Ajay Thapar, Gemma Hammerton, Stephan Collishaw, Robert Potter, Frances Rice, Gordon Harold, Nicholas Craddock, Anita Thapar and Daniel J Smith

# Detecting recurrent major depressive disorder within primary care rapidly and reliably using short questionnaire measures


## Abstract

**Background**
Major depressive disorder (MDD) is often a chronic disorder with relapses usually detected and managed in primary care using a validated depression symptom questionnaire. However, for individuals with recurrent depression the choice of which questionnaire to use and whether a shorter measure could suffice is not established.

**Aim**
To compare the nine-item Patient Health Questionnaire (PHQ-9), the Beck Depression Inventory, and the Hospital Anxiety and Depression Scale against shorter PHQ-derived measures for detecting episodes of DSM-IV major depression in primary care patients with recurrent MDD.

**Design and setting**
Diagnostic accuracy study of adults with recurrent depression in primary care predominantly from Wales

**Method**
Scores on each of the depression questionnaire measures were compared with the results of a semi-structured clinical diagnostic interview using Receiver Operating Characteristic curve analysis for 337 adults with recurrent MDD.

**Results**
Concurrent questionnaire and interview data were available for 272 participants. The one-month prevalence rate of depression was 22.2%. The area under the curve (AUC) and positive predictive value (PPV) at the derived optimal cut-off value for the three longer questionnaires were comparable (AUC = 0.86–0.90, PPV = 49.4–58.4%) but the AUC for the PHQ-9 was significantly greater than for the PHQ-2. However, by supplementing the PHQ-2 score with items on problems concentrating and feeling slowed down or restless, the AUC (0.91) and the PPV (55.3%) were comparable with those for the PHQ-9.

**Conclusion**
A novel four-item PHQ-based questionnaire measure of depression performs equivalently to three longer depression questionnaires in identifying depression relapse in patients with recurrent MDD.



**Keywords**
diagnosis; major depressive disorder; primary care; recurrent depression; ROC curve; sensitivity and specificity.


## INTRODUCTION

The importance of major depressive disorder (MDD) is well established.[1] Depression is difficult to diagnose[2] and manage,[3,4] despite major public health and education campaigns. The use of validated depression screening questionnaires in primary care is recommended in the UK.[3] These include the 9-item Patient Health Questionnaire (PHQ-9),[5] the 7-item Hospital Anxiety and Depression Scale depression subscale (HADS-D),[6] and the 21-item Beck Depression Inventory (BDI).[7]

The PHQ-9 and the HADS-D are the most widely used,[8] perform reasonably well as screening instruments for depression,[9–11] but agreement levels for depression severity seem poor,[10] and findings on effectiveness in correctly diagnosing MDD in primary care are mixed.[12–15] Research has shown, despite the widespread use of screening questionnaires in clinical practice, recognition of both first onset and relapse of MDD remains poor,[2] questionnaire depression scores correlate poorly with actual clinical management,[8] and GPs express reservations about the validity and utility of these questionnaire measures.[4,16] However, patients view the score as a tangible measure of their condition.[16,17]

The length of the questionnaires may be an issue given time constraints in general practice,[18] as they can take 3 to 5 minutes to complete, half the primary care consultation length in many countries, and this may be a barrier to integrating scores with clinical assessment.[8] The situation may be even worse for individuals with recurrent depression as GPs may have a broader range of physical and social health problems to deal with.[8,19]

Given that depression symptoms are quite highly correlated, briefer measures of depression may perform adequately to detect MDD and also improve the acceptability and utility of depression measures, especially for those individuals with recurrent depression.[19] PHQ-2 was developed for depression screening,[20] with some evidence for a role in diagnosing depression.[21,22] However, a meta-analysis has highlighted weaknesses in identifying depressive disorder with short measures,[23] and difficulties in establishing an optimal cut-off when used for screening.[24,25]

Finally, the instruments used for screening and picking up new cases of depression may not be the best instruments for use in those individuals with recurrent depression. Monitoring for relapse and remission in those with


**A Thapar**, PhD, MRCGP, GP, professor; **G Hammerton**, BSc, PhD student; **S Collishaw**, PhD, senior lecturer, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff. **R Potter**, MRCGP, MRCPsych, consultant in child and adolescent psychiatry, Child and Adolescent Mental Health Services, Swansea. **F Rice**, PhD, reader in social and emotional development, University College, London. **G Harold,** PhD, professor of developmental psychopathology, University of Leicester, Leicester. **N Craddock**, PhD, FRCPsych, professor; **A Thapar**, PhD, FRCPsych, professor, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff. **DJ Smith**, MD, MRCPsych, reader in psychiatry, Institute of Health and Wellbeing, Glasgow University, Glasgow, UK.

**Address for correspondence**
Ajay Thapar, Taff Riverside Practice, Riverside Health Centre, Wellington Street, Canton, Cardiff, CF11 9SH, UK.
**E-mail:** thaparak@cardiff.ac.uk



**Submitted:** 2 July 2013; **Editor's response:** 17 July 2013; **final acceptance:** 17 September 2013.

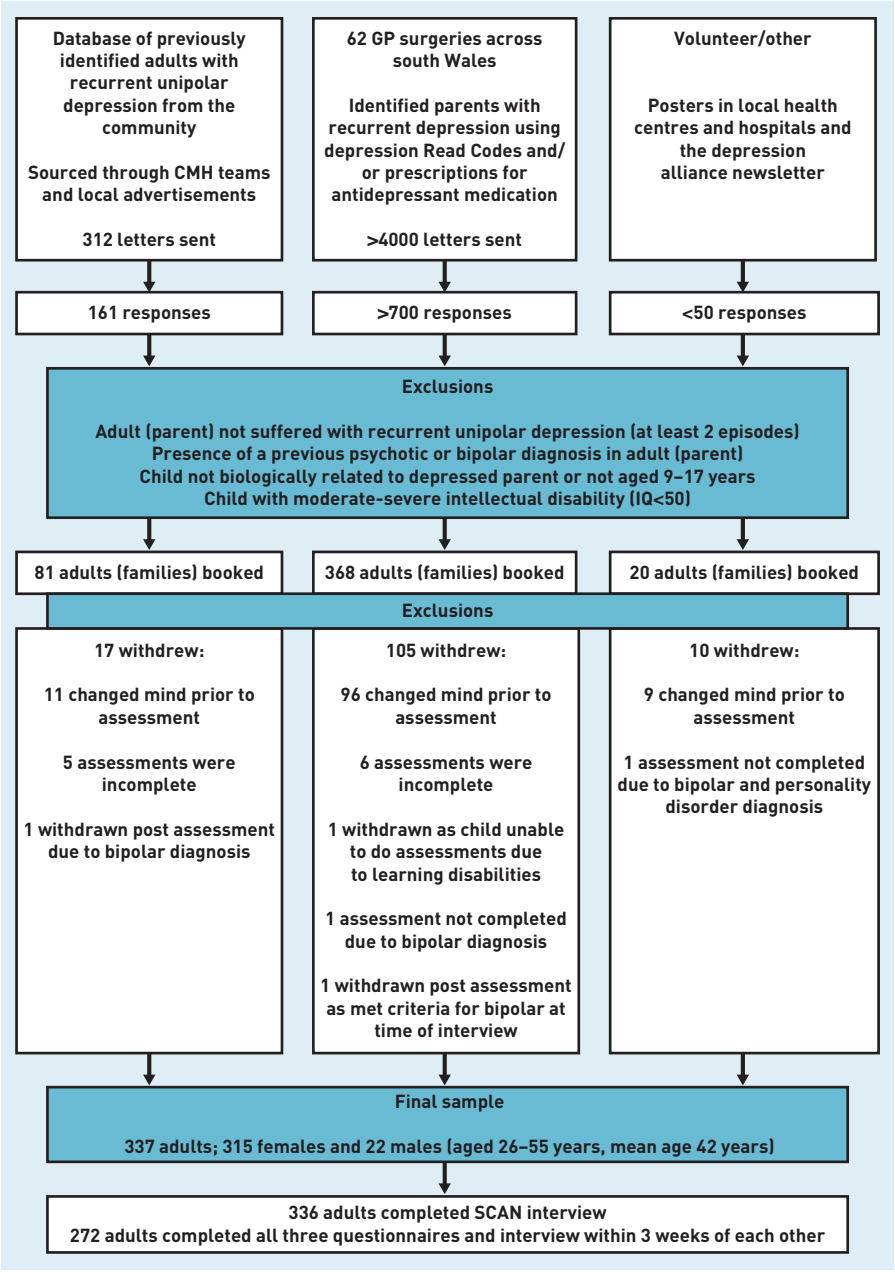


**©British Journal of General Practice**
This is the full-length article (published online 30 Dec 2013) of an abridged version published in print. Cite this article as: **Br J Gen Pract 2014; DOI: 10.3399/bjgp14X676438**.

## How this fits in

Individuals with recurrent depression, apart from being more likely to be depressed, are also more likely to have their depression undertreated. The use of standardised validated questionnaires in primary care has been promoted for adults with depression but whether these inform management is debatable and limited time availability is an issue. Depression relapse in adults with recurrent depression can be diagnosed accurately with a short four-item questionnaire derived from the PHQ-9. A shorter questionnaire may be more easily recalled and scores incorporated into day-to-day clinical management.

*Figure 1. CONSORT diagram: Recruitment.*




recurrent depression is another important role of primary care. To date, there has been no investigation as to what measures work best in primary care for this purpose.

## AIMS

The aims of this study are to establish whether, for a sample of individuals with a history of recurrent depression:

- the HADS-D, PHQ-9 and BDI perform equivalently in accurately identifying MDD relapse as established by a semi-structured psychiatric interview (Schedules for Clinical Assessment in Neuropsychiatry (SCAN)[26];
- the PHQ-2 performs as well as the above longer measures in identifying MDD;
- the performance of the PHQ-2 in identifying MDD can be enhanced using other PHQ-9 items.

## METHOD

### Sample

Three hundred and thirty nine families with a history of recurrent depression in the parent and with children aged 9–17 were recruited into a study examining intergenerational transmission of depression in families predominantly in South Wales. One of the aims of this study was to look at how best to monitor depression in adults with recurrent depression. Parents with recurrent depression were identified by searching their medical records and then carrying out a telephone assessment to ensure they met the criteria for inclusion in the study (at least two episodes of depression requiring antidepressant treatment in the past 5 years).

Two families were subsequently excluded as detailed assessment revealed a history of bipolar disorder. The 337 families included were recruited either through general practices (263 families), through an existing register of depressed adults (64 families) or by other means (10 families). Individuals with a history of bipolar disorder, schizophrenia or schizoaffective disorder were excluded. Ethical approval was obtained from the South East Wales Research Ethics Committee. The methodology has been previously described,[27] and recruitment as per CONSORT guidelines is summarised in Figure 1.

### Measures

All participating adults were mailed and self-completed the PHQ-9, BDI (version 1A) and HADS-D, and these individuals were all also interviewed using the SCAN

by research psychologists. The SCAN interview is a well-established, widely used psychiatric diagnostic interview that assesses symptoms over the last month. Training and monitoring/quality control was carried out by a research psychiatrist. The majority of interviews were conducted within 2 weeks of completing the questionnaires (252/318), 19 individuals did not complete questionnaires and those with an interval of more than 3 weeks were excluded from this analysis ($n$ = 42).

The HADS-D was developed for the assessment of depression in medical outpatients, and has been widely used worldwide.[6] HADS-D scores range from 0 to 21, with scores between 8 and 10 indicating borderline depression, and of 11 or above indicating probable major depressive illness. The PHQ-9 items are based on DSM-IV MDD symptoms and scores range from 0 to 27, with scores of 5–9 indicating mild depression, 10–14 moderate depression, 15–19 moderately severe depression, and scores of 20 or above indicating severe depression.[5] Two items from the PHQ-9, the items relating to 'low mood' and 'loss of interest', constitute the PHQ-2 scale.[21] The final questionnaire measure used in this study was the well-established Beck Depression Inventory Version 1A (BDI-IA).[7] Scores on this scale range from 0 to 63, with scores of 10–18 indicating mild–moderate depression, 19–29 indicating moderate–severe depression, and 30–63 indicating severe depression.

### Analysis

Questionnaire scores (index tests) were compared to a current (last month) episode of DSM-IV depressive disorder, diagnosed using the SCAN interview, using Receiver Operating Characteristic (ROC) curve analysis. ROC curve analysis not only plots the sensitivity and specificity values at different cut-off values on the questionnaire as a curve to identify an optimum cut-off score, but also calculates the area under the curve (AUC), sometimes termed C-index, to assess how well the questionnaire performs overall in correctly identifying depressive disorder (as diagnosed by the psychiatric interview). Scores of above 0.5 on ROC curve analysis indicate the questionnaire performs better than chance in identifying depression and a score of 1.0 indicates it performs perfectly. The AUC (C-index), sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (LR+), and negative likelihood ratio (LR–) were initially derived for different cut-off scores for the PHQ-9, HADS-D, and BDI-IA using STATA (version 11).

In the next stage, the responses on the 2 items constituting the PHQ-2 were compared with the PHQ-9, HADS-D and BDI using ROC curve analysis and reported as described above.

Finally the PHQ-2 scores were supplemented with additional items from the PHQ-9 and systematically compared for performance in identifying MDD against all other measures, again using ROC curve analysis and the measures already described.

A single cut-off score was proposed for each of these questionnaires based on the Youden index (sensitivity + specificity –1)

## RESULTS

Results from 337 families were eligible to be included in the analysis. Adult participants were aged 26–55 years (mean age = 42 years, SD = 5.44) with a history of DSM- IV recurrent MDD (at least two previous episodes of depression); 18 (6.5%) were male and 258 (93.5%) were female.

In total, 336 individuals fully completed a SCAN interview. The results from the first wave of this study were used for the analysis for this paper (December 2007 to April 2009).

Complete interview and questionnaire data were available for 274 individuals for the BDI, 273 individuals for the PHQ and 275 individuals for the HADS-D. The primary analysis was based on the 272 participants where information from all three questionnaires and the SCAN interview was available and interviews and questionnaires were completed within 3 weeks of each other. For these 272 individuals the prevalence rate of a current episode (in the last month) of DSM-IV major depression was 22.2% derived from the SCAN psychiatric interview.

The full results for the initial analyses for the BDI-1A, HADS-D, PHQ-9, PHQ-2, and the '4-item PHQ' are shown in Table 1. The AUCs were similar for the three questionnaires and confidence intervals overlapped, with the AUC for PHQ-9 being highest (0.90), the BDI-1A next (0.89) and finally the HADS-D (0.86).

Further analysis examining modified shortened versions of the PHQ-9 showed that the AUC was best for a 4-item scale (4-item PHQ), which consisted of the PHQ-2 items as well as the items on concentration and feeling slowed down or restless

**Table 1. The area under the curve, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive (LR+), and negative (LR–) likelihood ratios for scores above the cut-off points on the BDI-IA, HADS-D, PHQ-9, PHQ-2 scales, and the 4-item PHQ**

| Measure (*n*) Cut-off | Sensitivity % | Specificity % | PPV % | NPV % | LR+ | LR– | Area under curve (95% CI) Scale developers' categorisation of significance/ severity |
|---|---|---|---|---|---|---|---|
| **BDI-IA (274)** | | | | | | | **0.89 (0.84 to 0.93)** |
| ≥10 | 98.3 | 39.7 | 31.7 | 98.8 | 1.63 | 0.04 | Mild–moderate depression |
| ≥16 | 86.7 | 69.6 | 44.9 | 94.8 | 2.85 | 0.19 | |
| ≥17 | 83.3 | 72.9 | 46.7 | 93.9 | 3.07 | 0.23 | |
| ≥18 | 80.0 | 76.6 | 49.4 | 93.1 | 3.42 | 0.26 | |
| ≥19 | 78.3 | 79.9 | 52.6 | 92.8 | 3.90 | 0.27 | Moderate–severe depression |
| ≥20* | 78.3 | 84.5 | 59.0 | 93.2 | 5.08 | 0.26 | |
| ≥21 | 76.7 | 86.9 | 62.6 | 92.9 | 5.86 | 0.27 | |
| ≥30 | 46.7 | 97.2 | 82.6 | 86.5 | 16.65 | 0.55 | Severe depression |
| **HADS-D (275)** | | | | | | | **0.86 (0.81 to 0.91)** |
| ≥8 | 85.2 | 68.2 | 43.3 | 94.2 | 2.68 | 0.22 | Borderline depression |
| ≥9* | 82.0 | 75.2 | 48.5 | 93.6 | 3.31 | 0.24 | |
| ≥10 | 72.1 | 83.2 | 55.0 | 91.3 | 4.29 | 0.34 | |
| ≥11 | 63.9 | 86.4 | 57.3 | 89.3 | 4.72 | 0.42 | Probable clinical depression |
| ≥13 | 50.8 | 93.5 | 69.0 | 86.9 | 7.77 | 0.53 | |
| ≥14 | 36.1 | 96.7 | 75.7 | 84.1 | 11.0 | 0.66 | |
| **PHQ-9 (273)** | | | | | | | **0.90 (0.86 to 0.94)** |
| ≥5 | 98.4 | 43.4 | 33.1 | 99.0 | 1.74 | 0.04 | Mild depression |
| ≥8 | 91.8 | 65.1 | 42.9 | 96.5 | 2.63 | 0.13 | |
| ≥9 | 88.5 | 72.2 | 47.6 | 95.7 | 3.18 | 0.16 | |
| ≥10 | 85.2 | 76.9 | 51.2 | 94.8 | 3.69 | 0.19 | Moderate depression |
| ≥11* | 83.6 | 83.0 | 58.4 | 94.7 | 4.92 | 0.20 | |
| ≥13 | 77.0 | 87.3 | 63.4 | 93.0 | 6.05 | 0.26 | |
| ≥14 | 70.5 | 91.5 | 70.3 | 91.6 | 8.30 | 0.32 | |
| ≥15 | 62.3 | 92.0 | 69.0 | 89.5 | 7.77 | 0.41 | Moderately severe depression |
| ≥18 | 50.8 | 96.7 | 81.5 | 87.3 | 15.4 | 0.51 | |
| ≥19 | 37.7 | 97.2 | 79.3 | 84.5 | 13.3 | 0.64 | |
| ≥20 | 34.4 | 98.1 | 83.8 | 84.0 | 18.2 | 0.67 | Severe depression |
| **PHQ-2 (273)** | | | | | | | **0.87 (0.82 to 0.91)** |
| ≥2 | 96.7 | 50.9 | 36.0 | 98.0 | 1.97 | 0.06 | |
| ≥3* | 72.1 | 82.1 | 53.5 | 91.2 | 4.02 | 0.34 | |
| ≥4 | 62.3 | 91.5 | 67.7 | 89.5 | 7.34 | 0.41 | |
| ≥5 | 42.6 | 95.3 | 72.1 | 85.3 | 9.04 | 0.60 | |
| **'4-item PHQ' (273)** | | | | | | | **0.91 (0.87 to 0.95)** |
| ≥3 | 93.4 | 67.9 | 45.4 | 97.3 | 2.91 | 0.10 | |
| ≥4 | 83.6 | 80.7 | 55.3 | 94.5 | 4.32 | 0.20 | |
| ≥5* | 77.0 | 87.7 | 64.1 | 93.0 | 6.3 | 0.26 | |
| ≥7 | 59.0 | 95.8 | 80.0 | 89.1 | 13.9 | 0.43 | |

**Optimal cut-off point for defining a DSM-IV diagnosis of depressive disorder based on highest Youden index (sensitivity + specificity –1).*

(AUC = 0.91); the range for all other 4-item combinations was 0.87–0.90.

The only significant difference when the AUC for the PHQ-9 scale (the scale with the highest calculated AUC value) was compared with the AUC for the BDI-1A, HADS-D, and PHQ-2 was between the PHQ-2 and the PHQ-9 ($\chi^2 = 6.41$, $P = 0.011$). If the analysis was repeated using 4 items from the PHQ (4-item PHQ) instead of the PHQ-2 there were no longer any significant differences between any of the areas under the curve. The ROC curves for the BDI, HADS-D, PHQ-9, and the 4-item PHQ are

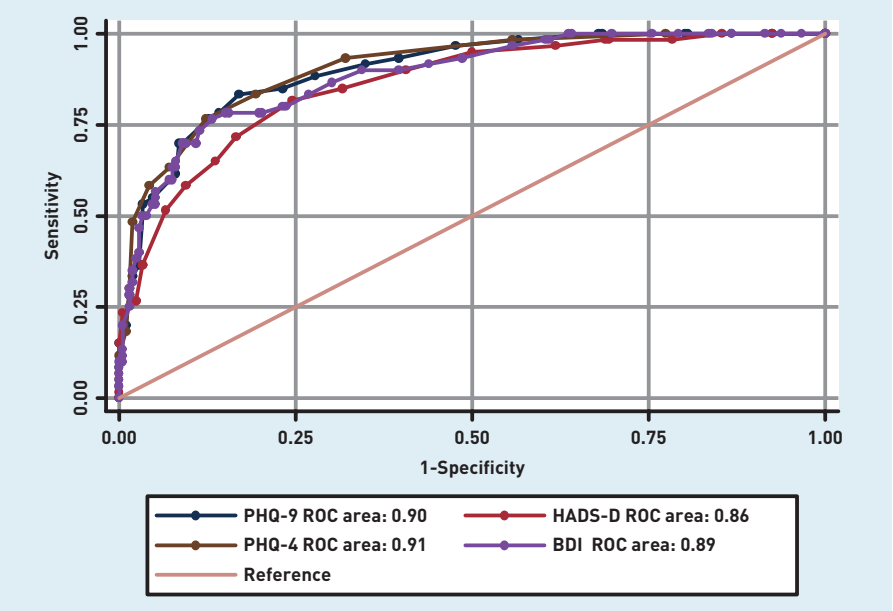



*Figure 2. ROC curves for the BDI 1-A, HADS-D, PHQ-9 and 4-item PHQ scales.*

**Table 2. ROC curve analysis comparing the PHQ-9, HADS-D, BDI, and 4-item PHQ**

| Questionnaire | *N* | Area under curve | Standard error | 95% CI |
|---|---|---|---|---|
| **PHQ-9** | 272 | 0.90 | 0.021 | 0.86 to 0.94 |
| **BDI** | 272 | 0.89 | 0.024 | 0.84 to 0.93 |
| **4-item PHQ** | 272 | 0.91 | 0.020 | 0.87 to 0.95 |
| **HADS-D** | 272 | 0.86 | 0.026 | 0.81 to 0.91 |

*Test for difference in areas of four questionnaires $\chi^2$ (3) = 4.12, P = 0.2483.*

displayed in Figure 2 and the ROC curve analysis is detailed in Table 2.

## DISCUSSION

### Summary

All three questionnaires (BDI, HADS-D, and PHQ-9) performed equally well in identifying MDD relapse, as diagnosed by the SCAN interview, in adults with a history of recurrent major depression, with a single cut-off value (showing both high sensitivity and specificity) for each questionnaire. The shorter PHQ-2 did not perform as well as the longer PHQ-9 in identifying depression. However, adding the items 'impaired concentration' and 'feeling slowed down or restless' to this scale resulted in equivalent performance to the longer questionnaires in identifying depression relapse and a single cut-off value.

### Strengths and limitations

This large primary care sample was carefully screened to ensure they fulfilled the criteria for recurrent depression and then also completed a semi-structured psychiatric interview. A high prevalence rate for a current episode of depressive disorder (22.2%) was reported, which is broadly in line with the view of depression as a chronic disease.[28]

The study had limitations. Only a small percentage of all the families initially approached consented to participate, most of the adults with depression who consented to participate in the study were female, and the sample included individuals with a range of patterns of depression symptoms (episodic to chronic). However, the rates of current depression reported are comparable with those reported for other studies of individuals with recurrent depression. A female excess in cases of depression is representative of GP attenders but as this was a study of parents with offspring aged 9–17, there may be an even greater over-representation of women, therefore caution should be used in applying these findings to other groups such as men with depression or adults without children. The study used the BDI-IA for the current study whereas the BDI-II[29] has generally been recommended (for example, in the UK GP contract).[3] However, scores[30] and the factor structure[31] for the BDI-IA and BDI-II scales are very similar. The order of the depression screening questionnaires in the questionnaire booklet was not randomised so this may potentially have affected results. However, the completion rate for the three questionnaires was very similar. The C-index for different combinations of items in the short scale were not very dissimilar (ranging from 0.87 to 0.91), so specific clinical recommendations would be premature and results would also need replication in another sample to exclude the possibility of over-fitting or chance. However, these results are broadly in line with the findings of other studies.

### Comparison with existing literature

All three standard full-length questionnaires (the BDI, PHQ-9, and the HADS-D) performed very well in screening for depression relapse and also moderately well in correctly identifying depression in the sample. These results are in accordance with the findings of some earlier work,[9–13] but other studies have suggested a role for these questionnaires only as screening instruments rather than as diagnostic tools,[14,15] but differences may relate to the prevalence rates of current depression.[32]

The optimal cut-off values the study determined (using the Youden index) were different from those suggested by the scale

developers. The optimal cut-off levels found in this study for the PHQ-9 (≥11) and the BDI (≥20) were higher than those recommended by the developers (PHQ-9 ≥10,[5] BDI ≥19[7]), whereas for the HADS-D the suggested cut-off (≥9) was considerably lower than that recommended (≥11).[6] Similar but not identical disparities have also been found in other studies.[8] However, there is no single method of determining the optimal cut-off scores on these questionnaires, and these depend on the purpose of the questionnaire (screening or diagnosis), the type of patient, the setting, and the costs and/or benefits of correctly making the diagnosis.[32] The PPV for these depression questionnaires is not often reported in the literature and the only values found were those for the PHQ-9. These results were similar to those found in the study.[13,25]

The study also investigated the possibility of using a shorter measure to identify depression in the sample. In the study the PHQ-2, the most widely evaluated measure,[21] did not perform as well as the PHQ-9 for depression detection. Some studies have found that it is an accurate measure of depression,[22,23,33,34] but other studies have found that very short measures such as the PHQ-2 are only adequate as screening tools.[20,34]

The study found; however, that a 4-item measure derived from the PHQ-9 performed as well as the full-length standard questionnaires in identifying depression in the sample. Other studies have also reported encouraging findings using similar length (4- to 5-item) measures.[35–37] A primary care-based study found that using a two-stage process, with initial screening using the PHQ-2 and subsequent questioning about the presence of other symptoms (sleep disturbance, anhedonia, low self-esteem, and decreased appetite), seemed to explain most of the variance in functioning associated with depression.[35] In another study, in a fairly heterogeneous population, using three of the five non-somatic DSM items (low mood, anhedonia, concentration/indecisiveness, guilt/worthlessness, and suicidal thoughts) accurately identified depression diagnosed by the Structured Clinical Interview for DSM-IV.[36] Finally, in a large outpatient-based study,[37] using the two PHQ-2 items along with 'reduced drive' (as a 'rule-in' item) and impaired concentration (as a 'rule-out' item) resulted in improved performance in detecting depression. Psychomotor retardation had been found to have the highest accuracy as a 'rule-in' item but because of a relatively low prevalence had not been included.[34] In summary, in the above studies three of the four depression symptoms the study identified (mood, anhedonia, and impaired concentration) have been consistently found to be useful in detecting depression, with mixed findings for the fourth item (retardation/restlessness).

### Implications for research and practice

Given that time is a major limiting factor in primary care consultations,[18] the advantages of using a short questionnaire measure for assessing (and monitoring) depression in those with a known history of the disorder are considerable.[23] Given that most patients with depression are entirely managed within primary care, a brief measure that is easy to recall and score may offer considerable advantages. In this study a 4-item scale of items from the PHQ performed as well as the full 9-item PHQ scale (and other standardised questionnaires) in screening for and correctly identifying depression. Two of the items selected (low mood and loss of interest) are used in day-to-day depression screening so are well known by clinicians, so only two other items (problems with concentration and a feeling of being either slowed down or restless) have to be recalled. Using a short measure allows the clinician time to incorporate these responses into the framework of a normal consultation so may be more acceptable and may result in a higher level of use. This brief measure may be even more valuable in the presence of multimorbidity. These findings need to be replicated; however, before recommending more widespread use.


### Funding

This work was supported by the Sir Jules Thorn Charitable Trust. The National Institute for Social Care and Health Research Academic Health Science Collaboration (NISCHR AHSC) funds a fellowship for the corresponding author. SC is funded by The Waterloo Foundation.

### Ethical approval

Ethical approval was obtained from the South East Wales Research Ethics Committee.

### Provenance

Freely submitted; externally peer reviewed.

### Competing interests

The authors have declared no competing interests.

### Acknowledgements

We thank Becky Mars for Figure 1, and along with Ruth Sellers for help in preparation of the data and their comments on the manuscript, and Professor Mike Owen for his comments on earlier versions of the manuscript.


### Discuss this article

Contribute and read comments about this article: **www.bjgp.org/letters**

## REFERENCES


1. Moussavi S, Chatterji S, Verdes E, *et al.* Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet* 2007; **370(9590):** 851–858.
2. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009; **374(9690):** 609–619.
3. British Medical Association and NHS Employers Association. *Quality and Outcomes Framework guidance for GMS Contract 2009/10. Delivering investment in general practice.* London: NHS Employers, 2009.
4. Mitchell C, Dwyer R, Hagan T, Mathers N. Impact of the QOF and the NICE guideline in the diagnosis and management of depression: a qualitative study. *Br J Gen Pract* 2011; **61(586):** e279–e289.
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. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67(6):** 361–370.
7. Beck AT, Steer RA. *Manual for the Beck Depression Inventory.* San Antonio, TX: The Psychological Corporation, 1993.
8. Kendrick T, Dowrick C, McBride A, *et al.* Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. *BMJ* 2009; **338:** b750.
9. Löwe B, Spitzer RL, Gräfe K, *et al.* Comparative validity of three screening questionnaires for DSM-IV depressive disorders and physicians' diagnoses. *J Affect Disord* 2004; **78(2):** 131–140.
10. Cameron IM, Crawford JR, Lawton K, Reid IC. Psychometric comparison of PHQ-9 and HADS for measuring depression severity in primary care. *Br J Gen Pract* 2008; **58(546):** 32–36.
11. Hansson M, Chotai J, Nordstöm A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. *Br J Gen Pract* 2009; **59(566):** e283–e288.
12. Gilbody S, Richards D, Barkham M. Diagnosing depression in primary care using self-completed instruments: UK validation of PHQ-9 and CORE-OM. *Br J Gen Pract* 2007; **57(541):** 650–652.
13. Löwe B, Gräfe K, Zipfel S, *et al.* Diagnosing ICD-10 depressive episodes: superior criterion validity of the Patient Health Questionnaire. *Psychother Psychosom* 2004; **73(6):** 386–390.
14. Wittkampf K, van Ravesteijn H, Baas K, *et al.* The accuracy of Patient Health Questionnaire-9 in detecting depression and measuring depression severity in high-risk groups in primary care. *Gen Hosp Psychiatry* 2009; **31(5):** 451–459.
15. Cameron IM, Cardy A, Crawford JR, *et al.* Measuring depression severity in general practice: discriminatory performance of the PHQ-9, HADS-D, and BDI-II. *Br J Gen Pract* 2011; **61(588):** e419–e426.
16. Dowrick C, Leydon GM, McBride A, *et al.* Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 2009; **338:** b663.
17. Malpass A, Shaw A, Kessler D, Sharp D. Concordance between PHQ-9 scores and patients' experiences of depression: a mixed methods study. *Br J Gen Pract* 2010; **60(575):** e231–e238.
18. Freeman GK, Horder J, Howie JG, *et al.* Evolving general practice consultation in Britain: issues of length and context. *BMJ* 2002; **324(7342):** 880–882.
19. Farmer A, Korszun A, Owen MJ, *et al.* Medical disorders in people with recurrent depression. *Br J Psychiatry* 2008; **192(5):** 351–355.
20. Kroenke K, Spitzer RL, Williams JB. The Patient Health Questionnaire-2: validity of a two-item depression screener. *Med Care* 2003; **41(11):** 1284–1292.
21. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). *J Psychosom Res* 2005; **58(2):** 163–171.
22. Arroll B, Khin N, Kerse N. Screening for depression in primary care with two verbally asked questions: cross sectional study. *BMJ* 2003; **327(7424):** 1144–1146.
23. Mitchell AJ, Coyne JC. Do ultra-short screening instruments accurately detect depression in primary care? A pooled analysis and meta-analysis of 22 studies. *Br J Gen Pract* 2007; **57(535):** 144–151.
24. Arroll B, Goodyear-Smith F, Crengle S, *et al.* Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Ann Fam Med* 2010; **8(4):** 348–353.
25. Zuithoff NP, Vergouwe Y, King M, *et al.* The Patient Health Questionnaire-9 for detection of major depressive disorder in primary care: consequences of current thresholds in a cross-sectional study. *BMC Fam Pract* 2010; **11:** 98.
26. Wing JK, Babor T, Brugha T, *et al.* SCAN. Schedules for Clinical Assessment in Neuropsychiatry. *Arch Gen Psychiatry* 1990; **47(6):** 589–593.
27. Mars B, Collishaw S, Smith D, *et al.* Offspring of parents with recurrent depression: which features of parent depression index risk for offspring psychopathology? *J Affect Disord* 2012; **136(1–2):** 44–53.
28. Judd LL, Akiskal HS, Maser JD, *et al.* A prospective 12-year study of subsyndromal and syndromal depressive symptoms in unipolar major depressive disorders. *Arch Gen Psychiatry* 1998; **55(8):** 694–700.
29. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II.* San Antonio, TX: The Psychological Corporation, 1996.
30. Beck AT, Steer RA, Ball R, Ranieri WF. Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. *J Pers Assess* 1996; **67(3):** 588–597.
31. Steer RA, Clark DA, Beck AT, Ranieri WF. Common and specific dimensions of self-reported anxiety and depression: the BDI-II versus the BDI-IA. *Behav Res Ther* 1999; **37(2):** 183–190.
32. Fletcher RW, Fletcher SW. *Clinical Epidemiology: The Essentials*, 4th ed. Baltimore, MD: Lippincott Williams & Wilkins, 2005.
33. de Lima Osório F, Vilela Mendes A, Crippa JA, Loureiro SR. Study of the discriminative validity of the PHQ-9 and PHQ-2 in a sample of Brazilian women in the context of primary health care. *Perspect Psychiatr Care* 2009; **45(3):** 216–227.
34. Li C, Friedman B, Conwell Y, Fiscella K. Validity of the Patient Health Questionnaire 2 (PHQ-2) in identifying major depression in older people. *J Am Geriatr Soc* 2007; **55(4):** 596–602.
35. Brody DS, Hahn SR, Spitzer RL, *et al.* Identifying patients with depression in the primary care setting: a more efficient method. *Arch Intern Med* 1998; **158(22):** 2469–2475.
36. Zimmerman M, Galione JN, Chelminski I, *et al.* A simpler definition of major depressive disorder. *Psychol Med* 2010; **40(3):** 451–457.
37. Mitchell AJ, McGlinchey JB, Young D, *et al.* Accuracy of specific symptoms in the diagnosis of major depressive disorder in psychiatric out-patients: data from the MIDAS project. *Psychol Med* 2009; **39(7):** 1107–1116.