Schedules of Clinical Assessment in Neuropsychiatry (SCAN) [30]. By relying on standardised questions and diagnostic algorithms, this approach fails to capture the less differentiated but clinically important symptom combinations characterising community-level CMD. In addition, the length of such interview schedules limits their feasibility where clinician researchers are in short supply and can be burdensome for non-literate informants. The extensive training which is required is also not easily accessed from low-income settings.

Dr. C. Hanlon (✉) · M. Dewey · M. Prince
King's College London, Institute of Psychiatry
Health Services and Population Research Dept.
P.O. Box 60, De Crespigny Park, Camberwell
London, SE5 8AF, UK
Tel.: +44-207/848-0136
Fax: +44-207/848-5450
E-Mail: charlotte.hanlon@iop.kcl.ac.uk

G. Medhin
Aklilu Lemma Institute of Pathobiology
Addis Ababa University
Addis Ababa, Ethiopia

A. Alem · M. Araya · A. Abdulahi · M. Tesfaye ·
D. Wondimagegn · H. Taha · B. Anbesse · Y. Baheretibeb
Dept. of Psychiatry, Faculty of Medicine
Addis Ababa University
Addis Ababa, Ethiopia

In our quest for an alternative measure of CMD that would be reliable, valid and feasible for use in Ethiopia, we trained local psychiatrists in use of the Comprehensive Psychopathological Rating Scale (CPRS) [4]. The CPRS is an observer-rated scale which was designed to identify the presence and change over time of a broad range of symptoms and signs of mental disorder, as well as to provide a global rating of the presence or absence of mental disorder. Although the CPRS has mainly been used in clinical settings, the scale has face validity for use in the community and does not constrain clinicians to use international diagnostic criteria in deciding on a “case” of CMD. Using the CPRS means that the evaluation of caseness can benefit from the clinician’s psychiatric interviewing skills, clinical expertise and local knowledge while ensuring comprehensive assessment and standardised rating.

The primary aim of this study was to evaluate the reliability and feasibility of CPRS rating of CMD in Ethiopia. We also undertook exploratory factor analysis of CPRS as a means of evaluating the construct validity of CMD in this setting.

Method

This study evaluated the measurement of CMD caseness using the CPRS as follows: (1) test-retest reliability, (2) inter-rater reliability and (3) construct validation.

■ Setting and sampling

The study was conducted among women aged 15–45 years recruited from governmental health facilities in Addis Ababa, the capital of Ethiopia.

- (1) Test-retest reliability study: Attendees at the Beletshachew Primary Health Care (PHC) clinic over 1 month in 2004 formed the sample population. Women were systematically sampled from a sampling frame composed daily from a list of all women registering with the clinic,
- (2) Inter-rater reliability study: A convenience sample of 99 women of reproductive age attending psychiatric (49.5%), medical (8.1%) and antenatal (36.4%) out-patient clinics held at Amanuel psychiatric hospital and St Paul’s general hospital over 1 month in 2005, and
- (3) Construct validation: Consecutive women attending the Addis Ketama and Selam Primary Health Care clinics for postnatal checks over 1 month in 2006.

■ Exclusion criteria

In all three studies women were excluded from participation if acutely unwell, that is requiring immediate medical attention or too mentally unwell to engage in an interview, or non-fluent in Amharic, the National language of Ethiopia.

■ Measurement of CMD using the CPRS

The CPRS has 66 items; 40 symptoms based on the subjective report of the interviewee, 25 signs rated on the basis of observation during the interview and a global rating indicating presence of significant mental disorder. The presence of clearly defined symptoms or signs of mental disorder is rated on a 4-point scale (0–3). The definitions of each scale

point were standardised as follows: 0 = not present; 1 = doubtful whether present, and not interfering with life; 2 = definitely present and of moderate severity; 3 = severe or incapacitating. Clinicians were asked to conduct a full psychiatric interview and then complete the CPRS ratings using all available information. The interviewers were free to (1) phrase questions to be understandable to respondents, (2) ask about symptoms not included within the CPRS, (3) screen out culturally acceptable beliefs and behaviours, and (4) probe for indications of culturally relevant clinical significance.

The judgment as to caseness of CMD did not depend on a simple tally of CPRS items, but rather on the basis of significance criteria, as follows:

- (1) Subjective report of significant distress, (2) Interfering with functioning (occupational, social or interpersonal), (3) Objectively significant disorder even if not considered so by the participant, for example due to lack of insight or (4) Response considered disproportionate to any adverse circumstances reported or representing a change for that individual even in the presence of ongoing adversity.

■ Training in CPRS

Initially four Ethiopian trainee psychiatrists, with a minimum of 6 months psychiatric experience, were trained and participated in the test-retest study. Two of these psychiatrists also participated in the inter-rater study and were supplemented by two new psychiatrists who received their own training in CPRS. The CPRS items were not translated into the National language of Ethiopia (Amharic) because they are not intended to be read out verbatim. All medical education in Ethiopia, including psychiatric training, is conducted in English, and therefore training was focused around ensuring full understanding of the concepts behind the items in English. Each item of the CPRS was discussed with three senior Ethiopian psychiatrists, and practice interviews and ratings carried out over 3 days. Discrepancies in rating were discussed and consensus reached.

■ Reliability

- (1) Test-retest reliability study: Assessment of test-retest reliability involves comparing ratings from separate interviews conducted on the same patient by different interviewers. Thus test-retest reliability includes the process of eliciting a feature of mental disorder as well as the rating of the elicited symptom or sign. In this study, each participant was interviewed consecutively by two different Ethiopian psychiatrists. The results of this test-retest study, together with the need to evaluate the reliability of new interviewers joining the project, led us to conduct an inter-rater reliability study.
- (2) Inter-rater reliability study: When assessing inter-rater reliability, the participant’s responses are independently and simultaneously rated by two observers, one of whom conducts the interview. This method only measures agreement in the way that symptoms and signs are rated.

Although a less stringent test than test-retest, the inter-rater reliability study was feasible and allowed the new psychiatrists direct experience of the interviewing style of the original interviewers. In this study, one psychiatrist conducted a full psychiatric examination of the participant and then both psychiatrists (interviewer and observer) independently rated the CPRS items.

In both reliability studies, allocation of the four psychiatrists was randomised using an incomplete blocks design to ensure equal numbers of possible psychiatrist pairs. The order in which the psychiatrists interviewed patients (test-retest study) and whether they were interviewing or observing (inter-rater study) was also randomly assigned. See Table 1 for the allocation of interviewers. Each psychiatrist was masked to the other ratings. CPRS item scores and a global CMD rating were recorded.

In order to understand how the CPRS cases of CMD related to international diagnostic criteria, the psychiatrists were also asked to document the presence of any axis I diagnoses according to DSM-IV, regardless of whether or not participants were categorised as CPRS cases of CMD. All of the psychiatrists were experienced in

Table 1 Randomly allocated interviewer pairs for the test-retest and inter-rater reliability studies (individual interviewer identified A to E)

Test-retest reliability		Inter-rater reliability	
A	B	A	B
A	C	A	E
A	D	A	F
B	C	B	E
B	D	B	F
C	D	E	F

application of the DSM-IV but were additionally supplied with the DSM-IV criteria and asked to record the following diagnoses where present: depressive disorders (296.2/296.3/296.9/300.4), anxiety disorders (300.x), acute stress reaction (308.x), adjustment disorder and post-traumatic stress disorder (309.x).

Construct validity

The construct validity of CMD in Ethiopia was assessed by examining the factor structure of the CPRS scale. We were interested in the presentation of CMD in primary health care/antenatal care. Therefore reliability study participants who were recruited from psychiatric or medical outpatients were excluded, leaving a sample $n = 138$. In addition, CPRS data was available on 100 consecutive postnatal women attending PHC services for vaccination of their new infant. The resulting combined sample size used for construct validation numbered 238.

Statistical methods

Sample size calculation

Assuming the prevalence of CMD in the Primary Healthcare setting to be 15%, in order to estimate κ with a 95% confidence interval of width 0.4 assuming a true value of 0.7, a sample size of 100 women was needed [8].

Data analyses

Data were analysed using Stata version 8.0 [26]. Cohen's kappa (κ) [9] was calculated to show the degree of agreement in categorisation of CMD caseness (CPRS global rating of 2 or 3) over and above that expected by chance alone. Agreement on rating individual CPRS items was estimated using weighted κ coefficients with weights: $1 - |i - j| / (k - 1)$, where i and j index the rows and columns of the ratings by the two raters and k is the maximum number of possible ratings. This accounted for the distance between ratings in calculating the level of agreement. Agreement on total CPRS score was evaluated as recommended by Bland and Altman [6]. Maximum likelihood factor analysis

with varimax rotation was carried out and factors extracted on the basis of the scree plot, the amount of variance explained by the factor and interpretability of the resulting factors [23].

Ethical considerations

Ethical approval was granted by Research Ethics Committees in the relevant academic institutions in Ethiopia and the UK. Participants gave informed and voluntary consent. Women with significant mental health problems were referred for free treatment.

Results

Sample characteristics

The mean age of participants in the test-retest and inter-rater reliability studies ($n = 99$ in each sub-sample) was 25.5 (standard deviation (SD) 6.25) and 27.0 (6.6) years, respectively, and in the final factor analysis sample ($n = 238$) was 25.7 years (5.5). In the test-retest study the majority of women ($n = 56$; 65.9%) underwent both interviews on the same day, the remainder less than 10 days apart. Participants for the inter-rater study came predominantly from psychiatric ($n = 43$; 45.3%) and antenatal clinics ($n = 39$; 41.0%).

Caseness for CMD

The estimated prevalence of CPRS cases of CMD was 33.3% in the test-retest study (first interview) and 23.3% in the inter-rater study (interviewer rating). In both studies the prevalence of "any DSM-IV diagnosis" was significantly higher than CPRS cases of CMD (Test-retest study $\chi^2 = 27.80$; $P < 0.001$. Inter-rater study $\chi^2 = 46.25$; $P < 0.001$, Table 2).

The distribution of DSM-IV diagnoses in CPRS cases of CMD is shown in Table 3.

The median total CPRS score in cases of CMD vs. non-cases was 21 (IQR 17) vs. 9 (IQR 12) for the test-retest study and 23 (IQR 10) vs. 1 (IQR 3) for the inter-rater study.

Reliability of assessment of CMD caseness

Test-retest κ for global CPRS caseness was fair (κ 0.29) but inter-rater reliability was excellent (κ 0.82).

Table 2 Prevalence of global CPRS caseness and DSM-IV diagnoses and estimated kappa for diagnostic agreement

	Test-retest reliability study ($n = 99$)			Inter-rater reliability study ($n = 99$)		
	Prevalence (%)		Kappa (SE)	Prevalence (%)		Kappa (SE)
	1st I/V	2nd I/V		I/V	Observer	
CPRS case of CMD	33.3	37.4	0.29(0.10)	23.2	19.2	0.82(0.07)
Any DSM-IV diagnosis	53.5	46.5	0.34(0.10)	27.3	22.2	0.60(0.09)
DSM-IV depression	29.3	19.2	0.44(0.10)	8.1	9.1	0.63(0.14)

I/V = interviewer

Table 3 Frequency distribution of primary DSM-IV diagnoses in cases of common mental disorder according to the CPRS

DSM-IV diagnostic category	Test-retest reliability (<i>n</i> = 99)		Inter-rater reliability (<i>n</i> = 99)	
	<i>n</i>	% categorised as CPRS case	<i>n</i>	% categorised as CPRS case
None	46	6.5	72	5.6
Depression	29	79.3	9	100.0
Dysthymia	4	75.0	1	0
Generalised anxiety disorder	16	25.0	2	100.0
Social phobia	1	0	1	0
Specific phobia	2	0	3	0
Somatisation	1	0	0	0
Anxiety (not otherwise specified)	0	0	2	100.0
Bipolar disorder	0	0	5	60.0
Schizophrenia	0	0	4	75.0
Any DSM-IV case	53	33.3	28	23.3

See Table 2. The κ for test-retest reliability was no different if the interviews were carried out on the same day (κ 0.27) compared to being up to 1 week apart (κ 0.26). Post-hoc inspection of κ for interviewer pairs indicated that agreement for presence or absence of CMD in pairs including interviewer A (A-B κ 0.16; A-C κ 0.33; A-D κ 0.03) was substantially lower than the other pair combinations (B-C κ 0.51; C-D κ 0.44; B-D κ 0.40).

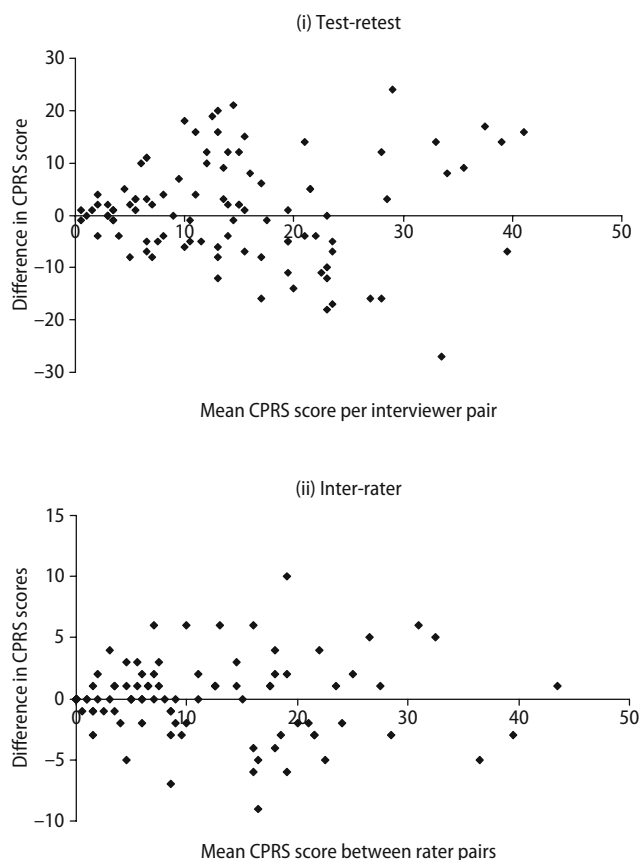
■ Agreement for total CPRS score

The mean difference in CPRS total score in the test-retest study was 1.1 (95%CI -0.9 to 3.0), although the limits of agreement (\pm 2SD from the mean) were -18.3 (95%CI -21.7 to -14.9) and 20.4 (95%CI 17.1 to 23.8) indicating substantial variation in the differences in CPRS score between interviewers. See Fig. 1 for the inter-rater study, the mean difference in CPRS score was 0.1 (95%CI -0.6 to 0.7). The limits of agreement were -6.0 (95%CI -7.1 to -5.0) and 6.1 (95%CI 5.0 to 7.2) indicating much less variation in agreement in total CPRS score between raters.

■ CPRS item frequencies

The prevalence of endorsement of individual CPRS items (dichotomised to indicate clinical significance: 0/1 vs. 2/3) is shown in Table 4.

In CMD cases, six of the top ten most prevalent CPRS items were the same in both studies (prevalence in test-retest vs. inter-rater studies): fatiguability (42.4% vs. 60.9%), pessimistic thoughts (42.4% vs. 39.1%), inability to feel (39.4% vs. 43.5%), reduced sleep (39.4% vs. 43.5%), reduced sexual interest (36.4% vs. 39.1%) and concentration difficulties (36.4% vs. 60.9%). Suicidal thoughts were present in 27.3% and 17.4% of cases in the test-retest and inter-rater studies, respectively. The only observational items rated as clinically significant were apparent sadness and reduced speech, both found exclusively in cases of CMD.

**Fig. 1** Difference in CPRS scores between assessing psychiatrists for (i) test-retest and (ii) inter-rater assessments

■ Agreement in rating individual CPRS items

In the test-retest reliability study the weighted κ for nearly two-thirds of CPRS items was 0.3 or higher, indicating moderate agreement (Table 4). For clinically significant CPRS items present at a prevalence of \geq 5%, the lowest weighted κ estimate was for aches and pains (0.12) and phobia (0.16) whereas the best agreement was found for suicidal thoughts (0.59), reduced appetite (0.50) and reduced sleep (0.49).

Table 4 Weighted kappa for CPRS items present with a prevalence of greater than one percent

CPRS item		Test-retest study		Inter-rater study	
		Prevalence (1st interview)	Weighted Kappa	Prevalence (1st rater)	Weighted Kappa
1	Sadness	24.2	0.38	8.1	0.81
3	Inner tension	6.1	0.22	9.1	0.73
4	Hostile feelings	16.2	0.38	7.1	0.76
5	Inability to feel	20.2	0.41	13.1	0.84
6	Pessimistic thoughts	21.2	0.38	11.1	0.90
7	Suicidal thoughts	12.1	0.59	5.1	0.89
9	Worrying over trifles	20.2	0.22	17.2	0.79
11	Phobias	12.1	0.16	12.1	0.71
13	Indecision	10.1	0.40	7.1	0.74
14	Lassitude	9.1	0.22	17.2	0.80
15	Fatiguability	18.2	0.32	27.3	0.85
16	Concentration difficulties	18.2	0.32	21.2	0.84
17	Failing memory	17.2	0.44	16.2	0.77
18	Reduced appetite	12.1	0.50	19.2	0.90
19	Reduced sleep	18.2	0.49	23.2	0.77
20	Increased sleep	2.0	0.49	4.0	0.79
21	Reduced sexual interest	21.2	0.36	32.3	0.73
23	Autonomic disturbances	8.1	0.31	7.1	0.76
24	Aches and pains	23.2	0.12	14.1	0.76
25	Muscular tension	2.0	0.38	0	0.89
26	Loss of sensation / movement	2.0	0.37	1.0	0.73
30	Disrupted thoughts	1.0	0.21	3.0	0.88
31	Ideas of persecution	1.0	0.48	2.0	0.76
32	Ideas of grandeur	0	–	2.0	0.91
36	Other delusions	0	–	2.0	0.81
37	Commenting voices	0	–	3.0	0.74
38	Other auditory hallucinations	0	–	6.1	0.83
39	Visual hallucinations	2.0	0.18	2.0	0.90
40	Other hallucinations	2.0	0.05	2.0	0.92
41	Apparent sadness	7.1	0.32	6.1	0.53
54	Reduced speech	3.0	0.06	0	0.26

In the inter-rater reliability study, the weighted κ for almost all CPRS items was above 0.70, indicating excellent agreement. The two exceptions were both observational items; apparent sadness (0.53) and reduced speech (0.26).

Factor analysis of CPRS

The frequency of CPRS item endorsement was inspected and items where 2.5% or fewer participants scored one or above were excluded. Likewise items that did not have a correlation of ≥ 0.30 with any other CPRS item were excluded. Factor analysis using Maximum Likelihood methods with varimax rotation gave a one factor solution explaining 24.7% of the variation. Item-factor correlations are shown in Table 5.

Table 5 Factor analysis of comprehensive psychopathological rating scale items in the combined sample ($n = 238$) using maximum likelihood with varimax rotation

CPRS items	Item-factor correlation
Sadness	0.78
Pessimistic thoughts	0.73
Apparent sadness	0.71
Inner tension	0.69
Inability to feel	0.69
Suicidal thoughts	0.64
Worrying over trifles	0.64
Failing memory	0.55
Lassitude	0.54
Reduced sleep	0.53
Reduced sexual interest	0.51
Aches and pains	0.51
Indecision	0.48
Reduced appetite	0.43
Muscular tension	0.42
Reduced speech	0.40

Discussion

Overall, measurement of CMD caseness in Ethiopia with local psychiatrists using the CPRS was shown to be reliable and feasible. Exploratory factor analysis supported the construct validity of CMD measured by

CPRS in this setting. Strengths of our study include the sample size and efforts to define caseness in a socioculturally relevant way.

The majority of reports evaluating the reliability of CMD measurement estimate inter-rater rather than

test-retest reliability and are often based on too small sample sizes [1, 3, 13, 19, 20, 28]. Agreement over global CPRS rating of caseness was only reported in one study from Japan, where κ for inter-rater agreement was lower than in the present study, ranging from 0.58 to 0.74 depending on the interviewer pair [13]. Other inter-rater studies examining the rating of individual CPRS items tend to show good to excellent agreement between different mental health professionals and across cultural settings within Europe [1, 19, 20], with poorer agreement for rating of observed items than self-reported [10, 18]. Kappa has also been noted to be lower for “neurotic” rather than “psychotic” CPRS items [10].

In this study, inter-rater reliability of CMD caseness was excellent and indicates that reliable application of CPRS as a gold standard measure of CMD in Ethiopia is possible with thorough training and clearly defined criteria for determining clinical significance. The results compare favourably with the reliability of more standardised clinical interviews; for example, the Structured Clinical Interview for DSM-III-R (SCID) [25] had κ of 0.37 for “any current diagnosis” in a non-patient sample [28]. Reliability of diagnostic agreement for SCAN diagnosis of depression has been variable across studies, varying from κ of 0.78 (test-retest) in field trials [27] to κ of 0.37 (inter-rater) in an Australian out-patient clinic sample [3], but more consistent for “any DSM diagnosis” (test-retest) κ of 0.62 [24] and inter-rater κ of 0.67 [27].

In the present study, the difference between estimated values of κ for detection of CMD in the test-retest and inter-rater studies was sizeable. There are two potential additional sources of variability when comparing test-retest to inter-rater studies: first, differences in the way clinicians elicit symptoms and signs, and second, variation in symptoms or the way the interviewee discloses symptoms between interviews. In the test-retest study, participants had presented to the primary care centre because of perceived ill health. The first CPRS interview sometimes took place prior to assessment by the primary care staff, whereas the second interview usually took place afterwards, and it is possible that participants were systematically less distressed once their physical health problems had been attended to. It was also noted by the study psychiatrists that most participants appreciated talking at length about their difficulties and appeared to derive therapeutic benefit from the research interview. This might also lead to a diminution of symptoms by the time of the second research interview. On the other hand, many of the participants were physically unwell and waiting around for the second interview could have accentuated their symptoms. Order effects for reporting of symptoms of mental disorder are well-established in the literature, and it is postulated that participants may learn to articulate their symptoms of mental

distress more clearly with practice, reflect upon their experiences and recall more symptoms or indeed modify their responses to decrease the interview duration [12].

Discussion of discordant cases in the test-retest reliability study revealed differing thresholds for deciding on whether a participant’s expressions of mental distress was understandable in view of the level of poverty and social adversity which they were experiencing. Further practice of joint rating of cases prior to the inter-rater reliability study to check application of the criteria for caseness is likely to have contributed to the observed difference in κ estimate in the second study.

The prevalence of CPRS cases of CMD differed markedly between the two reliability studies; 33.3% in the primary care sample vs. 23.2% in the sample from out-patient psychiatry, general medical and antenatal clinics. Random sampling was not employed in either study and so selection bias could have influenced the prevalence estimates in unpredictable ways. Although the sample including psychiatric out-patients might be expected to have a higher morbidity of mental disorder, most patients were receiving treatment and attending for routine follow-up and would therefore be expected to be relatively well. In contrast, at present there is no mental health care available to patients attending primary healthcare clinics and thus the burden of untreated disorder may well be higher. As the purpose of the study was to establish the reliability of measuring CMD caseness rather than to determine prevalence of CMD, the aforementioned differences should not have affected the veracity of our findings.

Determination of the presence of a DSM-IV diagnosis or caseness for CMD was carried out by the same assessor and thus caution is required in interpreting the relation between the categorisations. Nonetheless it is interesting that many women fulfilling criteria for DSM-IV diagnoses were not considered to be cases of CMD. It is possible that DSM-IV diagnoses unduly pathologise expressions of mental distress in this setting which are understandable with local expertise.

Exploratory factor analysis indicated that the construct of CMD in Ethiopia is unidimensional and includes depressive, anxiety and somatic symptoms. Most previous factor analytical studies of the CPRS have been conducted in clinical populations, included persons with psychotic disorders and were based on smaller sample sizes than the present study. These variations in design may explain the wide variation in proposed factor structures for CPRS [5, 15–17, 22]. Our findings most closely matched the Norwegian study conducted in persons with depression or anxiety disorders [16]. This lends support to the construct validity of CPRS in Ethiopia. However, the identified factor only accounted for a relatively small amount of the overall variability (24.7%). The reason for this

could have been our use of a primary healthcare sample where the number and range of symptoms of CMD would be expected to be lower than in psychiatric settings. In addition, CPRS items were originally selected as time-varying aspects of syndromes of mental disorder and thus do not include the full range of CMD characteristics recognised in Western settings, for example guilt. Likewise, culture-specific symptoms, which might be pertinent to the construct of CMD in Ethiopia, are not contained within the CPRS. Further examination of the construct of CMD in community samples from low-income settings is warranted.

Conclusions

Detection of socioculturally meaningful cases of CMD in Ethiopia can be reliably achieved with local psychiatrist assessment using CPRS, although thorough training is essential.

■ **Acknowledgments** This study was funded by a Wellcome Trust training fellowship for Dr Charlotte Hanlon (WT081504/Z/06/Z). We gratefully acknowledge the generosity of study participants in giving us their time, and the kind co-operation received from staff at the Primary Health Care centre and the clinics at Amanuel and St Paul's hospitals.

■ **Declaration of Interest** None.

References

- Amati A, del Vecchio M, Kemali D, Perris C, Vacca L (1978) The comprehensive psychopathological rating scale (CPRS): communicability to, and inter-rater reliability among untrained raters. *Acta Psychiatr Scand* 271(suppl):63–69
- American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders (DSM-IV), 4th edn. APA, Washington
- Andrews G, Peters L, Guzman AM, Bird K (1995) A comparison of two structured diagnostic interviews: CIDI and SCAN. *Aust NZ J Psychiatry* 29(1):124–132
- Asberg M, Montgomery SA, Perris C, Schalling D, Sedvall G (1978) A comprehensive psychopathological rating scale. *Acta Psychiatr Scand Suppl* 271:5–27
- Bertschy G, Viel JF, Ahyi RG (1992) Depression in benin: an assessment using the comprehensive psychopathological rating scale and the principal component analysis. *J Affect Disord* 25(3):173–80
- Bland JM, Altman DG (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* i:307–310
- Bolton P, Wilk CM, Ndogoni L (2004) Assessment of depression prevalence in rural Uganda using symptom and function criteria. *Soc Psychiatry Psychiatr Epidemiol* 39:442–447
- Cantor AB (1996) Sample size calculations for Cohen's kappa. *Psychol Methods* 1(2):150–153
- Cohen J (1960) A coefficient of agreement for nominal scales. *Educ Psychol Meas* 20:37–46
- Goekoop JG, Knoppert-Van der Klein EA, Hoeksema T, Klinkhamer RA, Van Gaalen HA, Van der Velde EA (1991) The interrater reliability of a Dutch version of the comprehensive psychopathological rating scale. *Acta Psychiatr Scand* 83(3):202–205
- Goldberg D, Goodyer I (2005) The origins and course of common mental disorders. London Routledge
- Jensen PS, Watanabe HK, Richters JC (1999) Who's up first? Testing for order effects in structured interviews using a counterbalanced design—statistical data included. *J Abnorm Child Psychol* 27(6):439–446
- Kasa M, Hitomi K (1985) Inter-rater reliability of the comprehensive psychopathological rating scale in Japan. *Acta Psychiatr Scand* 71(4):388–91
- Lewis G (1992) Dimensions of neurosis. *Psychol Med* 22:1011–1018
- von Luckner N, Woggon B, Asberg M, Wiesel FA, Bjerkenstedt L (1985) Scale construction of the Swedish CPRS version. *Neuropsychobiology* 13(4):180–186
- Martinsen EW, Friis S, Hoffart A (1989) A factor analytical study of the comprehensive psychopathological rating scale among patients with anxiety and depressive disorders. *Acta Psychiatr Scand* 80(5):492–8
- Maurer M, Kuny S, Dittrich A, Woggon B (1982) Scale construction of the German version of the comprehensive psychopathological rating scale (CPRS). *Int Pharmacopsychiatry* 17(4):338–353
- Maurer M, Kuny S, Woggon B, Dittrich A, von Luckner N (1984) Comparison of the AMP system and the CPRS with regard to interrater reliability. *Neuropsychobiology* 12(1):27–33
- Montgomery S, Asberg M, Jornestedt L, Thoren P, Traskman L, McAuley R, Montgomery D, Shaw P (1978) Reliability of the CPRS between the disciplines of psychiatry, general practice, nursing and psychology in depressed patients. *Acta Psychiatr Scand Suppl* 271:29–32
- Montgomery S, Asberg M, Traskman L, Montgomery D (1978) Cross-cultural studies on the use of CPRS in English and Swedish depressed patients. *Acta Psychiatr Scand Suppl* 271:33–37
- Patel V, Mann A (1997) Etic and emic criteria for non-psychotic mental disorder: a study of the CISR and care provider assessment in Harare. *Soc Psychiatry Psychiatr Epidemiol* 32:84–89
- Perris C, Eisemann M, von Knorring L, Perris H (1984) Presentation of a subscale for the rating of depression and some additional items to the comprehensive psychopathological rating scale. *Acta Psychiatr Scand* 70(3):261–274
- Pett MA, Lackey NR, Sullivan JJ (2003) Making sense of factor analysis. The use of factor analysis for instrument development in health care research. Sage Publications, Thousand Oaks
- Rijnders CA, van den Berg JF, Hodiament PP, Nienhuis FJ, Furer JW, Mulder J, Giel R (2000) Psychometric properties of the schedules for clinical assessment in neuropsychiatry (SCAN-2.1). *Soc Psychiatry Psychiatr Epidemiol* 35(8):348–352
- Spitzer RL, Williams JB, Gibbon M et al (1992) The structured clinical interview for DSM-III-R (SCID). *Arch Gen Psychiatry* 49:624–629
- Stata Corporation (2003) Intercooled Stata 8.1 for Windows, Texas
- Tomov T, Nikolov V (1990) Reliability of SCAN categories and scores. Results of the field trials. In: Stefanis CN (ed) *Psychiatry: a world perspective*. Elsevier, Amsterdam, pp 107–112
- Williams JBW, Gibbon M, First MB, Spitzer RL, Davies M, Borus J, Howes MJ, Kane J, Harrison GP, Rounsaville B, Wittchen HU (1992) The structured clinical interview for DSM-III-R (SCID) II: multi-site test-retest reliability. *Arch Gen Psychiatry* 49:630–636
- World Health Organisation (1992) International statistical classification of diseases and related health problems (10th Revision) (ICD-10). WHO, Geneva
- World Health Organisation (1992) Schedules for clinical assessment in neuropsychiatry. WHO, Geneva