

Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study

Christopher Dowrick, professor of primary medical care,¹ Geraldine M Leydon, senior research fellow,² Anita McBride, research fellow,² Amanda Howe, professor of primary care,³ Hana Burgess, academic foundation trainee,² Pamela Clarke, research assistant,¹ Sue Maisey, research associate,³ Tony Kendrick, professor of primary medical care²

¹University of Liverpool School of Population, Community and Behavioural Sciences, University of Liverpool, Liverpool L69 3GB

²University of Southampton Primary Medical Care Group, Aldermoor Health Centre, Southampton SO16 5ST

³University of East Anglia School of Medicine, Health Policy and Practice, University of East Anglia, Norwich NR4 7TJ

Correspondence to: C Dowrick
cfd@liverpool.ac.uk

Cite this as: *BMJ* 2009;338:b663
doi:10.1136/bmj.b663

ABSTRACT

Objective To gain understanding of general practitioners' and patients' opinions of the routine introduction of standardised measures of severity of depression through the UK general practice quality and outcomes framework.

Design Semistructured qualitative interview study, with purposive sampling and constant comparative analysis.

Participants 34 general practitioners and 24 patients.

Setting 38 general practices in three sites in England: Southampton, Liverpool, and Norfolk.

Results Patients generally favoured the measures of severity for depression, whereas general practitioners were generally cautious about the validity and utility of such measures and sceptical about the motives behind their introduction. Both general practitioners and patients considered that assessments of severity should be seen as one aspect of holistic care. General practitioners considered their practical wisdom and clinical judgment ("phronesis") to be more important than objective assessments and were concerned that the assessments reduced the human element of the consultation. Patients were more positive about the questionnaires, seeing them as an efficient and structured supplement to medical judgment and as evidence that general practitioners were taking their problems seriously through a full assessment. General practitioners and patients were aware of the potential for manipulation of indicators: for economic reasons for doctors and for patients to avoid stigma or achieve desired outcomes.

Conclusions Despite general practitioners' caution about measures of severity for depression, these may benefit primary care consultations by increasing patients' confidence that general practitioners are correct in their diagnosis and are making systematic efforts to assess and manage their mental health problems. Further education of primary care staff may optimise the use and interpretation of depression questionnaires.

INTRODUCTION

Since April 2006 the quality and outcomes framework of the UK general practice contract has offered

financial incentives to general practitioners to measure the severity of depression at the outset of treatment in all diagnosed cases by using validated questionnaires.¹ The questionnaire may be completed during or after the consultation. In line with national guidelines,² the aim is to improve the targeting of treatment for patients diagnosed as having depression, particularly the prescribing of antidepressants to those with moderate to severe depression. The need to improve quality of care for depression was shown by studies done in the first years of the quality and outcomes framework, when depression was not one of the incentivised conditions.³

How general practitioners and patients view the introduction of this quality indicator is as yet unclear. General practitioners tend to confer higher value on the benefits of their own practical wisdom and clinical judgment ("phronesis") than on evidence derived from external sources.⁴ They are less motivated to achieve performance indicators if they dispute the evidence on which these are based.⁵ However, they may also see benefit in the provision of formal evidence in support of diagnosis or assessment, in the same way that they make use of a peak flow reading in asthma or minimal state score in dementia. The organisational literature on outcomes measurement suggests possible unintended consequences of the introduction of quality indicators. Perverse consequences of well intentioned measurement exercises are possible,⁶ including tunnel vision (a concentration on those areas in the outcome set to the exclusion of other important areas), gaming (the alteration of behaviour to gain strategic advantage), and even misrepresentation (including creative accounting and fraud).⁷

Patients' views on this change in general practitioners' activity are equally important. Little evidence exists on how the measures are completed with patients and how useful patients perceive them to be in their diagnosis and management. Although general practice patients may find depression questionnaires acceptable, some have expressed concern about the validity of

the kinds of questions asked and the usefulness of such questionnaires in practice.⁸ We aimed to gain understanding of doctors' and patients' views of the introduction of severity questionnaires for depression and their interpretation in practice.

METHODS

Sampling

The sampling frame was 38 general practices in three locations in England (based around university centres in Southampton, Liverpool, and Norwich) that also took part in a quantitative study of the introduction of questionnaires to assess the severity of depression.⁹ Participating general practitioners identified potential patients to participate during routine consultations and gave them information about the study, including an invitation to contact the study team if they were interested. We recruited other patients by written invitation after a search of practice records, and others volunteered in response to notices in practice waiting areas. Within the total sample of participating doctors and patients, we used a maximum variation approach: for general practitioners, variation was by sex, years of experience, full time/part time practice, trainer/non-trainer status, geographical location, and size of practice; for patients, variation was by sex, age, self defined ethnicity, and socio-demographic background.

Data collection

Three researchers (PC, AMcB, and SM) did open ended, in-depth interviews at a site of the participant's choosing. Interviews used a semistructured topic guide providing broad prompts to explore key issues derived from the literature, focusing exclusively on quality indicators for depression. The prompts included views on intended consequences of the introduction of the depression severity indicator, such as aiming for more active treatment for moderate to severe cases and more watchful waiting for milder cases, and on unintended consequences, such as increase in alternative diagnostic labels. The interview followed the practitioner's or patient's agenda as far as possible, while remaining relevant to the introduction of severity assessment and diagnostic and management decisions related to depression. Prompts were introduced if needed towards the end of the interview to cover relevant areas not already discussed by the participant.

We asked general practitioner respondents for concrete examples to support responses about diagnostic and therapeutic decision making and invited them to discuss possible manipulation and economic considerations. We encouraged patients to describe their feelings at being invited to complete questionnaires to measure severity of depression, their understanding of the meaning of the scores, and their views on whether the treatment of depression should be guided in this way and on the impact of these measures on care. All interviews were audiotaped and transcribed verbatim.

Analysis

AMcB, HB, and GML used the principles of constant comparison to analyse the transcribed interviews.¹⁰ HB led on analysis of the general practitioners' data, and AMcB led on analysing patients' data. Analysis involved deconstructing each interview to identify primary concerns and categories (open coding). These categories were compared with others within the transcript, and across other transcripts, and with concepts within the existing literature. The next stage involved cross linking categories and concepts to generate new meanings and concepts (axial coding). The final stage cross linked the concepts to generate themes (selective coding).

Once HB and AMcB had derived an initial list of themes, each author was allocated 10 to 15 interview transcripts to test the robustness of the proposed thematic scheme. Iterative discussion over many weeks between GML, AMcB, and HB led to consensus themes. CD, TK, and AH then reviewed the clinical coherence and relevance of emergent themes. Data collection and analysis proceeded in an iterative manner, allowing progressive focusing on key themes that explained most of the data together with important deviant cases.

RESULTS

We interviewed 34 general practitioners and 24 patients. Table 1 gives their relevant characteristics. All but one of the general practitioners were interviewed in their own surgeries. Half of the patients were interviewed at home and the others either at work or in their doctor's practice.

We provide an overview of the key themes and concerns as regards participants' views on introducing the measures of severity, including their utility and validity, and on interpreting the measures in practice, including the importance of holism, clinical judgment versus objective assessment, and manipulation.

Introduction of severity measures

We asked general practitioners and patients to reflect on their views on the introduction of the severity measures. Respondents raised many issues about the advantages and disadvantages of using the measures: most felt that this was a complex intervention with considerable challenges to their current practices.

Utility

General practitioner participants could see the utility of introducing severity measures in terms of improving their assessment of depression: "The whole kind of detection and management of depression is something that primary care hasn't been enormously good at historically and I think if we've got a, a tool which helps us, collectively, do it better, then, then I think that's a good thing" (GP07, 273-5); matching interventions to severity and avoiding unnecessary prescribing: "Well maybe about, sort of, more active treatment for people scoring more highly and more watchful waiting for the lower scores, so from that, sort of influencing

prescribing patterns” (GP13, 170-2); and helping to standardise practice: “It probably, maybe it’s a standardisation, isn’t it? Um, it may help people who are less confident and less experienced” (GP17, 65-6).

However, general practitioners consistently set conditions on the possible utility of the measures for diagnosis and depression. The usefulness of the questionnaires was often expressed as contingent on the experience of the doctor—that is, useful for others but not for me: “I’ve got a reasonable amount of experience in mental health . . . I feel reasonably confident in assessing depression anyway. If you’ve got people who are er less um um less happy about their abilities in assessing depression this would be a useful tool” (GP10, 189-92).

Patients were much less likely to make equivocal comments about the utility of the severity instruments. Overwhelmingly, they considered the questionnaires to be useful in assisting the general practitioner in diagnosing and managing depression: “I didn’t understand how you could ask somebody questions and think whether they were depressed or not . . . but then . . . more recently I did it with the [measure] . . . they had a lot more questions and they did it like on the computer and it was a lot better and was more methodical” (PT16, 11-16). Some reflected that the score was valuable in improving targeting of treatment: “Oh yeah because it

depends . . . you should get treatment to . . . correspond with very depressed but if you’re only slightly depressed you know maybe you just need erm counselling or something” (PT01, 58-61).

Some patients reported that the questionnaire had helped to increase their self understanding, so that they could organise their thoughts and express themselves better to their doctor: “I think that [completing the questionnaire] helped me in my head as well . . . Well I started to think, you know about why I was getting depressed and that” (PT19, 45-48); “It helped ’cause some of the things you can’t say how you feel and with answering the questions, it, like, asked you how you felt and things like that and now I think that helped, just like, ticking the boxes, just so that he knew . . . how I felt as well” (PT21, 20-25).

Validity

We asked general practitioners to reflect on the ability of the tools to assist in accurate diagnosis of depression. They expressed uncertainty about the tools’ validity: “I don’t have sufficient confidence that it’s an objective enough tool, really, to measure trends” (GP11, 86-8); “I’ve read recently in the press about the fact that the PHQ-9 compared with the HADS tends to over-estimate depression as opposed to under-estimating depression, I read that yesterday” (GP09, 124-6).

Scepticism about the motives behind the introduction of tools, seeing them perhaps as “academically” oriented and supported by a government wishing to save money by decreasing prescriptions, also shaped some general practitioners’ views on their validity: “I have a horrible feeling that a few academics got together and said this is a good idea and someone at the Department of Health said, oh yes, this is another hoop to make GPs jump through” (GP06, 193-6). Some reflected on the broader and ever-shifting policy context, and this seemed to be an important factor that influenced general practitioners’ belief in the measures: “In the recent past . . . we were clouted round the ear for not prescribing enough in depression and now, three years later, we’re being clouted round the ear for prescribing too much in depression” (GP18, 190-3).

We did not ask patients to comment specifically on the reasons for the introduction of the severity measures, and most interviews with patients were dominated by concerns other than the questionnaires, such as why people become depressed and how doctors should react to this. Some patients expressed scepticism about the validity of the measures. In particular, one patient questioned the reductionist nature of scores in the context of a complex condition such as depression: “I suppose, yeah, it sort of quantifies that you are . . . you do have problems . . . but I, I still feel that it was, like . . . you’re trying to, like, tie a number to a thing which isn’t necessary, isn’t necessarily like a yes or a no . . . it’s a, it’s very difficult to . . . put a description to it, I think” (PT03, 98-101). Some also questioned how scientific the measures were: “I mean didn’t seem very scientific you know questioning like that” (PT13,

Characteristics of general practitioners and patients interviewed. Values are numbers (percentages) unless stated otherwise

Characteristic	General practitioners (n=34)	Patients (n=24)
Sex:		
Male	19 (56)	9 (38)
Female	15 (44)	15 (62)
Age:		
Range	31-62	20-77
Median	45	46
Mean	44	47
Ethnicity:		
White	30 (88)	22 (92)
Asian-Indian	4 (12)	2 (8)
Years in general practice:		
Range	0-28	
Median	15	NA
Mean	14	
Trainer		
	7 (21)	NA
Contract:		
Full time	25 (74)	
Part time	9 (26)	NA
Location:		
City	29 (85)	20 (83)
Town	2 (6)	1 (4)
Village/small town	3 (9)	3 (13)
Size of practice:		
Small	4 (12)	5 (21)
Medium	8 (24)	5 (21)
Large	22 (65)	14 (58)

NA=not applicable.

39-41). In contrast, some patients found the factual and impersonal nature of the tools useful in terms of being able to think about their problems in a new way: “You’ve got to think about . . . um . . . yourself a bit differently, I suppose. Whereas if it had been that the doctor had asked you the questions . . . then it might be a bit different. I think the impersonal nature of the questionnaire’s probably helpful” (PT11, 107-10).

Interpreting the measures in practice

General practitioners and patients were emphatic that severity scores should be understood within a broader context of care, although they tended to place differing emphases on the importance of assessment by questionnaire versus clinical judgment. They were also aware of the potential for manipulating scores.

Importance of holism

All the general practitioners stressed the importance of a holistic approach and evaluating the severity score in the context of the individual patient. They believed that patients’ coping abilities, social situation, and previous mental health battles are all important factors in determining management—specifically, whether to prescribe: “Mental health, mental illness . . . more than most other illnesses are so patient specific and . . . how it affects their lives depends on what they’re doing in their lives, depends on what their background is, might depend on family history and might depend on so many other factors, I think it’s um . . . (completely) impossible to, to mechanise the assessments” (GP04, 158-62). They could see benefits in situations where questionnaires facilitated discussion with patients about diagnosis and treatment options and involving patients in their care, particularly when scores were low and drug treatment was not indicated: “I think for those patients where . . . scores are low and, and they haven’t required medication” (GP07, 230-2).

Patients’ reports generally agreed with doctors on the importance of taking a holistic approach: “Each personality has varying situations and, like, a questionnaire is not going to necessarily give you the full picture of what that person’s circumstances are. It might gloss over some areas that would turn out to be more important than others, I think. So I think . . . probably talking to them, getting a more fuller picture of like how they . . . their lives before that depression and their lives during the depression and afterwards . . . what’s affecting them” (PT03, 199-205). Some patients also considered that the scores were helpful in assessing change in depression over time: “They can see then whether things have improved or not . . . by looking at those [previous results]” (PT23, 65).

Clinical judgment or questionnaire assessment?

General practitioners expressed a strong view that clinical judgment should be the gold standard. Many accepted the scores as part of the clinical assessment, but more as a confirmation of their findings than as an extension of their clinical skills. They expressed concern that the severity measure could be misused

by using it to replace a full history and examination, leading to de-skilling of doctors and a mechanistic approach to assessment: “Yeah I mean the threat the threat is that people will rely on the HAD score as opposed to their own clinical judgment” (GP01, 110-1); “I think we’ll be bringing up a whole . . . generation of doctors who will be very driven by ticking boxes” (GP32, 281-2).

Retaining the essential human elements of patient care was often seen as important for general practitioners: “So whilst I do feel . . . that kind of idea of recipe book medicine, or, or, um . . . if you get this score you do that, you know, is a bit . . . is a bit less human” (GP08, 113-5). Some general practitioners reported that severity scores usually tallied with their own judgment. Where they did not tally, some found the score useful in amending their clinical decisions: “Occasionally I will get a ‘very severe’ come through that I might consider secondary care sooner than previously. But they’re few and far between, um” (GP02, 6-8). However, most were inclined to report ignoring the score and relying on their clinical judgment: “They’re simply part of the assessment of the patient but I wouldn’t put all the weight on the HAD score and I’d be quite happy to ignore it if it didn’t fit” (GP19, 128-9).

Patients, in contrast, tended to see the severity instruments as more important than—or an important adjunct to—the general practitioner’s clinical judgment. They saw them as more efficient and structured, enhancing comprehensive questioning and objectivity: “It’s got to help ’cause then they can see a fuller picture, but yeah . . . so it’s bound to help, other than just asking things that come to mind at the time” (PT22, 86-89); “Well I think it gives them a more accurate picture . . . and as I’ve said already it’s laid out for the doctor so there’s no slip-ups of things being left out because it’s written in black and white” (PT05, 222-5).

Some patients reported a sense of validation, of being taken seriously, and of the benefits of general practitioners taking the initiative to explore the possibility of depression: “It’s probably something the doctor should have asked a long time ago you know ’cause blokes especially are never going to come in and say ooh I’m depressed it’s like come back with a proper illness you know” (PT13, 162-4); “It can be perceived as you’re being taken more seriously I suppose” (PT24, 367-8).

Manipulation

Some reports clearly showed that potential exists for manipulating the measures. Some general practitioners acknowledged being influenced by the implications of the new quality and outcomes framework indicator when recording their initial diagnosis. They reported avoiding coding patients’ symptoms as depression in favour of other diagnostic labels, in order to avoid completing the severity measures and to save time in the consultation: “I think we stop and pause a little bit before we actually put a depression code in. And, of course, there was a mad scramble around the Read codes to find a Read code that wouldn’t get picked up

WHAT IS ALREADY KNOWN ON THIS TOPIC

Under the quality and outcomes framework, UK general practitioners are rewarded for using validated questionnaire measures of severity of depression at the outset of treatment

Doctors tend to give preference to practical wisdom over external evidence, particularly when they are sceptical about the strength of such evidence

Patients welcome the introduction of depression questionnaires into routine practice but may be uncertain about their validity and usefulness

WHAT THIS STUDY ADDS

In relation to the diagnosis and management of depression, general practitioners have a strong preference for clinical judgment over scores on severity measures

In contrast, patients see severity measures as an important adjunct to doctors' clinical judgment because these are efficient and structured and show that problems are taken seriously

General practitioners and patients noted the potential for manipulation of severity scores

by the QOF" (GP28, 199-201); "Low mood, or stress-related problem. I mean, you know, the end point could still be SSRIs... If I haven't got time, if I see it is a recurrence, if I see oh it is a never-ending problem" (GP12, 104-7).

A few patients also referred to the possibilities for manipulation or gaming, although in their case this was related to how honestly they might complete the questionnaire, on the basis of concerns about stigma: "You're more likely to lie, well I found I'm more likely to lie... Because I still find a lot of stigma attached to depression" (PT16, 100-5). Some patients contextualised their "gaming" when completing the questionnaire in their desire to avoid unwanted treatment outcomes, or to achieve their desired outcome: "I said no to anti-depressants as I wasn't depressed. And then she gave me a questionnaire, I remember quite clearly... And I remember reading... and I thought, I'm not putting down how I really feel, and so I didn't" (PT24, 131-42). Two patients noted the financial benefit to general practitioners of introducing the measure, although both assumed that benefit to patients and professionalism would take priority: "If the GPs are totally financial-minded everybody who comes who is a bit tearful or low could get a questionnaire, but I don't know. They're professional people so I don't think they would" (PT24, 280-2).

DISCUSSION

General practitioners were generally more cautious about the validity and utility of measures of severity of depression than were patients, who seemed more enthusiastic. Sometimes, however, general practitioners were surprised by the scores and used them to change clinical behaviour. Both general practitioners and patients considered that assessments of severity should be part of holistic care. General practitioners tended to favour their clinical judgment over questionnaires, whereas patients placed more weight on the value of the score as an objective adjunct to medical judgment and as an expression of general practitioners taking time and trouble to assess them carefully.

Awareness of possibilities for manipulation of indicators existed among general practitioners and patients, although for differing reasons.

Strengths and weaknesses

This is the first study to investigate doctors' and patients' views of the introduction and impact of routine measurement of the severity of depression in primary care. It confirms the view that general practitioners generally place more weight on their own practical wisdom than on the application of externally derived evidence, particularly when they are sceptical about the validity of such evidence.¹¹ It shows that patients, however, may find greater value than their doctors in objective measurement of depression. This may reflect the very different situations of physician and patient—physicians are confident in their own diagnostic ability and expertise, whereas patients fear the stigma and suffering of depression and will see any evidence of their need for care as a supportive measure. The links between this qualitative study and the parallel quantitative study allow for comparison between sources of evidence.⁹ For example, they provide contrast between possible under-treatment of some target groups in the quantitative study and the relative sophistication and knowledge of assessment tools displayed by respondents in this study. This study also shows that the potential for gaming or manipulation of pay for performance indicators extends beyond explicit exception reporting.¹²

Potential biases existed in our recruitment methods. Only interested general practitioners were likely to take part, so they may have tended to express stronger opinions—either positive or negative—than would be the norm. Most of the patients were recruited through general practitioners, who may consciously or otherwise have selected a group of patients relatively sympathetic to the ethos and working practices of general practice. This study did not have access to depression scores for participating patients, so we could not assess whether responses may have varied according to current severity of depression.

As regards study methods, thematic analysis can result in the de-contextualisation of speakers' words, which may misrepresent the intended meaning as they appeared in the original sequential talk. However, we took care to analyse the participants' words in the broader context of the surrounding utterances, to ensure a fair interpretation. As with all interview studies, the kinds of data generated are limited. Interviews provide useful perspectives on events or experiences but cannot be windows on events as they occur. Moreover, interviewees seek to manage the impression they make.¹³ We need to be aware of the complexity of general practitioners' narratives, not least in their tendency towards caution and downgrading of responses in suggesting, for example, that indicators may be more useful for others than for themselves.

Future research

Further analysis will focus on how tools for assessing the severity of depression are deployed in practice—for example, whether they are completed within the consultation or taken away by the patient, and the impact of these different approaches on diagnosis and treatment. More generally, the discrepancy between doctors and patients in the value attached to objective measurement of symptoms is unlikely to be confined to depression: testing this finding in relation to other clinical conditions would be useful.

Impact on policy and practice

Formal scoring systems and assessment tools are becoming more common in primary care: they are used internationally not only in mental health but also in the management of chronic conditions such as diabetes.^{14,15} The findings of both convergence and divergence between doctors' and patients' perspectives are therefore likely to have relevance well beyond the confines of the United Kingdom's quality and outcomes framework indicators for care of patients with depression.

Negative medical views might lead to calls to abandon questionnaires for assessing the severity of depression. However, patients' favourable responses indicate that the quality and outcomes framework's severity measures for depression may have a beneficial impact on primary care consultations, insofar as they confer increased confidence among patients that general practitioners are taking their mental health seriously. Patients may feel an increased sense of validation when such questions are introduced in the consultation. That the severity measures could not stand alone, however, is clear from both patients and doctors—caring and attentive relationships are crucial if the routine use of questionnaires is to be seen as acceptable. Adding a sentence to standard questionnaires to confirm that they are only one aspect of the assessment process may be useful. Education of primary care staff may be necessary to optimise the use and interpretation of depression questionnaires. In future, quality indicators should be piloted before their introduction, so that the effects on patients and practitioners are understood and the necessary education is provided.

We are grateful for the advice we received on the design, conduct, and interpretation of the study from Simon Gilbody, Michael Moore, Robert Peveler, Deborah Sharp, and Andre Tylee. Becky Rowles, clinical studies officer at the Mental Health Research Network North West hub, helped with patient recruitment.

Contributors: TK and CD devised the idea of the study and designed the methods. TK raised the funding. AMcB, PC, and SM were responsible for implementing the study. GML led the analysis with HB and AMcB, with

contributions from CD, TK, and AH. CD prepared the first draft of the manuscript, and all authors contributed to each section of the final draft of the manuscript. CD is the guarantor.

Funding: This study was funded by an unrestricted educational grant from Lilly, Lundbeck, Servier, and Wyeth pharmaceuticals. It also received funding from Southampton City Primary Care Trust and the Mental Health Research Network, East Anglia, West, and North West hubs. None of the above bodies had any role in study design; the collection, analysis, and interpretation of data; the writing of the paper; or the decision to submit this paper for publication. The study was sponsored by the University of Liverpool.

Competing interests: The study was funded by Lilly, Lundbeck, Servier, and Wyeth pharmaceuticals, manufacturers of antidepressants. TK has received fees for presenting at educational meetings from Lilly, Lundbeck, Wyeth, and Pfizer pharmaceuticals. TK and CD are members of the mental health expert panel of advisors for the UK GP contract quality and outcomes framework, which recommended the inclusion of the incentives for using depression severity questionnaires in the contract.

Ethical approval: Liverpool Paediatric Research Ethics Committee gave ethical permission for the study (reference 07/Q1502/23), and approval was obtained from local ethics committees and NHS trust research governance offices at all three sites.

- 1 British Medical Association, NHS Employers. *Revisions to the GMS contract, 2006/7: delivering investment in general practice*. London: BMA, 2006.
- 2 National Institute for Clinical Excellence. *Depression: management of depression in primary and secondary care*. London: British Psychological Society and Gaskell, 2004. (Clinical guideline 23.)
- 3 Steel N, Maisey S, Clark A, Fleetcroft R, Howe A. Quality of clinical primary care and targeted incentive payments: an observational study. *Br J Gen Pract* 2007;57:449-54.
- 4 Hilton S. Education and the changing face of medical professionalism: from priest to mountain guide? *Br J Gen Pract* 2008;58:353-61.
- 5 Maisey S, Steel N, Marsh R, Gillam S, Fleetcroft R, Howe A. Effects of payment for performance in primary care: qualitative interview study. *J Health Serv Res Policy* 2008;13:133-9.
- 6 Nutley S, Smith P. League tables for performance improvement in health care. *J Health Serv Res Policy* 1998;3:50-7.
- 7 Davies HTO, Crombie IK. Interpreting health outcomes. *J Eval Clin Pract* 1997;3:187-9.
- 8 Wood F, Pill R, Prior L, Lewis G. Patients' opinions of the use of psychiatric case-finding questionnaires in general practice. *Health Expect* 2002;5:282-8.
- 9 Kendrick T, Dowrick C, McBride A, Howe A, Clarke P, Maisey S, et al. Management of depression in UK general practice in relation to scores on depression severity questionnaires: analysis of medical record data. *BMJ* 2009;339:b750.
- 10 Strauss A, Corbin J. *Basics of qualitative research: techniques and procedures for developing grounded theory*. 2nd ed. London: Sage, 1998.
- 11 Lorenz KA, Ryan GW, Morton SC, Chan KS, Wang S, Shekelle PG. A qualitative examination of primary care providers' and physician managers' uses and views of research evidence. *Int J Qual Health Care* 2005;17:409-14.
- 12 Doran T, Fullwood C, Reeves D, Gravelle H, Roland M. Exclusion of patients from pay-for-performance targets by English physicians. *N Engl J Med* 2008;359:274-84.
- 13 May C, Dowrick C, Richardson M. The confidential patient: the social construction of therapeutic relationships in general practice. *Sociol Rev* 1996;44:187-203.
- 14 Bushnell J, McLeod D, Dowell A, Salmond C, Ramage S, Collings S, et al. The treatment of common mental health problems in general practice. *Fam Pract* 2006;23:53-9.
- 15 Glasgow R, Wagner E, Schaefer J, Mahoney L, Reid R, Greene S. Development and validation of the patient assessment of chronic illness care (PACIC). *Med Care* 2005;43:436-44.

Accepted: 29 Jan 2009