

# Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM

Simon Gilbody, David Richards and Michael Barkham

## ABSTRACT

There is increased emphasis on routine assessment of depression in primary care. This report is the first UK validation of two self-completed measures: the Patient Health Questionnaire (PHQ-9) and the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM). Optimum cut-off points were established against a diagnostic gold standard in 93 patients. PHQ-9 sensitivity = 91.7% (95% confidence interval [CI] = 77.5 to 98.3%) and specificity 78.3% (95% CI = 65.8 to 87.9%). CORE-OM sensitivity = 91.7% (95% CI = 77.5 to 98.2%) and specificity = 76.7% (95% CI = 64.0 to 86.6%). Brief self-rated questionnaires are as good as clinician-administered instruments in detecting depression in UK primary care.

## Keywords

depression; diagnosis; primary care; screening.

## INTRODUCTION

Recent guidelines, such as those issued by the National Institute for Health and Clinical Excellence,<sup>1</sup> recommend standardised instruments to improve the recognition and management of depression. The role of standardised assessment of depression is emphasised in the Quality and Outcomes Framework,<sup>2</sup> where routine assessment in established depression, and depression case-finding in diabetes and ischaemic heart disease, are rewarded. Two specific instruments have been proposed:<sup>3,4</sup> the Patient Health Questionnaire (PHQ-9)<sup>5</sup> and the Clinical Outcomes in Routine

Evaluation – Outcome Measure (CORE-OM).<sup>6</sup> The PHQ-9 is a self-administered nine-item depression-specific questionnaire developed in the US.<sup>5</sup> The CORE-OM is a longer 34-item generic instrument developed in the UK, which measures common mental health problems (including four items tapping depression), functional capacity, and risk. Both of these have clear advantages over other instruments, in that they are self-completed and are freely available to end-users. However, there is no published UK primary care validation study of these instruments assessed against a diagnostic gold standard of depression.

## METHOD

A randomised trial was conducted of collaborative care for depression in a UK primary care setting where both the PHQ-9<sup>5</sup> and CORE-OM<sup>6</sup> were used at 3 months' follow-up. At 3 months' follow-up a trained interviewer carried out a diagnostic interview using the Structured Clinical Interview for DSM (SCID),<sup>7</sup> without foreknowledge of PHQ-9 or CORE-OM scores. One hundred and fourteen patients were recruited, of whom 96 were followed up and received a diagnostic interview, PHQ-9, and CORE-OM (22 males, 74 females; mean age 42.5 years, standard deviation 13.6 years).

Sensitivity, specificity, and likelihood ratios for various PHQ-9 scores were calculated (including  $\geq 10$ , as recommended in US primary care<sup>5</sup>) and for the CORE-OM-clinical and depression scores.<sup>6</sup> Receiver operator characteristic (ROC) analysis was also conducted.

## RESULTS

Thirty-six of 96 patients were diagnosed as having major depressive disorder on the SCID. Forty-seven out of the 60 SCID non-depressed patients scored below the recommended PHQ-9 cut off of  $\geq 10$ , giving PHQ-9 sensitivity of 91.7% (95% CI = 77.5 to 98.3%), specificity 78.3% (95% CI = 65.8 to 87.9%), positive likelihood ratio 4.2 (95% CI = 2.6 to 6.9), and negative likelihood ratio of 0.11 (95% CI = 0.04 to 0.32). Increasing the cut-off point of the PHQ-9 to  $\geq 12$  improved the specificity slightly, without compromising

S Gilbody, MD, DPhil, MRCPsych, professor of psychological medicine & health services research; D Richards, PhD, professor in mental health, Department of Health Sciences, University of York, York; M Barkham, PhD, professor of clinical and counselling psychology, Psychological Therapies Research Centre, Leeds.

## Address for correspondence

Professor Simon Gilbody, Department of Health Sciences, University of York, York, YO10 5DD.  
E-mail: sg519@york.ac.uk

Submitted: 14 June 2006; Editor's response: 25 November 2006; final acceptance: 7 December 2006.

©British Journal of General Practice 2007; 57: 650–652.

sensitivity (Table 1). For the CORE-OM-clinical score the optimum cut-off point was  $\geq 13$ , giving the same sensitivity of 91.7% (95% CI = 77.5 to 98.2%), specificity 76.7% (95% CI = 64.0 to 86.6%), positive likelihood ratio 3.9 (95% CI = 2.5 to 6.3), and negative likelihood ratio of 0.10 (95% CI = 0.04 to 0.32). Using the depression subscale (CORE-OM-D) improved the psychometric properties marginally (optimum cut off  $\geq 17$  sensitivity = 94.4%, 95% CI = 81.3 to 99.3%; specificity = 78.3%, 95% CI = 65.8 to 87.9%).

ROC curve analysis indicated that the PHQ-9 and CORE-OM-clinical performed well: PHQ-9 area under the curve (AUC) = 0.94 (95% CI = 0.89 to 0.98); CORE-OM-clinical AUC = 0.92 (95% CI = 0.87 to 0.97); CORE-OM-D AUC = 0.92 (95% CI = 0.86 to 0.97) (Figure 1 and Table 1; see Supplementary Table 1 for detailed psychometric properties, including CIs). At cut offs of  $\geq 10$  and  $\geq 13$  respectively, PHQ-9 and CORE-OM-clinical delivered almost identical sensitivity and specificity.

By applying Bayes' theorem<sup>8</sup> to these performance data, at a commonly encountered baseline prevalence of 10% for depressive disorders, a positive screen of  $\geq 10$  on the PHQ-9 will increase the post-test probability of depression from 10 to 32% (95% CI = 22 to 43%). At a 20% prevalence of depression, typical of that encountered among those with chronic diseases such as diabetes, a positive screen on the PHQ-9 will increase the post-test probability from 20 to 51% (95% CI = 39% to 63%).

## DISCUSSION

This is the first UK validation of the PHQ-9 and CORE-OM against a diagnostic gold standard for depression in a UK primary care population. There are relative merits to both instruments: the PHQ-9 can be self-completed in less than 2 minutes, while the CORE-OM measures a greater range of mental health problems in addition to depression, and assesses

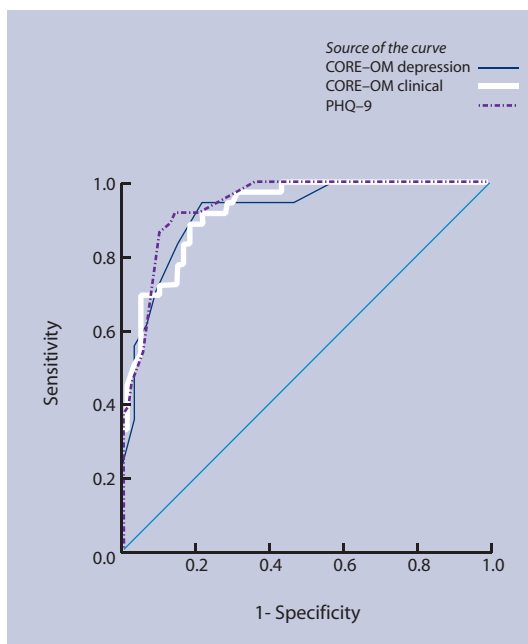


Figure 1. Receiver Operator Characteristic (ROC) curve of CORE-OM (Clinical Outcomes in Routine Evaluation – Outcome Measure) depression scores; CORE-OM clinical scores; and Patient Health Questionnaire (PHQ-9) scores in the presence or absence of major depression diagnosed by the Structured Clinical Interview for DSM.

functional capacity and risk. The cut-off points of  $\geq 10$  for the PHQ-9 and  $\geq 13$  for the CORE-OM achieve good performance, if not better than other longer and clinician-competed instruments.<sup>9</sup> As this is a relatively small cross-sectional study, replication of these findings would be helpful.

While these performance characteristics are

Table 1. Sensitivity, specificity, and likelihood ratios at various cut-off points of the PHQ-9 and CORE-OM.

	Cut off	Sensitivity Specificity		LR+	LR-	AUC (95% CI)
		%	%			
PHQ-9						0.94 (0.89 to 0.98)
	$\geq 9$	94.4	73.3	3.542	0.076	
	$\geq 10$	91.7	78.3	4.231	0.106	
	$\geq 11$	91.7	81.7	5.000	0.102	
	$\geq 12$	91.7	85.0	6.111	0.098	
	$\geq 13$	88.9	86.7	6.667	0.128	
CORE-OM-clinical						0.92 (0.87 to 0.97)
	$\geq 12$	91.7	71.7	3.235	0.116	
	$\geq 13$	91.7	76.7	3.929	0.109	
	$\geq 14$	88.9	80.0	4.444	0.139	
	$\geq 15$	83.3	83.3	4.545	0.204	
	$\geq 16$	75.0	85.0	5.000	0.294	
CORE-OM-depression						0.92 (0.86 to 0.97)
	$\geq 17$	94.4	78.3	4.359	0.071	
	$\geq 18$	88.9	81.7	4.848	0.136	
	$\geq 19$	88.9	81.7	4.848	0.136	
	$\geq 20$	88.9	81.7	4.848	0.136	
	$\geq 21$	83.3	85.0	5.556	0.196	

AUC = area under the curve. CORE-OM = Clinical Outcomes in Routine Evaluation – Outcome Measure. LR+ = positive likelihood ratio. LR- = negative likelihood ratio. PHQ-9 = Patient Health Questionnaire.

## How this fits in

There is an increased emphasis on the recognition and management of depression in the UK under the Quality and Outcomes Framework. PHQ-9 and CORE-OM are brief self-completed instruments, which have been recommended as screening and assessment tools. There is no published report from UK primary care on the performance of the PHQ-9 or CORE-OM as assessment tools for depression. The PHQ-9 and CORE-OM perform as well in UK primary care as clinician-rated instruments. GPs may consider using either the PHQ-9 or CORE-OM alongside enhancements of care for depression.

impressive according to accepted criteria,<sup>10</sup> it is unlikely that case-finding instruments, by themselves, will improve the quality and outcome of primary care for depression.<sup>1,11</sup>

#### Supplementary information

Additional information accompanies this article at <http://www.rcgp.org.uk/bjgp-supinfo>

#### Ethics committee

Not applicable

#### Competing interests

Michael Barkman was involved in the development of the CORE-OM instrument, but does not gain financially from its use. The CORE-OM instrument is free within the NHS.

#### Acknowledgements

We are grateful to the Medical Research Council who funded the initial trial from which these data are drawn, and to our co-investigators of the MRC Collaborative care platform trial group.

#### REFERENCES

1. National Institute for Clinical Excellence. *Depression: core interventions in the management of depression in primary and secondary care*. London: The Stationery Office, 2004.
2. BMA and NHS Employers. *Revisions to the GMS contract, 2006/7. Delivering investment in General Practice*. London: British Medical Association, 2006.
3. Fletcher J, Bower P, Richards D, Saunders T. *Enhanced services specification for depression under the new GP contract*. Manchester: NIMHE Northwest, 2004.
4. Fletcher J, Bower P, Gask L, *et al.* (2006) Primary care service for depression — a guide to best practice. Manchester: Care Services Improvement Partnership (CSIP)/NPCRDC, 2006. <http://www.nimhe.csip.org.uk/silo/files/pcs-for-depressionpdf.pdf> (accessed 9 Jul 2007).
5. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001; **16**(9): 606–613.
6. Barkham M, Margison F, Leach C, *et al.* Service profiling and outcomes benchmarking using the CORE-OM: towards practice-based evidence in the psychological therapies. *J Consult Clin Psychol* 2001; **69**(2): 184–196.
7. Spitzer RL, Williams JB, Gibbon M, First MB. The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Arch Gen Psychiatry* 1992; **49**(8): 624–629.
8. Sackett DL, Haynes RB. Evidence base of clinical diagnosis: the architecture of diagnostic research. *BMJ* 2002; **324**(7336): 539–541.
9. Williams JW, Pignone M, Ramirez G, Stellato CP. Identifying depression in primary care: a literature synthesis of case-finding instruments. *Gen Hosp Psychiatry* 2002; **24**(4): 225–237.
10. Streiner D, Norman G. *Health measurement scales: a practical guide to their development and use*. 3rd edn. Oxford, UK: Oxford University Press, 2003.
11. Gilbody S, Whitty P, Grimshaw JG, Thomas R. Improving the recognition and management of depression in primary care. *Effective Health Care Bulletin* 2002; **7**(Number 5): 1–12.