Clinical Psychology and Psychotherapy
Clin. Psychol. Psychother. **16**, 444–449 (2009)
Published online 21 August 2009 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/cpp.646

Assessment

Validation of the Italian version of the Clinical Outcomes in Routine Evaluation Outcome Measure (CORE-OM)

Gaspare Palmieri,[1]* Chris Evans,[2] Vidje Hansen,[3,4] Greta Brancaleoni,[3] Silvia Ferrari,[1] Piero Porcelli,[5] Francesco Reitano[6] and Marco Rigatelli[1]

[1]*Casa di Cura Villa Igea, Modena, Italy*
[2]*Nottinghamshire Healthcare NHS Trust, Nottingham, UK*
[3]*Institute of Clinical Medicine, University of Tromsø*
[4]*Psychiatric Department, University Hospital Northern Norway*
[5]*Psychosomatic Unit, IRCCS De Bellis Hospital, Bari, Italy*
[6]*Psychological Service (Mental Health Department), APSS Trento, Italy*

The Clinical Outcomes in Routine Evaluation—Outcome Measure (CORE-OM) was translated into Italian and tested in non-clinical (n = 263) and clinical (n = 647) samples. The translation showed good acceptability, internal consistency and convergent validity in both samples. There were large and statistically significant differences between clinical and non-clinical datasets on all scores. The reliable change criteria were similar to those for the UK referential data. Some of the clinically significant change criteria, particularly for the men, were moderately different from the UK cutting points. The Italian version of the CORE-OM showed respectable psychometric parameters. However, it seemed plausible that non-clinical and clinical distributions of self-report scores on psychopathology and functioning measures may differ by language and culture.

Key Practitioner Message:

- A good quality Italian translation of the CORE-OM, and hence the GP-CORE, CORE-10 and CORE-5 measures also, is now available for use by practitioners and anyone surveying or exploring general psychological state. The measures can be obtained from CORE-IMS or yourself and practitioners are encouraged to share anonymised data so that goodclinical and non-clinical referential databases can be established for Italy.

*Corresponding author: Dr Gaspare Palmieri, Casa di Cura Villa Igea, via Stradella 73, 41040, Saliceta San Giuliano, Modena, Italy.
E-mail: gasparepalmieri@yahoo.com

Keywords: Outcome Measures, CORE-OM, Validation, Transcultural Psychology

INTRODUCTION

The Clinical Outcomes in Routine Evaluation–Outcome Measure (CORE-OM; Evans et al., 2000) is a 34-item self-report measure, which is widely used in the UK and shown to be reliable, valid and acceptable in a range of settings (Barkham et al., 2002; Evans, Connell, Barkham, Marshall, & Mellor-Clark, 2003; Shepherd et al., 2005). The items cover four domains: subjective well-being, problems/symptoms, life functioning and risk. These domains are not separate linear factors but different areas of expression of distress and dysfunction as shown in (Lyne, Barrett, Evans, & Barkham, 2006). Higher scores on all domains indicate more problems by reversing scoring on eight positively keyed items; total score has been reported as mean across items completed. More recently, Connell and Barkham (2007) recommend multiplying that score by 10 to avoid decimal fractions, we report raw means for comparability with Evans et al. (2002) (this paper is referred to as 'the UK data'). The CORE-OM measure is copyleft, i.e., can be reproduced on paper without a fee provided it is not changed in any way (all CORE measures are copyleft, see www.coreims.co.uk).

There is a great need for outcome measures in Italian psychotherapy services as routine evaluation is not common in Italy and should be implemented (Chiappelli et al., 2007; Gallo & Rucci, 2000; Lomazzi et al., 1997). The most popular measures currently used in psychological therapies in Italy are the Global Assessment of Functioning (GAF) (Spitzer et al., 1994) and the SCL-90 (Derogatis, 1977), the latter despite the absence of any published psychometric data on the Italian version. The creation and validation of an Italian CORE-OM could help close gaps between clinical practice and research in psychological treatments.

METHOD

The Italian CORE-OM was produced starting with independent forward translations by seven mental health professionals and three professional translators, then all translations were reviewed by a subset of translators and one of the original authors (CE), seeking the best translation holding the meaning of the original version but a comparable informal style in Italian. The final version and the derived shortened forms (i.e., the CORE-SFA/SFB, GP-CORE, CORE-10 and CORE-5) are all available from the first author and www.coreims.co.uk.

The non-clinical sample included volunteer medical students ($n = 189$) approached by GB at the end of lectures and invited to take part in a study of seasonal affective disorder; the response rate was 100%. To extend the sample, GP approached colleagues, psychiatric trainees, occupational therapy students and administrative staff members, with very few refusals ($n = 74$). It was assumed that all would be sufficiently fluent in Italian that there were no exclusion criteria. The clinical sample included inpatients ($n = 68$) and outpatients ($n = 579$) recruited from 17 psychotherapeutic settings across Italy. The two exclusion criteria were, first, the reasonable suspicion that the patient could not read or write Italian fluently and, second, the clinical judgement that it would be inappropriate to ask the patient to participate given their current mental state. There were very few exclusions or refusals.

Analyses largely following those in (Evans et al., 2002) and assessed: acceptability; internal consistency (Cronbach, 1951); principal component analyses (PCA), to assess and compare with the UK/English dimensionality; discriminant validity (assessed by comparing clinical and non-clinical subjects); and convergent validity against the Global Severity Index (GSI) of the Hopkins Symptom Checklist (SCL-90-R) (Derogatis, 1977).

Using the internal reliabilities, and clinical and non-clinical means and standard deviations, we calculated criteria for reliable and clinically significant change (Jacobson & Truax, 1991; Evans, Margison, & Barkham, 1998). The methods of classifying change as 'reliable' and as 'clinically significant' address individual change and complement analyses of group mean change. Reliable change is that found only in 5% of cases if change were simply due to unreliability of measurement. Clinically significant change is what moves a person from a score more characteristic of a clinical population to a score more characteristic of a non-clinical population.

Analyses were conducted in Statistical Package for the Social Sciences (SPSS), version 14 or R, version 2.9.0 (R Development Core Team, 2009).

RESULTS

The dataset consisted of data from 263 non-clinical and 647 clinical participants. Gender was

DOI: 10.1002/cpp

not given by 8 participants from the clinical dataset (0.9%) leaving 514 women (57%) and 388 men (43%); women outnumbered men in both samples (clinical: $n = 443$, 69%; non-clinical sample $n = 192$, 73%).

Age was missing for nine people and ranged from 15 to 80 years (Mean = 33, SD = 11.6). The age for the non-clinical participants ranged from 18 to 59 (Mean = 25; SD = 5.8); for the clinical participants, it ranged from 15 to 80 (Mean = 36, SD = 11.9). The difference was highly statistically significant ($t(874) = 19.1$, $p < 0.0001$). Within the non-clinical sample 212 (81%) completed all the items, 44 (16.7%) omitted item 1, six (2.3%) omitted item 2 and one (0.4%) omitted item 3 (i.e., all returns were usable if prorating up to the usual maximum of three missing items). In the clinical sample, 623 (96%) of returns were complete and the missing item counts were: one item only ($n = 10$, 1.5%), two items ($n = 8$, 1.2%), three items ($n = 5$, 0.8%) and four items ($n = 1$, 0.2%) (i.e., 99.8% usable). Internal consistency did not differ significantly between clinical and non-clinical samples and all domains showed $\alpha > 0.7$ with $\alpha > 0.9$ for the total score as shown in Table 1 with comparisons with the original UK parameters.

Spearman *rho* correlations with the GSI in a clinical sample of 49 inpatients ranged from 0.79 (Well-being) to 0.87 (Symptoms), somewhat stronger than the correlations in the UK data between the CORE-OM scores and the BSI, the most similar measure to the SCL and the GSI scores.

The correlations between age and any of the CORE-OM scores were very small and non-significant (largest *rho* = 0.05 with risk in the clinical sample, $p = 0.22$). Gender effects in mean scores are summarized in Table 1, which replicates the format of table 7 in the UK data except that the *d* values are signed here to make it easier to see which way the gender differences were: negative where the mean for the women was higher (more clinical) than for the men. For the clinical group, the gender differences are markedly greater than in the UK data except for the risk score; for the non-clinical data the differences are very similar to the UK data.

As expected, all domain scores were significantly positively inter-correlated as shown in Table 1, both in the non-clinical and in clinical sample with the risk items showing lower correlations with the other scores (Table 2). As an indicator of dimensionality, we plotted the scree plot (Cattell, 1966) from the PCA (Figure 1), which shows the variance in each independent component of variation in the data of which there are 34 in these two samples.

Table 1. Gender difference in scores for non-clinical and clinical samples

Domain	Non-clinical				95% CI for non-clinical		Clinical				95% CI for clinical	
	Male ($n = 71$)		Female ($n = 192$)		Difference	*d*	Male ($n = 196$)		Female ($n = 443$)		Difference	*d*
	Mean	(SD)	Mean	(SD)			Mean	(SD)	Mean	(SD)		
Well-being	0.98	(0.69)	1.26	(0.85)	−0.49 to −0.08	−0.37 to −0.33	1.95	(0.93)	2.47	(0.94)	−0.67 to −0.36	−0.56 to −0.54
Symptoms	0.84	(0.60)	0.99	(0.65)	−0.32 to 0.02	−0.25 to −0.21	1.65	(0.78)	2.00	(0.86)	−0.49 to −0.22	−0.43 to −0.42
Functioning	1.04	(0.56)	1.01	(0.51)	−0.12 to 0.18	0.05 to 0.08	1.56	(0.63)	1.68	(0.69)	−0.23 to −0.01	−0.18 to −0.17
Risk	0.16	(0.42)	0.11	(0.33)	−0.06 to 0.16	0.13 to 0.17	0.38	(0.62)	0.44	(0.63)	−0.16 to 0.05	−0.09 to −0.08
Non-risk items	0.95	(0.55)	1.03	(0.56)	−0.24 to 0.06	−0.18 to −0.14	1.65	(0.66)	1.93	(0.72)	−0.39 to −0.16	−0.40 to −0.39
All items	0.81	(0.50)	0.87	(0.49)	−0.20 to 0.07	−0.15 to −0.11	1.43	(0.60)	1.66	(0.67)	−0.34 to −0.13	−0.37 to −0.35

DOI: 10.1002/cpp

Table 2. Correlations between domain scores

Spearman's *rho*	W	S	F	R	–R	All
Non-clinical (n = 251)						
Well-being	1.00					
Problems	0.78	1.00				
Functioning	0.73	0.65	1.00			
Risk	0.33	0.43	0.43	1.00		
All non-risk items	0.89	0.92	0.88	0.45	1.00	
All items	0.88	0.92	0.87	0.49	1.00	1.00
Alpha	0.75	0.86	0.76	0.74	0.92	0.92
95% CI alpha	0.70 to 0.79	0.84 to 0.88	0.71 to 0.80	0.69 to 0.78	0.90 to 0.93	0.91 to 0.93
Alpha in UK data	0.77	0.90	0.86	0.79	0.94	0.94
Clinical (n = 632)						
Well-being	1.00					
Problems	0.78	1.00				
Functioning	0.71	0.67	1.00			
Risk	0.50	0.57	0.49	1.00		
All non-risk items	0.88	0.93	0.88	0.59	1.00	
All items	0.87	0.93	0.87	0.66	0.99	1.00
Alpha	0.71	0.87	0.77	0.77	0.91	0.92
95% CI alpha	0.68 to 0.74	0.86 to 0.88	0.75 to 0.79	0.75 to 0.79	0.91 to 0.92	0.91 to 0.93
Alpha in UK data	0.75	0.88	0.87	0.79	0.94	0.94

W = well-being; S = problems/symptoms; F = functioning; R = risk; –R = all items except the risk items; All = all 34 items; CI = confidence interval.

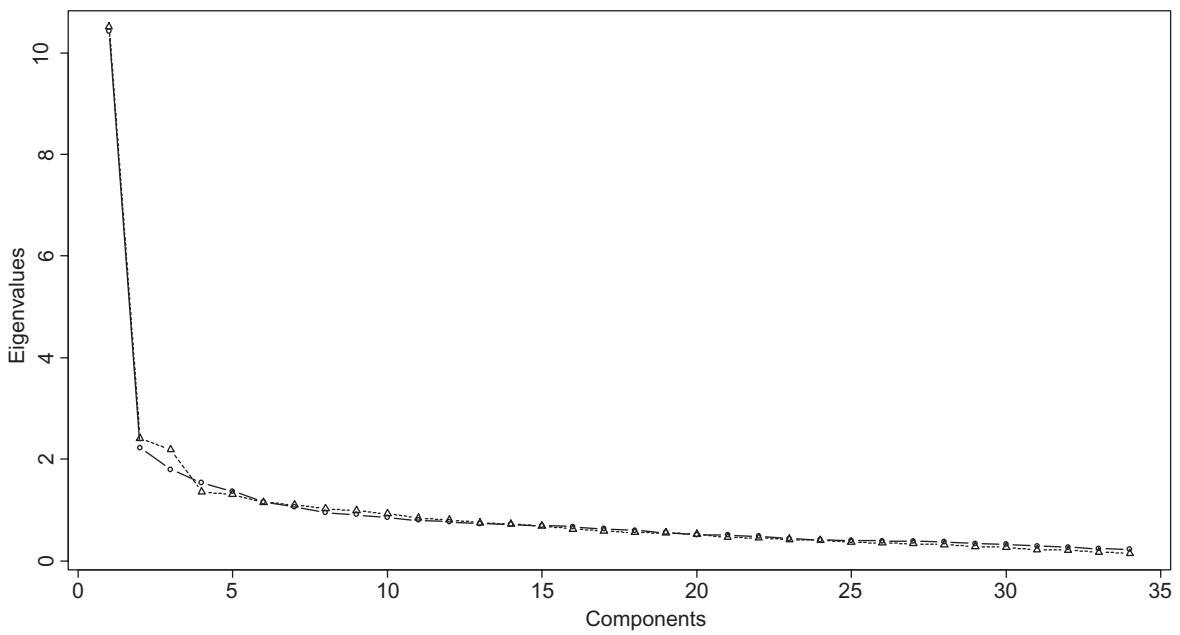

Figure 1. Scree plot of principal component analyses (PCA) item
Triangles and dashed lines for non-clinical data, circles and continuous lines for clinical

DOI: 10.1002/cpp

Table 3. Criteria of reliable (RC) and clinically significant (SC) change

	Italy			UK		
	RC	SC Male	SC Female	RC	SC Male	SC Female
Well-being	1.45	1.40	1.84	1.33	1.37	1.77
Symptoms	0.85	1.30	1.43	0.85	1.44	1.62
Functioning	0.89	1.29	1.30	0.84	1.29	1.30
Risk	0.83	0.25	0.22	0.95	0.43	0.30
All non-risk items	0.60	1.27	1.42	0.55	1.36	1.50
All items	0.52	1.09	1.20	0.51	1.19	1.29

As in the UK data for both clinical and non-clinical datasets, both plots showed a dominant first component though slightly smaller than in the UK data, here accounting for 33% of the variance in both samples. The plots suggested an elbow after three components, more clearly for the non-clinical than the clinical data and are similar to the scree plots from the UK data.

There were statistically significant differences between clinical and non-clinical datasets on all scores, with effect size (Cohen's *d*) ranging from 1.01 to 1.34 for the total score. The criteria for reliable and clinically significant change are shown in Table 3.

DISCUSSION

The Italian version of CORE-OM was well accepted in respectably sized non-clinical and clinical samples. As we had expected and hoped, the translated version of the CORE-OM showed respectable basic psychometric parameters of internal consistency, discriminant and convergent validity.

Internal consistency for the whole measure and all domain scores was lower than that in the UK referential data (Evans et al., 2002). However, the consistencies are still very respectable for short scales and all above 0.7.

For convergent validity, because this was an unfunded study, we were only able to make a comparison with the SCL-90R and only in a selected subsample. These correlations, particularly for the overall mean and non-risk scores, were very similar to those in the British sample and, again, the correlation with the problems/symptoms score was the highest across the CORE-OM domain scores.

Though there were large and statistically significant differences between clinical and non-clinical datasets on all scores, the differences tended to be smaller in the Italian than in the UK data. This may partly reflect interesting and quite marked differences between men and women in the Italian clinical sample which do not parallel the UK gender differences.

Exploratory PCA showed a strong first component similar to that in the UK sample but, though large, the dataset is not sufficiently large to support detailed confirmatory factor analysis such as that of Lyne et al. (2006) which has explored the factor structure of the CORE-OM in a very large UK clinical sample. As expected of a short, multi-domain measure, this was a complex structure and we would expect to find similar complexity if not necessarily exactly the same structure when we can collect the much larger Italian datasets needed to explore differences in factor structures of that kind.

For Jacobson's reliable change criterion, based on internal consistency and variance, not repeat data, showed very similar criteria to those from the original UK data since the slightly lower variance in the Italian clinical data largely cancelled out the effects of slightly lower internal consistency of the translation. By contrast, the clinically significant change criteria showed some moderate differences from the UK values with differences varying across domain scores and between men and women. Interestingly, the cutting points for the women showed rather smaller differences between Italy and the UK. This may reflect vagaries of sampling at this stage when the Italian referential data are still relatively small in size. However, it seems entirely plausible that non-clinical and clinical distributions of self-report scores on psychopathology and functioning measures may well differ by language and culture. Further analyses, comparing much larger and representative datasets, both UK and Italian, will throw more light on this and we encourage others working with Italian-speaking clients to contact the first author about possible

DOI: 10.1002/cpp

collaboration in accumulating large anonymous datasets that would start to support the full referential possibilities now being explored in the UK and other countries with the CORE system and other similar systems.

REFERENCES

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., Benson, L., Connell, J., Audin, K., & McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology*, *69*, 184–196.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.

Chiappelli, M., Grigoletti, L., Albanese, P., Taras, M.A., Tulli, P., Grassi, A., & Gruppo I-Psycost. (2007). The cost and utilization of psychotherapy in community-based mental health services. A multicentre study in five Italian areas. *Epidemiologia e Psichiatria Sociale*, *16*(2), 152–62.

Connell, J., Barkham, M., Stiles, W.B., Twigg, E., Singleton, N., Evans, O., & Miles, J. (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.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 297–34.

Derogatis, L.R. (1977). *The SCL-R-90 manual I: Scoring, administration and procedures for the SCL-90*. Baltimore, MD: Clinical Psychometric Research.

Evans, C., Connell, J., Barkham, M., Margison, F., McGrath, G., Mellor-Clark, J., & Audin, K. (2002). Towards a standardised brief outcome measure: Psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, *180*(1), 51–60.

Evans, C., Connell, J., Barkham, M., Marshall, C., & Mellor-Clark, J. (2003). Practice-based evidence: Benchmarking NHS primary care counselling services at national and local levels. *Clinical Psychology & Psychotherapy*, *10*, 374–388.

Evans, C.E., Margison, F., & Barkham, M. (1998) The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-Based Mental Health*, *1*, 70–72.

Evans, C., Mellor-Clark, J., Margison, F., Barkham, M., McGrath, G., Connell, J., & Audin, K. (2000). CORE: Clinical outcomes in routine evaluation. *Journal of Mental Health* , *9*(3), 247–255.

Gallo, E., & Rucci, P. (2000). Supply, demand and predictive factors of psychotherapies in 10 community mental health services in Emilia Romagna. *Epidemiologia e Psichiatria Sociale*, *9*(2), 103–112.

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.

Lomazzi, L., Fava, E., Landra, S., D'Angelo, P., Lammoglia, M., Pazzi, E., Calini, P., Arduini, L., Barattini, D., & Carta, I. (1997). Psychotherapies in psychosocial centers in Lombardia: a point of view of psychiatrists and psychologists, *Epidemiologia e Psichiatria Sociale*, *6*(3), 184–193.

Lyne, J., Barrett, P., Evans, C., & Barkham, M. (2006). Dimensions of variation on the CORE-OM amongst patients attending for psychological therapy. *British Journal of Clinical Psychology*, *45*(2), 185–203.

R Development Core Team (2009). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Shepherd, M., Ashworth, M., Evans, C., Robinson, S., Rendall, M., & Ward, S. (2005). What factors are associated with improvement after brief psychological interventions in primary care? Issues arising from using routine outcome measurement to inform clinical practice. *Counselling and Psychotherapy Research*, *5*(4), 273–280.

Spitzer, R.L, Gibbon, M., Williams, J.B.W., & Endicott, J. (1994). Global Assessment of Functioning (GAF) Scale. In L.I. Sederer & B. Dickey (Eds.), *Outcomes assessment in clinical practice* (pp. 76–78, 182). Baltimore: Williams and Wilkins.

DOI: 10.1002/cpp