Research

Michael Moore, Saima Ali, Beth Stuart, Gerry M Leydon, Jessica Ovens, Chris Goodall and Tony Kendrick

Depression management in primary care:

an observational study of management changes related to PHQ-9 score for depression monitoring

Abstract

Background
Since 2009 UK GPs have been incentivised to use depression severity scores to monitor patients' response to treatment after 5–12 weeks of treatment.

Aim
To examine the association between the severity scores obtained and follow-up questionnaires to monitor depression and subsequent changes made to the treatment of it.

Design and setting
A retrospective cohort study utilising routine primary care records was conducted between April 2009 and March 2011 in 13 general practices recruited from within Hampshire, Wiltshire, and Southampton City primary care trusts.

Method
Records were examined of 604 patients who had received a new diagnosis of depression since 1 April 2009, and who had completed the nine-item depression scale of the Patient Health Questionnaire (PHQ-9) at initial diagnosis and a subsequent PHQ-9 within 6 months. The main outcome measure was the odds ratio (OR) for a change in depression management. Change in management was defined as change in antidepressant drug prescription, dose, or referral.

Results
Controlling for the effects of potentially confounding factors, patients who showed an inadequate response in score change at the time of second assessment were nearly five times as likely to experience a subsequent change to treatment in comparison with those who showed an adequate response (OR 4.72, 95% confidence interval = 2.83 to 7.86).

Conclusion
GPs' decisions to change treatment or to make referrals following a second PHQ-9 appear to be in line with guidance from the National Institute for Health and Clinical Excellence for the monitoring of depression in primary care. Although the present study demonstrates an association between a lack of change in questionnaire scores and treatment changes, the extent to which scores influence choice and whether they are associated with improvements in depression outcomes is an important area for further research.

Keywords
depression; patient monitoring; primary care.

INTRODUCTION

Depression is a common illness; in the UK it is largely managed in primary care and has an estimated prevalence of 5–10%.[1] It is recommended that management plans for depression take into account the severity of the illness[2] but has been suggested that, although GPs do this, they are not able to accurately identify those most likely to benefit from treatment.[3]

Since 2006 the UK Quality and Outcomes Framework (QOF) has rewarded GPs for using questionnaire assessments of depression severity at the outset of treatment for patients with a new diagnosis.[4] The rationale is that accurate assessment of severity will allow guidelines from the National Institute for Health and Clinical Excellence (NICE), which recommend different interventions for moderate-to-severe depression than for mild depression, to be implemented. In April 2009, a new indicator, DEP 3, was added to the QOF, rewarding follow-up assessments of severity 5–12 weeks after initial assessment;[5] the rationale for this is that depression is frequently a chronic or relapsing condition that needs monitoring over time. This indicator was modified in June 2011 — the timescale for follow-up was extended slightly from the original 5–12 weeks to 4–12 weeks.

Analyses of primary care records following initial assessment of depression have suggested GPs do not decide on drug treatment or referral on the basis of questionnaire scores alone.[6,7] This may be appropriate as the validity of these tools is measured at the group level, in comparison with a 'gold standard' diagnostic interview, and there will always be false positives and false negatives due to individual variation.[8] There are also a number of other reasons why initial treatment and referral decisions may not be made exactly in line with that which symptom scores suggest.[3,7,9]

Qualitative data from interviews with patients suggest that some consider questionnaires to be useful in terms of providing an insight into the nature and severity of their depression, as well as aiding the diagnosis and an appropriate treatment plan.[7] In addition, some patients requested subsequent assessments in order to monitor their own treatment response and recovery process.[7] Although patients' views of their own progress are broadly in line with follow-up questionnaire scores, there is some concern that tools may fail to recognise, or may miss, symptoms that are relevant to patients. In qualitative interviews, some GPs also reported that they saw value in using scores to monitor patients' progress, although that was not the specific focus of the interviews.[7] Therefore, the trajectory of scores over time may be more important in individual patients than the individual value at any given point.

M Moore, MSc, MRCP, FRCGP, reader; **S Ali**, BSc, MSc, PhD, strategic training fellow in interdisciplinary primary health care research; **B Stuart**, PhD, primary care research fellow; **GM Leydon**, BA, MA, PhD, senior research fellow; **J Ovens**, BMedSci, medical student; **C Goodall**, BMedSci, medical student, Primary and Population Science, University of Southampton, Southampton. **T Kendrick**, MD FRCGP FRCPsych, professor of primary care and dean, Hull York Medical School, University of York, York.

Address for correspondence
Michael Moore, University of Southampton, Primary and Population Science, Aldermoor Close, Southampton, SO16 5ST.

E-mail: mvm198@soton.ac.uk

How this fits in

Depression is a common, and often relapsing, condition. Monitoring of depression is recommended in a number of developed countries and, since April 2009, GPs in the UK have been incentivised to do this using symptom count questionnaires. Inadequate improvements in depressive symptoms, as suggested by scores obtained on the PHQ-9 at follow-up, are associated with changes in depression management. GPs' decisions to change treatment or make referrals appear to be in line with guidance from the National Institute for Health and Clinical Excellence and the rationale for the introduction of the DEP 3 indicator.

To date, there has been little empirical examination of the use of depression severity assessment tools, as specified by the QOF DEP 3 indicator for monitoring depression in primary care. The present study, therefore, aimed to determine whether there is any evidence that GPs change treatment, or decide to refer, on the basis of a change in scores, in line with the rationale for the introduction of the DEP 3 indicator.

METHOD

A retrospective cohort study utilising routine primary care records was conducted between April 2009 and March 2011. Primary care practices were recruited from the Hampshire, Southampton City, and Wiltshire primary care trusts (PCTs). In order to capture a broad range of practice demographics, all 69 practices within these trusts were contacted via post with an invitation to participate in the study. Practices failing to respond to this initial invitation were followed up by telephone. Recruitment was limited to practices using the nine-item depression scale of the Patient Health Questionnaire (PHQ-9) as this questionnaire was used with 75% of patients in a previous study,[6] the instrument has been shown to be sensitive to change for monitoring,[10] and guidance on score interpretation that has been published or is online is readily available.[11,12]

Anonymised data were extracted for patients registered at the recruited practices who had:

- been identified with a new diagnosis of depression since 1 April 2009;
- completed a PHQ-9 questionnaire score

at initial diagnosis; and
- completed a subsequent PHQ-9 questionnaire in line with contract requirements.

Patients were excluded if:

- examination of their medical records suggested a diagnosis of postnatal depression;
- the second PHQ-9 was not completed within 26 weeks of the initial score;
- PHQ-9 scores were non-feasible (that is, a score of >27); or
- scores on both PHQ-9 questionnaires fell below the lower cut-off point for mild depression (PHQ-9 score of <5).

Patients were identified via the practice's QOF DEP 3 indicator patient list; different strategies were used to cope with the coding associated with specific computer systems.

Extracted data included the dates and scores for the initial and follow-up questionnaires, initial management of depression (including antidepressant medication, dosage, and any referrals), and any change in treatment or referral for specialist mental healthcare up to 6 months after the first measure. Additional data included sex, date of birth, the presence of comorbid physical illnesses (diabetes, heart disease, osteoarthritis, chronic renal failure, inflammatory bowel disease, coeliac disease, cancer, chronic respiratory disease, infectious disease, chronic fatigue, and neurological disorders), and whether there was a past history of depression.

Analysis

Logistic regression analysis was used to determine the odds of a change in antidepressant drug prescribing or referral in relation to a change in PHQ-9 score controlling for baseline factors. A fall of five points in the PHQ-9 score is regarded as the minimal clinically significant change,[12,13] which is supported by empirical observation.[14] Changes in questionnaire score between initial and follow-up measures were, therefore, categorised as:

- adequate change — a drop of ≥5 points;
- borderline — a drop of 2–4 points; and
- inadequate change — a drop of 1 point, no change, or an increase in score.

This categorisation allowed the data to be

analysed to compare changes of score with the subsequent management of depression (antidepressant drug prescription, dose, or referral) in the 4 weeks following the second score.

Logistic regression analysis was used to determine the odds of a change in antidepressant drug prescribing or referral between groups with adequate and inadequate treatment responses on the PHQ-9. Models were controlled for age, sex, history of depression, comorbidities, and practice.

Examination of treatment changes related to depression 'caseness', defined by follow-up score, were also completed. A standard cut-off score of <10 was used to determine treatment change according to non-recovery (PHQ-9 score of ≥10) to a non-case level of score (PHQ-9 score of <10).

To map the likely operation and interpretation of the scores in clinical practice, those in the borderline change group were re-categorised as responders if the second score fell below the case threshold (PHQ-9 score of <10), while those with a second score of ≥10 were categorised as non-responders.

Sample size calculation
The incidence of new diagnoses of depression were estimated to be a mean of around 60 per practice per year, as found in a previous study;[6] based on early anecdotal evidence, it was assumed that around 40 per practice would have follow-up scores. It was planned that patient-level information from a minimum of 12 practices would be collected, yielding approximately 500 patients.

RESULTS
Of the 69 practices contacted, 14 (20%) agreed to participate. As one practice had only recently started using the PHQ-9, and the number of patients identified ($n = 4$) was too low for inclusion, data were extracted from the medical records of 13 of the 14 practices. Of the participating practices, nine were in Wiltshire PCT, two in Southampton City PCT, and two in Hampshire PCT. The total list size for all practices that were included was 77 820 (ranging from 3000–15 000 registered patients). The incidence of depression for the QOF year 2010 ranged from 0.3% to 1.5% and, on average, 79% of patients who were eligible for a second PHQ-9 assessment were followed up in accordance with the DEP 3 indicator (ranging from 23% to 100% across practices).
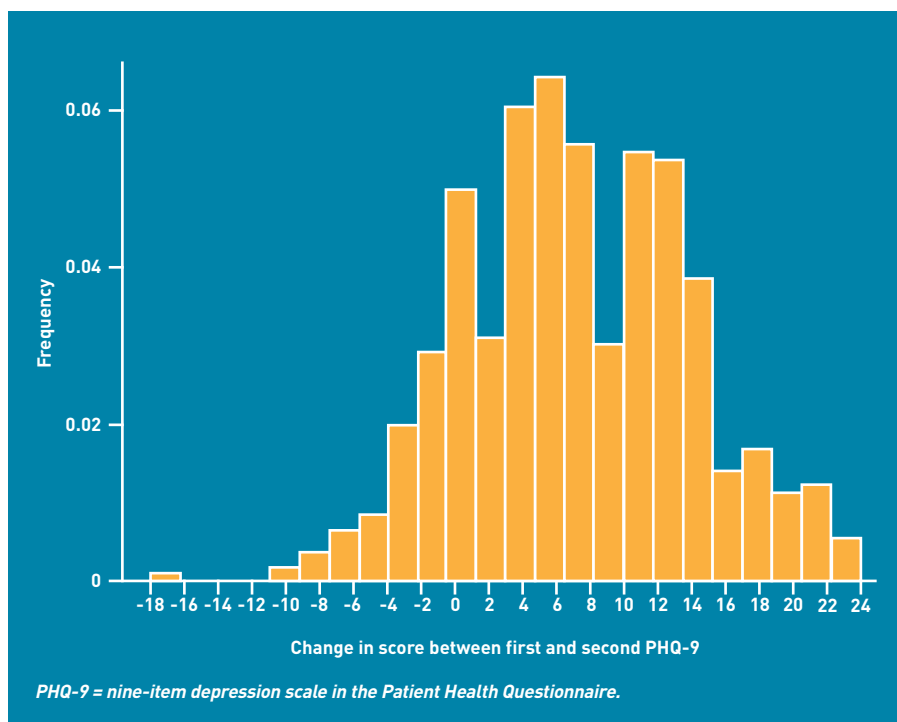
Anonymised data were extracted from 608 patients with a record of two valid PHQ-9 scores in the agreed time frame. Four patients scored below the lower cut-off point of five on the PHQ-9 and, hence, were excluded; this left a final sample of 604 patients.
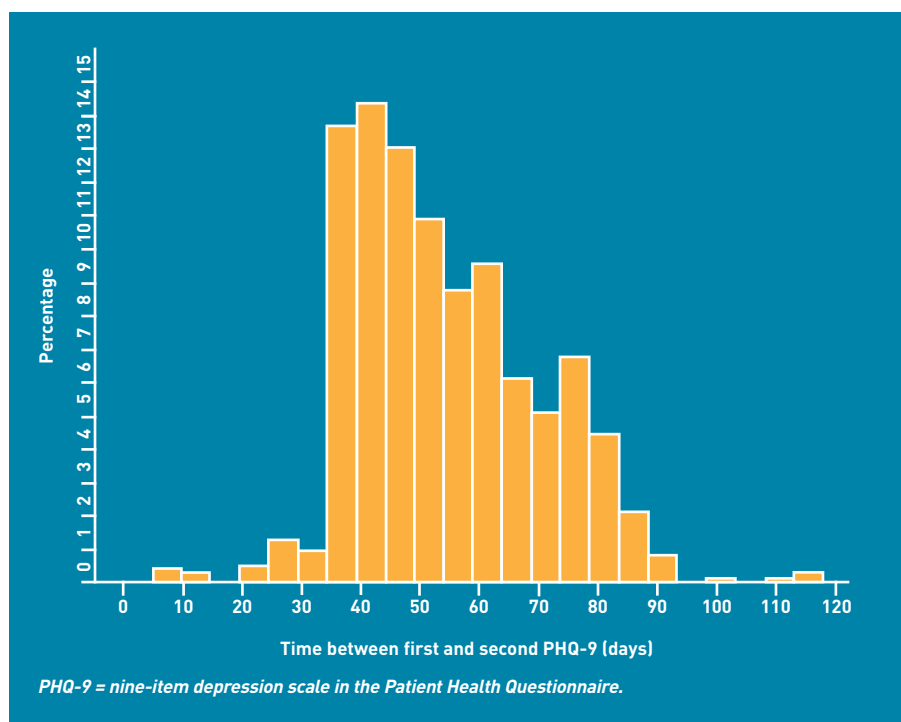
The mean age of the sample was 44.4 years. In total, 418 (69%) patients were female and 216 (36%) had a previous history of depression. One or more comorbidities were present in 106 (18%) patients of the population; 15 (2%) had two comorbidities. No patients were identified as having >3 comorbidities. Using χ^2 tests, no significant differences were observed with regard to treatment response between males and females, those with prior history of depression, or those with comorbidity.

Of the sample, 421 (70%) patients had a follow-up appointment within 4 weeks; the mean number of follow-up appointments in the first 12 weeks was 3.5, and 1.2 in weeks 13–26. The majority of the participants were treated with antidepressant medication — 572 (95%) received at least one drug prescription in the first 16 weeks — and 129 (21%) were referred for a mental health appointment within 16 weeks of diagnosis.

The majority of the sample (95%, $n = 576$) satisfied the case threshold for depression at the initial assessment, whereas, at follow-up, the number reaching case threshold fell to 318 (53%). At follow-up, 379 (63%) showed an adequate treatment response, 97 (16%) a borderline response, and 128 (21%) an inadequate response,

Figure 1. A frequency histogram illustrating the absolute change in PHQ-9 score observed between the first and second scores recorded.

PHQ-9 = nine-item depression scale in the Patient Health Questionnaire.

PHQ-9 = nine-item depression scale in the Patient Health Questionnaire.

Figure 2. A frequency histogram illustrating the recorded time between the first and second of the PHQ-9 pairs.

Figure 3. The odds of experiencing a management change in relation to the absolute score change between the first and second PHQ-9 scores.ᵃ

PHQ-9 = nine-item depression scale in the Patient Health Questionnaire. ᵃControlling for age, sex, comorbid illness, practice, and prior history of depression
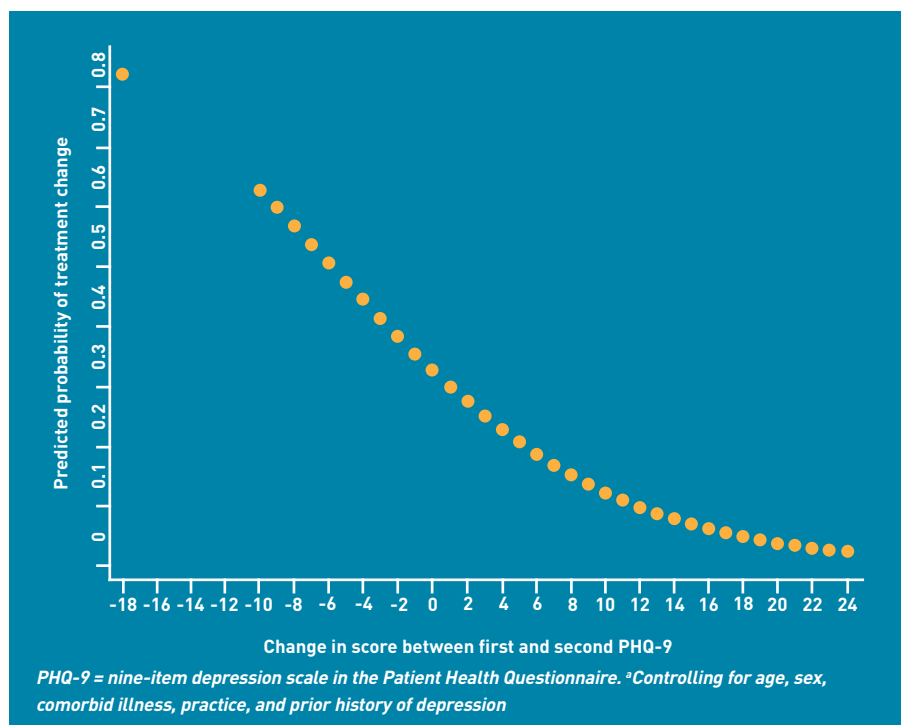
the second PHQ-9 questionnaire was done within 12 weeks of the first and rarely in <35 days (5 weeks, range 5–118 days). Figure 2 shows the time between first and second PHQ-9 questionnaires being done in days.

A management change was recorded in 308 (51%) patients in the 26 weeks of observation following the first PHQ-9 score; 129 (21%) of the total study sample experienced at least one referral, 160 (26%) one drug change, and 118 (20%) at least one dose change. Management changes within 4 weeks of the follow-up PHQ-9 being completed were observed in 119 (20%) patients; these consisted of referral (5%), change in drug (14%), and change in dose (8%), with 20 (3%) patients experiencing >1 management change. On average, the relevant management change following the second PHQ-9 was made after 9 days. However, this mean figure is somewhat skewed by a few changes that were made a considerable time after the PHQ-9 questionnaire was given out.

The median time to treatment change was 0 days — that is, the treatment changes were made on the day that the PHQ-9 was administered. In fact, 87% of changes were made on the same day as the second PHQ-9 and 95% were made within 8 weeks. The majority of changes were made on the same day in all groups: adequate 92%, borderline 84%, and inadequate 76%. Only management changes recorded in the 4 weeks following the second PHQ-9 were included in the subsequent analysis.

Results from the logistic regression, controlling for baseline factors, demonstrated a relationship between the change in PHQ-9 score and management change — for each 1-point increase in the absolute difference between the first and second PHQ-9 scores, the odds of experiencing a management change were reduced by about 12% (Figure 3). Patients who showed an inadequate response in score change at the time of second assessment were nearly five times as likely to experience a management change in the 4 weeks following the second assessment (Table 1).

Similar findings were observed when examining caseness following the second PHQ-9 measure. Looking at those patients who were classified as a case at baseline, those remaining above the case threshold were more than six times more likely to experience a management change in the 4 weeks following the second assessment, compared with those who fell below the threshold (Table 2).

according to the specified definition. Figure 1 illustrates the frequency of the absolute changes observed between patients' first and second PHQ-9 scores.

The second PHQ-9 was administered, on average, 54 days after the first. As the median time between the two was 52 days (interquartile range [IQR] 42–64), on average the second PHQ-9 was completed 7–8 weeks after the first. In 95% of cases,

Table 1. Odds ratios of experiencing a management change in the 4 weeks following the second depression measure

Treatment response	Odds of a treatment change OR (95% CI)	Adjusted odds of a treatment change[a] OR (95% CI)
Adequate[b]	1.00	–
Borderline	2.44 (1.40 to 4.26)	2.38 (1.44 to 3.96)
Inadequate	4.75 (2.97 to 7.63)	4.72 (2.83 to 7.86)

[a]Adjusted for sex, age, comorbid illness, previous history of depression, clustering by practice. [b]Used as a baseline. OR = odds ratio.

Table 2. Odds ratios for experiencing a management change in the 4 weeks following the second depression measure according to case or non-case status

	Odds of a treatment change OR (95% CI)	Adjusted odds of a treatment change[a] OR (95% CI)
Non-case at follow-up[b]	1.00	–
Case at time of second PHQ-9	6.93 (4.07 to 11.80)	6.49 (4.26 to 9.91)

[a]Adjusted for sex, age, comorbid illness, previous history of depression, clustering by practice. [b]Used as a baseline. OR = odds ratio.

Table 3. Odds ratios of experiencing a management change in the 4 weeks following the second depression measure according to treatment response if a borderline response is recategorised according to case status

	Unadjusted odds of a treatment change OR (95% CI)	Adjusted odds of a treatment change[a] OR (95% CI)
Adequate[b]	1.00	–
Inadequate	4.49 (2.95 to 6.84)	4.47 (2.80 to 7.14)

[a]Adjusted for sex, age, comorbid illness, previous history of depression, clustering by practice. [b]Used as a baseline. OR = odds ratio.

The analysis was repeated, recategorising those with borderline response using treatment response defined as PHQ-9 <10 (adequate) versus PHQ-9 ≥10 (inadequate). Those with an inadequate response were still nearly five times more likely to experience a management change in the 4 weeks following the second assessment (Table 3).

All regression analyses controlled for baseline factors, including a prior history of depression and comorbid physical illness. Compared with those patients who had no comorbid condition, those who had a comorbid condition were no more likely to experience a treatment change (adjusted odds ratio 1.11; 95% confidence interval [CI] = 0.41 to 3.14). However those with a previous history of depression were 1.59 times (95% CI = 1.11 to 2.28) more likely to have a treatment change. Looking only at the subgroups, including those with comorbid illness or prior history of depression, the observed relationship between inadequate treatment change and management change still held (data not shown)

DISCUSSION

Summary

Data from this observational study of depression management in primary care demonstrate that there appears to be a clear temporal relationship between treatment changes and an inadequate treatment response when recorded using the PHQ-9. This relationship holds regardless of the method used to estimate treatment failure. Patients with an inadequate score change were significantly more likely to experience a change in their treatment. Patients who scored above the case threshold (score of ≥10) at the time of their second PHQ-9 were also significantly more likely to experience a change to their treatment than those who were not classified as a case at that second time point.

Strengths and limitations

To date there has been little empirical evidence regarding the use of depression monitoring instruments in routine care, and no data examining how scores obtained using such tools are related to the subsequent management of depression in primary care in the UK. The present study, therefore, addresses an important gap in the literature. This study is the first to examine the association between follow-up measurements of depression, as advocated by the QOF DEP 3 indicator and subsequent management of patients. The study involved extracting routine data from 13 practices recruited from within three PCTs and, as such, is likely to reflect standard use in clinical practice.

It is important to be aware of the limitations of the data, which are restricted to patients with paired PHQ-9 scores that are available in the clinical record. Although a precise estimate of the number of patients who were followed for depression is not known, national figures show that follow-up questionnaires were recorded in 70% of depression cases in 2009–2010. The sample probably includes patients likely to be compliant with treatment and the rate of antidepressant prescribing (98%) is higher than that seen in other observational primary care data (80%).[15] As it is not possible to comment on the treatment of those with a single PHQ-9 measure, or draw conclusions with regard to the management of depression in patients who were not followed up, the data should not be taken to represent the global treatment of depression in primary care.

Only 19% of practices that were approached participated in the study. It is likely that participation may have been due

to staff in these practices having a greater interest in depression management than the average. Practices, however, were recruited from a range of locations, encompassing both inner-city and rural areas. Furthermore, it was possible to control for a number of potentially confounding factors.

Comparison with existing literature

The majority of patients were seen within 4 weeks of initial diagnosis, in line with clinical guidelines, and they experienced, on average, 3.46 appointments in the 12 weeks following the first score being completed. A similar figure was observed in a pragmatic study of mild-to-moderate depression in primary care, in which patients in the arm receiving drug treatment were seen on average 4.1 times in the 12 weeks after randomisation.[13]

The results from the present study also support findings from previous research, which demonstrated the influence of questionnaire monitoring from the US. In that study, the PHQ-9 was introduced for monitoring symptoms of depression in a diverse group of psychiatric practices and concluded that scores on the PHQ-9 at follow-up influenced clinical decision making for the majority of 6096 patient contacts.[14]

The PHQ-9 has been used to monitor depression in other studies, but only as a part of a more complex intervention involving care management approaches.[16,17] The benefits of feeding back symptom scores have been shown in specialist psychiatric and psychological care, where a positive effect on mental health was seen in the short term in those patients receiving feedback on their progress.[18,19] However, a recent qualitative study showed that GPs may doubt the validity of these symptom questionnaires,[7] and analysis of records following initial assessment suggested GPs do not decide on initial drug treatment or referral on the basis of questionnaire scores alone, so there are likely to be additional factors influencing treatment choice.

Another study has demonstrated some value in repeated use of the PHQ-9 — following the introduction of the new depression indicator, patients with depression were followed after diagnosis and interviewed at baseline, 3, and 6 months. Although participants described some mismatch between the domains of the PHQ-9 and their personal illness experience, they did request its use to monitor change in illness severity.[20]

It is important to note, however, that, due to the cross-sectional nature of the present study, it is not possible to make interpretations with regard to cause and effect, or to determine the value of follow-up scores to GPs and the extent to which they may impact on clinical decisions at follow-up. Although a previous study has shown poor agreement between practitioners' judgement and formal measures,[3] it is plausible that practitioners used clinical judgment to guide management and that the finding of an association between score and management decision relates to the agreement between the score and clinical judgment. The observed association between the absolute score change and odds of experiencing a treatment change, however, point to a more direct relationship with the score.

Implications for practice

The present study provides evidence regarding the use of depression severity measures to monitor illness severity 5–12 weeks after diagnosis of depression. Those with a poor response to treatment (that is, with either an inadequate change in score or a follow-up score remaining above the case threshold) were five times more likely to experience management changes. The findings show only an association between a lack of change in questionnaire scores and treatment changes and, therefore, cannot determine cause and effect in this study.

Further research is required to determine whether these associations result in improved outcomes for people with depression. Recently updated NICE guidelines emphasise that symptom counts alone are inadequate to assess the severity of depression and that additional factors should be assessed, including the degree of functional impairment and disability.[2] There remains scope to formally test, in a randomised controlled trial, the use of follow-up questionnaires to examine whether they are likely to change practice and improve outcome, and what instrument or combination of instruments should be used.

REFERENCES

1. Singleton N, Bumpstead R, O'Brien M, *et al.* Psychiatric morbidity among adults living in private households, 2000. *Int Rev Psychiatry* 2003; **15(1–2):** 65–73.

2. National Institute for Clinical Excellence. *Depression: the treatment and management of depression in adults.* Report No: CG90. London: NICE, 2009.

3. Kendrick T, King F, Albertella L, Smith PW. GP treatment decisions for patients with depression: an observational study. *Br J Gen Pract* 2005; **55(513):** 280–286.

4. British Medical Association. *Revisions to the GMS contract, 2006/7: delivering investment in general practice.* London: BMA, 2006. http://www.bma.org.uk/employmentandcontracts/independent_contractors/general_medical_services_contract/revisionnGMSFeb20062.jsp#.T3Gnvo6wM_U (accessed 27 Mar 2012).

5. NHS Employers. *Changes to QOF 2009/10.* Leeds: NHS Employers, 2011. http://www.nhsemployers.org/PayAndContracts/GeneralMedicalServicesContract/QOF/Pages/ChangesToQOF200910.aspx (accessed 27 Mar 2012).

6. Kendrick T, Dowrick C, McBride A, *et al.* Management of patients assessed with depression severity questionnaires in UK general practice: analysis of medical record data. *BMJ* 2009; **339:** b750.

7. 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.

8. 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.

9. Johnston O, Kumar S, Kendall K, *et al.* Qualitative study of depression management in primary care: GP and patient goals, and the value of listening. *Br J Gen Pract* 2007; **57(544):** 872–879.

10. Lowe B, Unutzer J, Callahan CM, *et al.* Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care* 2004; **42(12):** 1194–1201.

11. 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.

12. *Patient health questionnaire (PHQ-9).* GP notebook, 2011. http://www.gpnotebook.co.uk/simplepage.cfm?ID=x20090201195702749131 (accessed 10 May 2012).

13. Kendrick T, Chatwin J, Dowrick C, *et al.* Randomised controlled trial to determine the clinical effectiveness and cost-effectiveness of selective serotonin reuptake inhibitors plus supportive care, versus supportive care alone, for mild to moderate depression with somatic symptoms in primary care: the THREAD (THREshold for AntiDepressant response) study. *Health Technol Assess* 2009; **13(22):** iii–iv, ix–xi, 1–159.

14. Duffy FF, Chung H, Trivedi M, *et al.* Systematic use of patient-rated depression severity monitoring: is it helpful and feasible in clinical psychiatry? *Psychiatr Serv* 2008; **59(10):** 1148–1154.

15. Moore M, Yuen HM, Dunn N, *et al.* Explaining the rise in antidepressant prescribing: a descriptive study using the general practice research database. *BMJ* 2009; **339:** b3999.

16. Dietrich AJ, Oxman TE, Williams JW Jr., *et al.* Going to scale: re-engineering systems for primary care treatment of depression. *Ann Fam Med* 2004; **2(4):** 301–304.

17. Williams JW Jr., Gerrity M, Holsinger T, *et al.* Systematic review of multifaceted interventions to improve depression care. *Gen Hosp Psychiatry* 2007; **29(2):** 91–116.

18. Knaup C, Koesters M, Schoefer D, *et al.* Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry* 2009; **195(1):** 15–22.

19. Lambert MJ, Harmon C, Slade K, *et al.* Providing feedback to psychotherapists on their patients' progress: clinical results and practice suggestions. *J Clin Psychol* 2005; **61(2):** 165–174.

20. 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; DOI: 10.3399/bjgp10X502119.