Patient-reported outcome measures for monitoring primary care patients with depression: the PROMDEP cluster RCT and economic evaluation

Tony Kendrick, Christopher Dowrick, Glyn Lewis, Michael Moore, Geraldine M Leydon, Adam WA Geraghty, Gareth Griffiths, Shihua Zhu, Guiqing Lily Yao, Carl May, Mark Gabbay, Rachel Dewar-Haggart, Samantha Williams, Lien Bui, Natalie Thompson, Lauren Bridewell, Emilia Trapasso, Tasneem Patel, Molly McCarthy, Naila Khan, Helen Page, Emma Corcoran, Jane Sungmin Hahn, Molly Bird, Mekeda X Logan, Brian Chi Fung Ching, Riya Tiwari, Anna Hunt and Beth Stuart

Patient-reported outcome measures for monitoring primary care patients with depression: the PROMDEP cluster RCT and economic evaluation

Tony Kendrick[1*] Christopher Dowrick[2] Glyn Lewis[3]
Michael Moore[1] Geraldine M Leydon[1]
Adam WA Geraghty[1] Gareth Griffiths[4] Shihua Zhu[1]
Guiqing Lily Yao[5] Carl May[6] Mark Gabbay[2]
Rachel Dewar-Haggart[1] Samantha Williams[1] Lien Bui[1]
Natalie Thompson[1] Lauren Bridewell[1] Emilia Trapasso[2]
Tasneem Patel[2] Molly McCarthy[2] Naila Khan[2]
Helen Page[2] Emma Corcoran[3] Jane Sungmin Hahn[3]
Molly Bird[3] Mekeda X Logan[3] Brian Chi Fung Ching[3]
Riya Tiwari[1] Anna Hunt[1] and Beth Stuart[7]

[1]School of Primary Care, Population Health and Medical Education, University of Southampton, Southampton, UK
[2]Department of Primary Care and Mental Health, Institute of Population Health, University of Liverpool, Liverpool, UK
[3]Division of Psychiatry, Faculty of Brain Sciences, University College London, London, UK
[4]Southampton Clinical Trials Unit, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK
[5]Leicester Clinical Trials Unit, University of Leicester, Leicester, UK
[6]Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK
[7]Centre for Evaluation and Methods, Wolfson Institute of Population Health, Faculty of Medicine and Dentistry, Queen Mary University of London, London, UK

*Corresponding author

Disclosure of interests

Disclaimer: This report contains transcripts of interviews conducted in the course of the research, which include language which may offend some readers.

Health Technology Assessment

Criteria for inclusion in the *Health Technology Assessment* journal
Manuscripts are published in *Health Technology Assessment* (HTA) if (1) they have resulted from work for the HTA programme, and (2) they are of a sufficiently high scientific quality as assessed by the reviewers and editors.

Reviews in *Health Technology Assessment* are termed 'systematic' when the account of the search appraisal and synthesis methods (to minimise biases and random errors) would, in theory, permit the replication of the review by others.

HTA programme

Health Technology Assessment (HTA) research is undertaken where some evidence already exists to show that a technology can be effective and this needs to be compared to the current standard intervention to see which works best. Research can evaluate any intervention used in the treatment, prevention or diagnosis of disease, provided the study outcomes lead to findings that have the potential to be of direct benefit to NHS patients. Technologies in this context mean any method used to promote health; prevent and treat disease; and improve rehabilitation or long-term care. They are not confined to new drugs and include any intervention used in the treatment, prevention or diagnosis of disease.

The journal is indexed in NHS Evidence via its abstracts included in MEDLINE and its Technology Assessment Reports inform National Institute for Health and Care Excellence (NICE) guidance. HTA research is also an important source of evidence for National Screening Committee (NSC) policy decisions.

This manuscript

Abstract

Patient-reported outcome measures for monitoring primary care patients with depression: the PROMDEP cluster RCT and economic evaluation

Tony Kendrick,[1*] Christopher Dowrick,[2] Glyn Lewis,[3] Michael Moore,[1]
Geraldine M Leydon,[1] Adam WA Geraghty,[1] Gareth Griffiths,[4]
Shihua Zhu,[1] Guiqing Lily Yao,[5] Carl May,[6] Mark Gabbay,[2]
Rachel Dewar-Haggart,[1] Samantha Williams,[1] Lien Bui,[1]
Natalie Thompson,[1] Lauren Bridewell,[1] Emilia Trapasso,[2]
Tasneem Patel,[2] Molly McCarthy,[2] Naila Khan,[2] Helen Page,[2]
Emma Corcoran,[3] Jane Sungmin Hahn,[3] Molly Bird,[3] Mekeda X Logan,[3]
Brian Chi Fung Ching,[3] Riya Tiwari,[1] Anna Hunt[1] and Beth Stuart[7]

[1]School of Primary Care, Population Health and Medical Education, University of Southampton,
  Southampton, UK
[2]Department of Primary Care and Mental Health, Institute of Population Health, University of Liverpool,
  Liverpool, UK
[3]Division of Psychiatry, Faculty of Brain Sciences, University College London, London, UK
[4]Southampton Clinical Trials Unit, University of Southampton and University Hospital Southampton NHS
  Foundation Trust, Southampton, UK
[5]Leicester Clinical Trials Unit, University of Leicester, Leicester, UK
[6]Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK
[7]Centre for Evaluation and Methods, Wolfson Institute of Population Health, Faculty of Medicine and
  Dentistry, Queen Mary University of London, London, UK

*Corresponding author A.R.Kendrick@southampton.ac.uk

Background: Guidelines on the management of depression recommend that practitioners use patient-reported outcome measures for the follow-up monitoring of symptoms, but there is a lack of evidence of benefit in terms of patient outcomes.

Objective: To test using the Patient Health Questionnaire-9 questionnaire as a patient-reported outcome measure for monitoring depression, training practitioners in interpreting scores and giving patients feedback.

Design: Parallel-group, cluster-randomised superiority trial; 1 : 1 allocation to intervention and control.

Setting: UK primary care (141 group general practices in England and Wales).

Inclusion criteria: Patients aged ≥ 18 years with a new episode of depressive disorder or symptoms, recruited mainly through medical record searches, plus opportunistically in consultations.

Exclusions: Current depression treatment, dementia, psychosis, substance misuse and risk of suicide.

Intervention: Administration of the Patient Health Questionnaire-9 questionnaire with patient feedback soon after diagnosis, and at follow-up 10–35 days later, compared with usual care.

Primary outcome: Beck Depression Inventory, 2nd edition, symptom scores at 12 weeks.

Secondary outcomes: Beck Depression Inventory, 2nd edition, scores at 26 weeks; antidepressant drug treatment and mental health service contacts; social functioning (Work and Social Adjustment Scale) and quality of life (EuroQol 5-Dimension, five-level) at 12 and 26 weeks; service use over 26 weeks to calculate NHS costs; patient satisfaction at 26 weeks (Medical Informant Satisfaction Scale); and adverse events.

Sample size: The original target sample of 676 patients recruited was reduced to 554 due to finding a significant correlation between baseline and follow-up values for the primary outcome measure.

Randomisation: Remote computerised randomisation with minimisation by recruiting university, small/ large practice and urban/rural location.

Blinding: Blinding of participants was impossible given the open cluster design, but self-report outcome measures prevented observer bias. Analysis was blind to allocation.

Analysis: Linear mixed models were used, adjusted for baseline depression, baseline anxiety, sociodemographic factors, and clustering including practice as random effect. Quality of life and costs were analysed over 26 weeks.

Qualitative interviews: Practitioner and patient interviews were conducted to reflect on trial processes and use of the Patient Health Questionnaire-9 using the Normalization Process Theory framework.

Results: Three hundred and two patients were recruited in intervention arm practices and 227 patients were recruited in control practices. Primary outcome data were collected for 252 (83.4%) and 195 (85.9%), respectively. No significant difference in Beck Depression Inventory, 2nd edition, score was found at 12 weeks (adjusted mean difference −0.46, 95% confidence interval −2.16 to 1.26). Nor were significant differences found in Beck Depression Inventory, 2nd Edition, score at 26 weeks, social functioning, patient satisfaction or adverse events. EuroQol-5 Dimensions, five-level version, quality-of-life scores favoured the intervention arm at 26 weeks (adjusted mean difference 0.053, 95% confidence interval 0.013 to 0.093). However, quality-adjusted life-years over 26 weeks were not significantly greater (difference 0.0013, 95% confidence interval −0.0157 to 0.0182). Costs were lower in the intervention arm but, again, not significantly (−£163, 95% confidence interval −£349 to £28). Cost-effectiveness and cost–utility analyses, therefore, suggested that the intervention was dominant over usual care, but with considerable uncertainty around the point estimates. Patients valued using the Patient Health Questionnaire-9 to compare scores at baseline and follow-up, whereas practitioner views were more mixed, with some considering it too time-consuming.

Conclusions: We found no evidence of improved depression management or outcome at 12 weeks from using the Patient Health Questionnaire-9, but patients' quality of life was better at 26 weeks, perhaps because feedback of Patient Health Questionnaire-9 scores increased their awareness of improvement in their depression and reduced their anxiety. Further research in primary care should evaluate patient-reported outcome measures including anxiety symptoms, administered remotely, with algorithms delivering clear recommendations for changes in treatment.

Study registration: This study is registered as IRAS250225 and ISRCTN17299295.

Funding: This award was funded by the National Institute for Health and Care Research (NIHR) Health Technology Assessment programme (NIHR award ref: 17/42/02) and is published in full in *Health Technology Assessment*; Vol. 28, No. 17. See the NIHR Funding and Awards website for further award information.

Contents

CONTENTS

CONTENTS

List of tables

List of figures

List of supplementary material

Report Supplementary Material 1　Statistical Analysis Plan

Supplementary material can be found on the NIHR Journals Library report page (https://doi.org/10.3310/PLRQ4216).

Supplementary material has been provided by the authors to support the report and any files provided at submission will have been seen by peer reviewers, but not extensively reviewed. Any supplementary material provided at a later stage in the process may not have been peer reviewed.

List of abbreviations

BDI-II	Beck Depression Inventory, 2nd edition	MISS	Medical Informant Satisfaction Scale
CEAC	cost-effectiveness acceptability curve	NICE	National Institute for Health and Care Excellence
CI	confidence interval	NIHR	National Institute for Health and Care Research
CONSORT	Consolidated Standards of Reporting Trials	NP	nurse practitioner
CRN	Clinical Research Network	NPT	normalisation process theory
CSRI	Client Service Receipt Inventory	PHQ-9	Patient Health Questionnaire, 9-item
DSM	*Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition	PPI	patient and public involvement
		PROM	patient-reported outcome measure
EQ-5D-5L	EuroQol-5 Dimensions, five-level	PSS	Personal Social Services
GAD-7	Generalised Anxiety Disorder scale, 7-item	PSYCHLOPS	Psychological Outcomes Profile
		QALY	quality-adjusted life-year
GP	general practitioner	QOF	Quality and Outcomes Framework
HRA	Health Research Authority		
IAPT	improving access to psychological therapies	RCT	randomised controlled trial
		REC	Research Ethics Committee
ICER	incremental cost-effectiveness ratio	SD	standard deviation
		SOP	standard operating procedure
MCID	minimal clinically important difference	WSAS	Work and Social Adjustment Scale

Plain language summary

Depression is common, can be disabling and costs the nation billions. The National Health Service recommends general practitioners who treat people with depression use symptom questionnaires to help assess whether those people are getting better over time. A symptom questionnaire is one type of patient-reported outcome measure. Patient-reported outcome measures appear to benefit people having therapy and mental health care, but this approach has not been tested thoroughly in general practice. Most people with depression are treated in general practice, so it is important to test patient-reported outcome measures there, too.

In this study, we tested whether using a patient-reported outcome measure helps people with depression get better more quickly. The study was a 'randomised controlled trial' in general practices, split into two groups. In one group, people with depression completed the Patient Health Questionnaire, or 'PHQ-9', patient-reported outcome measure, which measures nine symptoms of depression. In the other group, people with depression were treated as usual without the Patient Health Questionnaire-9. We fed the results of the Patient Health Questionnaire-9 back to the people with depression themselves to show them how severe their depression was and asked them to discuss the results with the practitioners looking after them.

We found no differences between the patient-reported outcome measure group and the control group in their level of depression; their work or social life; their satisfaction with care from their practice; or their use of medicines, therapy or specialist care for depression. However, we did find that their quality of life was improved at 6 months, and the costs of the National Health Service services they used were lower.

Using the Patient Health Questionnaire-9 can improve patients' quality of life, perhaps by making them more aware of improvement in their depression symptoms, and less anxious as a result. Future research should test using a patient-reported outcome measure that includes anxiety and processing the answers through a computer to give practitioners clearer advice on possible changes to treatment for depression.

Scientific summary

Background

Depression is common and costly. It can lead to chronic disability, poor quality of life, suicide, and high service use and costs. National Institute for Health and Care Excellence guidelines recommend different treatments for more severe and less severe depression, but general practitioners, who treat more the majority of people with depression in primary care, are often inaccurate in their global clinical assessments of depression severity, and treatment is not targeted to patients most likely to benefit. The National Institute for Health and Care Excellence recommends that practitioners consider using validated patient-reported outcome measures to inform treatment at diagnosis and follow-up of people with depression, but there is insufficient evidence that these measures improve depression management and outcomes for patients in primary care.

Aim and objectives

The aim of the study was to answer the research question: What is the effectiveness and cost-effectiveness of assessing primary care patients with depression or low mood soon after diagnosis and again at follow-up 10–35 days later, using the Patient Health Questionnaire-9 combined with patient feedback and practitioner guidance on treatment?

The objectives were to (1) carry out a cluster-randomised controlled trial to compare the intervention with usual care; (2) provide intervention arm patients with written feedback on their Patient Health Questionnaire-9 scores, indicating evidence-based treatments relevant to the level of severity of depression to discuss with practitioners; (3) train practitioners to interpret Patient Health Questionnaire-9 scores and their implications for choice of treatment, taking into account contextual factors; (4) follow up participants for 26 weeks, with research assessments at 12 and 26 weeks; (5) determine the primary outcome of depressive symptoms on the Beck Depression Inventory, 2nd edition, at 12-week follow-up; (6) examine secondary outcomes, including depressive symptoms on the Beck Depression Inventory, 2nd edition, at 26 weeks, and social functioning and quality of life at both 12- and 26-week follow-ups; (7) measure patient satisfaction, adverse events, antidepressant treatment, secondary care contacts, service use, and costs over 26 weeks' follow-up, and perform cost-effectiveness and cost–utility analyses; and (8) carry out a qualitative process analysis to explore participants' reflections on the use of the Patient Health Questionnaire-9 and the potential for implementing it in practice.

Methods

The study design was a parallel-group, cluster-randomised superiority trial with 1 : 1 allocation to intervention and control arms. The setting was UK primary care (141 group general practices in England and Wales). Inclusion criteria were age ≥ 18 years with a new episode of depressive disorder or symptoms. Patients were recruited mainly through regular medical records searches but also opportunistically at consultations for new episodes of depression. Exclusion criteria were current treatment for depression; dementia; psychosis; substance misuse; or a significant risk of suicide.

The intervention was administration of the Patient Health Questionnaire-9 questionnaire as a PROM soon after diagnosis and at follow-up 10–35 days later. Patients were given written feedback on their Patient Health Questionnaire-9 scores and potential treatments to discuss with their general practitioners. Practitioners were trained in interpreting Patient Health Questionnaire-9 scores and taking them into account in treatment decisions.

The primary outcome was depressive symptoms on the Beck Depression Inventory, 2nd edition, at 12 weeks. Secondary outcomes were Beck Depression Inventory, 2nd edition, scores at 26 weeks; social functioning (on the Work and Social Adjustment Scale) and quality of life (on the EuroQol-5 Dimensions, five-level) at 12 and 26 weeks; service use including antidepressant treatment and primary and secondary care contacts over 26 weeks to calculate NHS costs; and patient satisfaction at 26 weeks (on the Medical Informant Satisfaction Scale).

For our sample size calculation, we assumed a baseline mean Beck Depression Inventory, 2nd edition, score of 24.0 with a standard deviation of 10.0 (derived from a feasibility study), and mean scores at 12-week follow-up of 14.0 in the intervention arm and 17.0 in the control arm. The anticipated difference of 3.0 points (effect size of 0.3) represented the minimum clinically important difference on the Beck Depression Inventory, 2nd edition. At the 5% level of significance, to have 90% power to detect that difference we calculated we needed 235 patients analysed per arm. We aimed to recruit a mean of six patients per practice and assumed an intracluster correlation coefficient of 0.03 (from the feasibility study), which gave a cluster design effect of 1.15, meaning we needed 270 per arm. We assumed a 20% loss to follow-up at 12 weeks, so the total sample size needed was 270 × 2/0.8 and our original target sample size was a total of 676 patients recruited, from 113 practices, by three recruitment centres (the University of Southampton, the University of Liverpool and University College London). We subsequently revised the target sample size on finding a significant correlation coefficient of > 0.5 between baseline and follow-up values for the primary outcome, which meant that we needed only 222 patients analysed per arm and, therefore, a target sample size of 554 patients recruited (revised 10 June 2021).

Cluster randomisation of practices to intervention and control arms was carried out remotely by a Clinical Trials Unit statistician using computerised sequence generation, with minimisation by recruiting centre, size of practice and urban or rural location. Blinding of participating practitioners and patients to allocation was impossible given the nature of the intervention and the cluster-randomised design, but self-report outcome measures were used to prevent researcher rating bias, and statistical analysis was blind to allocation.

Differences between intervention and control arms in the outcomes of depressive symptoms, social functioning and quality of life measured at 12- and 26-week follow-up were analysed using linear mixed models, adjusting for baseline depression; duration of depression; history of depression; baseline anxiety; sociodemographic factors (gender, age, socioeconomic position, housing, education, marital status and dependants), and clustering including a random effect for practice. Patient satisfaction, quality of life (quality-adjusted life-years) and costs were compared between the arms over the 26 weeks' study follow-up period. Differences between the arms in the process of care for depression were also analysed, including patients' self-reported use of antidepressants at the 12- and 26-week follow-up

points, and medication and contacts with mental health services (community mental health nurses, counsellors, psychologists, psychiatrists, other therapists and social workers) recorded in practice medical records over the 26 weeks' follow-up.

A health economic evaluation was undertaken from an NHS and Personal Social Services perspective. The outcomes were expressed as incremental cost per point improvement in the Beck Depression Inventory, 2nd edition, clinical outcome (cost-effectiveness analysis), and incremental cost per quality-adjusted life-year gained (cost–utility analysis). The primary analysis at 26 weeks used a generalised linear mixed model to estimate the differences in costs and quality-adjusted life-years (using the EuroQol-5 Dimensions, five-level to calculate patient utilities), adjusted for baseline quality of life; baseline anxiety; sociodemographic factors; and practice as a random effect. Incremental cost-effectiveness ratios and a cost-effectiveness acceptability curve were generated using non-parametric bootstrapping.

Qualitative interviews with participating practitioners and patients in both arms were conducted to reflect on their involvement in the trial and analysed using reflexive thematic analysis. Intervention arm participants were asked about barriers, facilitators, benefits and problems related to using the Patient Health Questionnaire-9, including questions derived from the normalisation process theory framework.

Results

Practices and patients
As the number of patients recruited per practice was smaller than anticipated, we recruited significantly more than our target of 113 practices, eventually reaching a total of 189, but 48 practices subsequently withdrew (24 in each arm), so the final number of active practices was 141: 72 intervention and 69 control (28 above our original target). Practice characteristics were well balanced by arm.

Of 11,468 patients approached in consultations or through mailed invitations, 1058 (9.2%) returned reply slips about the study: 574 (10.6% of those approached) in the intervention arm and 484 (8.0% of those approached) in the control arm. After the exclusion of patients declining to participate, ineligible at screening or uncontactable, 529 patients were assessed at baseline: 302 (5.5% of those approached) in the intervention arm and 227 (3.8% of those approached) in the control arm. The ratio of intervention to control arm patients recruited was, therefore, 1.3 to 1, which may have reflected lower motivation to take part among control arm practices. Of 529 patients recruited, 453 (85.6%) were followed up at 12 weeks: 254 intervention arm (84.1%) and 199 control arm (87.7%) patients. At the 26-week point, 414 patients (78.3%) were followed up: 230 intervention arm (76.2%) and 184 control arm (81.1%). Medical records data were collected for 259 intervention arm patients (85.8%) and 201 control arm patients (88.5%).

The mean BDI-II score for depressive symptoms at baseline was higher in the intervention arm, at 24.1 (standard deviation 8.89) than in the control arm, at 22.4 (standard deviation 9.52). Baseline anxiety and quality-of-life scores were also worse in the intervention arm. Control arm patients were more likely to have had two or more previous depressive episodes. Demographic characteristics were relatively well balanced.

Clinical outcomes
At the 12-week follow-up, the mean Beck Depression Inventory, 2nd edition, score was 18.5 (standard deviation 10.2) in the intervention arm and 16.9 (standard deviation 10.3) in the control arm. The adjusted mean score was slightly lower in the intervention arm, but this was not statistically significant (mean adjusted difference −0.46, 95% confidence interval −2.16 to 1.26; p = 0.60). At 26 weeks, the mean Beck Depression Inventory, 2nd edition, scores were 15.1 (standard deviation 10.8) in the intervention arm and 14.7 (standard deviation 10.6) in the control arm (mean adjusted difference −1.63,

95% confidence interval −3.48 to 0.21; *p* = 0.08). Social functioning on the Work and Social Adjustment Scale and Medical Informant Satisfaction Scale satisfaction with care scores favoured the intervention, but the differences found were not statistically significant.

A post hoc analysis at 26 weeks showed similar proportions improving by ≥ 50% on the Beck Depression Inventory, 2nd edition, in the intervention and control arms (45.1% vs. 37.3%), but the proportion remitting to a score of < 13 was significantly greater in the intervention arm (49.8% vs. 39.9%; adjusted odds ratio 2.18, 95% confidence interval 1.12 to 4.24; *p* = 0.02).

Process of care
In the intervention arm, 190 out of 261 patients (72.8%) had Patient Health Questionnaire-9 results in their medical records. In the control arm, 35 out of 201 patients (7.4%) had Patient Health Questionnaire-9 results recorded.

More patients in the intervention arm had antidepressant prescriptions recorded in their medical records over the 26 weeks' follow-up (67.4% vs. 55.7%), but the adjusted difference was not statistically significant. There was also no significant adjusted difference found in the proportions with mental health or social services contacts over the 26 weeks (34.6% vs. 33.8%, respectively).

Health economic outcomes
The adjusted mean difference in utility score between the arms was not statistically significant at the 12-week follow-up, but a statistically significant difference favouring the intervention arm was found at 26 weeks (0.053, 95% confidence interval 0.093 to 0.013; *p* = 0.01). However, quality-adjusted life-years over 26 weeks were not significantly greater (adjusted mean difference 0.0013, 95% confidence interval −0.0157 to 0.0182). Costs were lower in the intervention arm, but again not significantly (adjusted mean difference −£163, 95% confidence interval −£349 to £28). Cost-effectiveness and cost–utility analyses therefore suggested that the intervention was dominant over usual care, but with considerable uncertainty around the point estimates. The cost-effectiveness acceptability curve showed the probabilities of the intervention being cost-effective compared with usual care, at societal willingness-to-pay thresholds of £20,000 and £30,000 per quality-adjusted life-year, were 77% and 72%, respectively.

Qualitative interviews
Practitioners and patients interviewed described various benefits of using the Patient Health Questionnaire-9, including the provision of information about the range of symptoms and severity categories of depression; highlighting particular symptoms, including suicidal thoughts; identifying changes in mood over time; and informing treatment plans. However, a number of practitioners stated that their own clinical judgement was more important in making management decisions.

Some patients described the Patient Health Questionnaire-9 as oversimplifying their complex experiences of depression; and some practitioners did not like the rigidity of the severity categories and their associated suggested treatments, which they referred to as 'tick-box medicine'. Several practitioners expressed resistance to using the Patient Health Questionnaire-9 for these reasons, although some suggested it could be used as a guide for general practitioners with less experience.

Barriers to using the questionnaire in routine practice included the time taken up in the consultation, which practitioners considered could be reduced if administering the questionnaire were automated through technological integration into practice communication and computerised records systems. Practitioners wanted an evidence base that the questionnaire was effective; clearer guidance on what to do depending on patient scores; and remuneration for the extra time taken in consultations.

Limitations

Baseline differences in depression, anxiety and quality-of-life scores may have reflected selection bias due to the cluster randomised design.

We did not quite achieve the revised sample size target of 554 patients, falling short by 25, but the follow-up rate of 84.5% was better than the 80% predicted and so we gathered primary outcome data on 447 patients, exceeding the target of 444 and sufficient to answer the main research question with precision.

It was not possible to blind participants and researchers to allocation to intervention or control arms given the pragmatic open and cluster randomised design, but we used self-report measures to avoid observer bias, and the analyses were carried out blind to allocation.

We endeavoured to carry out the baseline assessments and administer the first Patient Health Questionnaire-9 as soon as possible after the patients first presented symptoms, but this was sometimes delayed by 2–3 weeks. In the meantime, treatment had already been started by the general practitioner/nurse practitioner of around half of the patients, which meant that the first Patient Health Questionnaire-9 score could not be taken into account when choosing initial treatments.

Conclusions

We found no benefit from using the Patient Health Questionnaire-9 in relation to the primary outcome of depression on the Beck Depression Inventory, 2nd edition, at the 12-week follow-up. There were also no significant differences found between the arms in the secondary outcomes of Beck Depression Inventory, 2nd edition, scores at 26 weeks, work and social functioning, patient satisfaction, medication use, or contacts with mental health services, although all the differences found in these measures were in the direction of favouring the intervention arm.

However, we did find a significant benefit in terms of improved quality of life at 26 weeks, for similar overall service costs. We also found evidence of benefit in a categorical analysis comparing rates of remission of depression at 26 weeks, although this result should be treated with caution as it was from a post hoc analysis. Cost-effectiveness acceptability curves showed that the probability of the intervention being cost-effective, at the lower and higher thresholds adopted by the National Institute for Health and Care Excellence of £20,000 and £30,000 per quality-adjusted life-year, was 77% and 72%, respectively.

We found that patients valued using the Patient Health Questionnaire-9 to identify changes in their scores. The mechanism by which feedback of scores might improve patients' quality of life, despite not changing the management of their depression, might be through increasing their awareness of improvement in their symptoms over time, supporting personal reflection on their progress to recovery.

Future work

Further research should be conducted in primary care evaluating (1) longer patient-reported outcome measures including anxiety symptoms, (2) administered remotely before and between consultations, with (3) algorithm-driven interpretation, delivering recommendations for changes in treatment.

Study registration

This study is registered as IRAS250225 and ISRCTN17299295.

Funding

Chapter 1   Introduction

Some text in this chapter has been adapted from the study protocol published as Kendrick T, Moore M, Leydon G, Stuart B, Geraghty AWA, Yao G, *et al*. Patient-reported outcome measures for monitoring primary care patients with depression (PROMDEP): study protocol for a randomised controlled trial. *Trials* 2020;**21**:441 (https://doi.org/10.1186/s13063-020-04344-9). This article is published under licence to BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution and reproduction in any medium provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article unless otherwise stated.

Description of the health problem

Depression is common and costly. In 2014, the 1-week prevalence among adults in the UK was 11.1%, comprising 3.3% major depression and 7.8% mixed depression and anxiety,[1] and since then it has increased, particularly during the coronavirus disease 2019 (COVID-19) pandemic.[2] Depression can lead to chronic disability, poor quality of life, suicide, and high service use and costs. The King's Fund estimates that 1.45 million people will have depression in England by 2026 and that the annual costs of care, social services and lost employment will be £12.2 billion.[3]

The prevalence of depression has continued to increase despite big increases in antidepressant prescribing[4] and psychotherapy[5] for depression in England, as in other countries. One of the main reasons for this increase is a lack of application of evidence-based treatments to those who would benefit – the so-called 'quality gap'.[6] Only around half of the people in the 2014 community survey found to have depression and anxiety were receiving treatment.[1]

National Institute for Health and Care Excellence (NICE) guidelines recommend different treatments for more severe depression (e.g. combined antidepressants and cognitive–behavioural therapy) from those for less severe depression (e.g. guided self-help or exercise).[7] However, general practitioners (GPs), who treat > 70% of cases in primary care, are often inaccurate in their global clinical assessments of depression severity, and so treatment is not targeted to patients most likely to benefit.[8] Some patients receive treatment they do not need (medicalising self-limiting illness and exposing them to antidepressant side effects) and others do not get treatment they do need, significantly contributing to the 'quality gap'. A systematic review concluded that there are many false diagnoses as well as missed cases, which could be improved by reassessment of individuals who might have depression.[9] As a result, NICE recommends that practitioners consider using a depression symptom questionnaire as a measure of severity at diagnosis and follow-up to inform and evaluate treatment.[7] A symptom questionnaire is an example of a patient-reported outcome measure (PROM), and PROMs have been promoted to increase patient involvement in their care.[10]

Current evidence on the effectiveness of patient-reported outcome measures in depression

Meta-analyses[11,12] of trials of using PROMs routinely in mental health and psychological therapy settings in Europe and the USA until 2010 reported benefits in terms of depression outcomes through the improved identification of clients whose progress was slower than expected ('not on track') early in the course of therapy and targeting extra therapy accordingly.

A 2016 Cochrane review[13] led by PROMDEP chief investigator Tony Kendrick, comprising 17 studies and 8787 participants, evaluated the effect of using PROMs to monitor progress in patients with common mental health disorders across all settings, including primary care. Overall, there was a lack of evidence that using PROMs improved patient symptoms or led to changes in symptom management over the course of treatment. The quality of evidence was low, as all included studies were assessed as being at high risk of bias with considerable attrition at follow-up. There was virtually no evidence on the effects on health-related quality of life, social functioning, adverse events or costs. A post hoc subgroup meta-analysis[13] of studies in mental health and psychological therapy settings (10 studies, 923 participants) found evidence of improvement in symptom scores in 'not on track' cases, although the benefit was small, with an effect size of around 0.2. In another subgroup analysis, PROM use also appeared to result in shorter treatment for 'on track' patients, thereby increasing service efficiency.

More recent research conducted in the improving access to psychological therapies (IAPT) services in England found similar benefits of routine outcome monitoring with PROMs combined with systematic feedback to practitioners in terms of targeting treatment, potentially improving outcomes for 'not on track' clients[14] and saving resources by shortening therapy for 'on track' clients.[15]

Evidence in primary care

Evidence of benefit from psychological therapy and mental health service settings may not, however, generalise to primary care, where only a proportion of patients have mental health problems and so embedding the routine administration of PROMs for depression to all patients presenting to a practice is not justified.

Only two studies in the 2016 Cochrane review were conducted in primary care; both took place in the USA, and they reported conflicting findings. One study of using the Hopkins Symptom Checklist-90[16] at diagnosis and follow-up with 587 patients with depression and anxiety found changes in the process of care in terms of increased referrals for mental health treatment but no improvement in outcomes for patients.[17] The other trial, in which 642 patients with depression were monitored using the Patient Health Questionnaire-9 (PHQ-9),[18] found an improved outcome in terms of depressive symptoms at follow-up[19] but no obvious changes in antidepressant prescribing or referrals to secondary care to explain the improvement.[20]

More recently, a cluster-randomised trial in 22 Swedish primary healthcare centres comprising 258 patients with depression found that those who were monitored with the Montgomery–Åsberg rating scale[21] were more likely to adhere to antidepressant treatment, but there was no demonstrable improvement in the outcome of their depression.[22] The study was relatively small and might have been underpowered to detect a small but still potentially clinically important change in outcome.

Observational research in English general practices also suggested that using depression symptom questionnaires could improve the process of care for patients. Following NICE guidance, from 2006 to 2013 the general practice contract Quality and Outcomes Framework (QOF) paid GPs to use symptom questionnaires to assess depression severity at diagnosis of a new episode. Questionnaire assessments at follow-up were also rewarded in the QOF from 2009 to 2013. Analysis of medical record data of 2294 patients from 38 practices, conducted in the year following the introduction of the QOF incentivisation of questionnaire use, found that higher questionnaire scores at diagnosis were associated with a greater likelihood of treatment with an antidepressant, or referral to psychology, or both.[23] In addition, qualitative interviews with 24 patients suggested that many valued the use of questionnaires to confirm their diagnosis and monitor their progress, although only a minority of the 34 GPs interviewed also valued them for monitoring patients.[24]

A later study of using the PHQ-9 at diagnosis and again at follow-up comprising 604 patients in 13 practices found that decisions to change treatment at follow-up were significantly associated with a lack of improvement in symptom questionnaire scores.[25] Patients who showed a response defined as inadequate in terms of change of PHQ-9 score at the second assessment were nearly five times as likely to experience a subsequent change in treatment as those who showed an adequate response.

However, some GPs disliked the use of questionnaires, saying that they intruded in consultations and undermined their autonomy. Some doubted the validity of the questionnaires, preferring to use their own judgement to assess severity and response to treatment.[24] In response to criticisms, NICE commissioned a review[26] that concluded that the evidence was not strong enough to require the use of questionnaires in QOF depression indicators. Currently, the QOF rewards follow-up reviews of patients with depression 10–56 days after diagnosis, but the use of symptom questionnaires is optional and not necessary to receive payments.

While the relationships found between PHQ-9 scores and treatment and referral were in the direction expected if questionnaire scores informed the process of care, observational research cannot demonstrate cause and effect, and could simply reflect the fact that questionnaire scores were consistent with clinical judgement and were not primarily influential in determining treatment choices.

Therefore, we carried out a feasibility study for a randomised controlled trial (RCT) of PROMs for monitoring depression in UK primary care to test the relationship between PROM scores, subsequent treatment decisions and outcomes for patients.[27] We tested individual patient and cluster randomisation in nine practices comprising 47 adults with new episodes: 22 intervention and 25 control. Three PROMs were administered following diagnosis and again 10–35 days later: the PHQ-9, the brief Distress Thermometer 0–10 analogue scale[28] and the longer Psychological Outcomes Profile (PSYCHLOPS) problem profile.[29] The feeding back of PROM scores to patients was left to the practitioners to manage as they would.

Our feasibility trial found the mean Beck Depression Inventory, 2nd edition (BDI-II), score at 12 weeks was lower among intervention arm patients than control patients by 5.8 points [95% confidence interval (CI) –11.1 to –0.50 points], adjusted for baseline differences and practice.[27] Social functioning scores were not significantly different. At 26 weeks, there were no significant differences in symptoms, social functioning, quality of life or costs, but the mean score for satisfaction with medical care received was lower in the intervention arm by 22.0 points (95% CI –40.7 to –3.29 points).[27] Qualitative interviews suggested this was because patients were disappointed when their GPs did not use the PROM scores to inform their treatment.

In qualitative interviews, some participating GPs reported the PROMs were not useful, and others wanted more guidance on treatment actions in response to the scores. Some described the Distress Thermometer's simple rating of 1–10 as too blunt, and some reported that there was not enough time in consultations for the considerably longer PSYCHLOPS to be undertaken. Most preferred the PHQ-9 as they understood the scores on that more than those on the other two measures. Some considered the PHQ-9 to be more valid, and all were more used to using it due to its previous incentivisation in the QOF.

We concluded from our feasibility trial that PROMs might improve depression outcomes, even if they do not always inform management, in line with the findings of the trial of using the PHQ-9 in the USA,[19] which suggested that patients might feel more involved in their care and more motivated to adhere to treatment and follow-up.[20] That view was also supported to an extent by a qualitative interview study carried out with 27 GPs in Sweden.[30]

In summary, there was no consistent evidence that using symptom questionnaires as PROMs improved the targeting of treatment and outcomes for depression in primary care. However, if they were effective

in improving management or outcomes, then they would likely be cost-effective given their low cost, and the benefits at a population level could be considerable in public health terms given the high costs of depression. Hence the PROMDEP definitive RCT was needed.

Our feasibility study informed its design and conduct including the PHQ-9 as the PROM of choice, the provision of more feedback to patients on the meaning of their scores and the provision of training to GPs in the use of the PHQ-9 to guide treatment choices.

Aim and objectives

The aim was to answer the following research question: What is the clinical effectiveness and cost-effectiveness of assessing primary care patients with depression or low mood soon after diagnosis and again at follow-up 10–35 days later, using the PHQ-9 questionnaire combined with patient and practitioner feedback and guidance on treatment?

The objectives were:

1. to carry out a parallel-group, cluster-randomised controlled trial that will compare (1) getting patients to complete the PHQ-9, for use as a PROM in their consultations with GPs or nurse practitioners (NPs) treating them for depression, with (2) usual practitioner care, uninformed by PHQ-9 scores
2. to motivate and train participating practitioners to reflect on the best use of the PHQ-9, improving their capability to interpret symptom scores, taking into account patients' responses to open-ended global enquiries, their level of functioning, history and social context including life events and difficulties
3. to provide patients in the intervention arm with written feedback on their PHQ-9 scores, including a 'traffic light' indication of the severity of their depression, a 100-manikin representation of the proportion of people in the population with that level of depression, and a brief list of evidence-based treatments relevant to the severity, which they will be asked to discuss with their GP/NP
4. to follow up participants for 26 weeks, with research assessments at 12 and 26 weeks
5. to determine the primary outcome of depressive symptoms on the Beck Depression Inventory, 2nd edition (BDI-II), at the 12-week follow-up
6. to examine secondary outcomes including depressive symptoms on the BDI-II at 26 weeks, and social functioning, quality of life and changes in drug treatment and referrals, at both 12- and 26-week follow-up
7. to measure service use and costs over the 26-week follow-up period and perform cost-effectiveness and cost–utility analyses based on the results of the trial
8. to carry out a qualitative process analysis to explore participants' reflections on the conduct of the trial and the potential for implementing the use of PROMs in practice.

Chapter 2　Methods

Setting

The study was carried out in primary care and recruited general practices in England and Wales from three sites: the University of Southampton, the University of Liverpool and University College London.

Design

The study design was a parallel-group, cluster-randomised trial, with patients clustered by participating practices, and 1 : 1 allocation of practices to intervention and control arms. We chose a cluster-randomised design, as in the feasibility trial we found that randomising patients individually within practices risked contamination between study arms (GPs or NPs taught to use a symptom questionnaire with intervention arm patients could use similar questions in a systematic way with control patients). In addition, patients do not always see the same GP/NP at diagnosis and follow-up, so all practitioners in a practice needed to follow the same protocol to optimise adherence to intervention or control arm procedures, and that was optimised using a cluster design.

Ethics approval and research governance

Independent peer review through the National Institute for Health and Care Research (NIHR) Health Technology Assessment panel ensured scientific quality and rigor. Ethics Committee and Health Research Authority (HRA) approvals were obtained prior to commencement of work with patients and health professionals, and subsequent issues were addressed with the Research Ethics Committee (REC) or HRA offices with applications to approve study amendments as necessary. The study was approved by the NHS REC West of Scotland REC 5, on 21 September 2018 (reference 18/WS/0144).

We ensured that the study aims were relevant to patients and the public through patient and public involvement (PPI) colleagues' input in the design, and their involvement continued throughout to ensure that participation was voluntary, that easily understood patient information was provided and that fully informed consent was obtained, ensuring confidentiality at all times. The information emphasised that participation in the trial was voluntary and that the participant could withdraw from the trial at any time and for any reason. The participant was given the opportunity to ask any questions that may have arisen and provided with the opportunity to discuss the study with family members, friends or an independent healthcare professional outside the research team. They were also given time to consider the information prior to agreeing to participate.

Participants

The target population was patients aged ≥ 18 years diagnosed by a GP or NP with a new episode of depression disorder or depressive symptoms. A new episode meant no diagnosis or treatment within the previous 3 months of presenting with the new episode of depression.

Inclusion criteria
The main inclusion criteria were adult patients seen in the practice within the last 2 weeks and assigned medical records computer codes by GPs or NPs for new presentations with diagnoses or symptoms of depression. There was no upper age limit, and patients with coexisting physical health problems were not excluded.

Exclusion criteria

Patients were excluded if they had been treated for depression in the 3 months prior to presenting to their GP, or if they had comorbid dementia, psychosis or substance misuse (as a main problem), or if they were judged to be at significant at risk of suicide, in which case their GP was informed immediately.

Recruitment of practices

We approached practices to discuss recruiting them to the study with the help of 13 NIHR Clinical Research Networks (CRNs), namely Wessex; North West Coast; North Thames; Kent, Surrey and Sussex; North West London; South London; Greater Manchester; Thames Valley and South Midlands; Yorkshire and Humber; West Midlands; Betsi Cadwaladr University Local Health Board; North-East and North Cumbria; and West of England. The study was advertised to practices signed up for research projects through sending them written research information sheets for practices (RISPs) summarising the study, and through a short introductory video from the chief investigator TK on YouTube available at www.youtube.com/watch?v=rSS29yIMBL4 (accessed November 2023).

Initially research staff visited interested practices for face-to-face site initiation visits to go through the trial processes in detail, but as this was not possible during the COVID-19-related lockdowns the initiation visits were also conducted remotely. Participating practice staff were sent a link to a set of site initiation visit slides on YouTube available at www.youtube.com/watch?v=ic5xGqdIrH0 (accessed November 2023). This was then followed up with e-mail and telephone correspondence to set up each practice's trial processes and documentation. Remote initiation visits enabled recruitment of practices from a much larger geographical area than was possible with face-to-face visits, which were generally limited by how far researchers could travel to them.

Recruitment of patients

In-consultation patient recruitment

Where possible, patients who were seen with a new episode of depressive symptoms or disorder were recruited opportunistically during consultations by participating GPs and NPs in both arms of the study. Patients identified through this method were given the information sheet by hand, together with a reply slip and a Freepost envelope, and asked to post it to the study team if they wished to take part. From May 2020, patients consulting by telephone or video call (due to the COVID-19 pandemic restrictions) were sent the information sheet by text or e-mail before, during or after the consultation at which they presented with a new episode of depression.

Medical records searches

However, because recruitment in consultations could have been subject to selection bias by the GP/NP, patients presenting with a new episode of depressive symptoms or disorder were also identified through searches of practice medical records databases carried out every 1–2 weeks to identify patients whom the GP/NP had not selected for possible participation. In the feasibility trial, both methods were used, and 79% of patients were recruited in consultations opportunistically and 21% through the weekly database searches, but this varied by practice, and some practices actually recruited the majority of patients through the weekly searches.

Our experience gained identifying people treated for depression in a previous general practice study[31] had shown that in around 120 medical records computer codes were used by GP/NPs, including both diagnostic codes (e.g. *major depressive disorder*) and symptom codes (e.g. *low mood*). Practices were given a search strategy that used the full list of both diagnostic and symptom codes for searching their databases weekly. Patients identified through this method received a study information sheet from the practice by post and were asked to contact the study team if they wished to take part or decline,

using the reply slip and Freepost envelope included. If they did not respond, the research team had no knowledge of them, maintaining patient confidentiality.

Telephone screening prior to recruitment

If patients did respond positively to either approach, a member of the research team contacted them to screen them over the telephone for any exclusion criteria and to arrange to see them for the baseline assessment either face to face or remotely, using Microsoft Teams (Microsoft Corporation, Redmond, WA, USA) or over the telephone (instigated during the COVID-19 pandemic).

Consent

For the baseline assessments, the patients were asked to give initially verbal, and subsequently written, consent to take part (by post or online). Online consent was given by completing an electronic copy of the consent form on Microsoft Forms (Microsoft Corporation, Redmond, WA, USA). For the patient and health professional qualitative interviews, participants gave initial verbal consent prior to the interviews. Consent was audio-recorded prior to the interviews and saved as a separate file from the interview to ensure anonymity. Subsequent written consent was again obtained online using Microsoft Forms.

The baseline visit was offered at patients' general practice premises, at their home or remotely, depending on patient preference or on necessity during the COVID-19 pandemic. The researchers attempted to meet with the patient within a week of receiving their reply slip indicating their interest in participating, in order to see them within 2–3 weeks of their initial presentation to the GP/NP. At the initial contact the researcher went over the patient information sheet again, sought initial verbal consent and carried out the baseline research assessment. The participant was then asked to confirm their consent in writing afterwards, by post, by e-mail or online using Microsoft Forms. Patients were advised that all information they provided was confidential and would not be shared with anyone else, the only exception being that information might be shared with their GP if significant risk of harm was suspected but that would be after discussion with them and ideally with their consent.

Intervention

The intervention consisted of getting patients to complete the PHQ-9 for depression symptoms so that this could be used as a PROM during their consultations with GPs or NPs treating them for depression.

The PHQ-9 is a nine-question self-report measure of depression symptoms that takes approximately 3 minutes to complete.[32] It asks about the nine diagnostic symptoms of major depressive disorder in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition (DSM-IV), and scores on each symptom range from 0 (not at all) to 3 (nearly every day). Total scores are categorised into minimal or no depression (0–4), mild (5–9), moderate (10–14) and severe (15–27). The questionnaire was developed and originally validated against diagnostic interviews in the USA and can be downloaded free of charge from https://patient.info/doctor/patient-health-questionnaire-phq-9 (accessed November 2023).

The initial PHQ-9 was administered by the researcher and completed by participating patients as soon as possible after diagnosis, and then administered again, but by the GP or NP, at a follow-up consultation 10–35 days after that. (This follow-up period was chosen as it was the interval specified for financially incentivised follow-up assessments in the GP contract QOF at the beginning of the study period.) Patients were given feedback by the researcher at the first administration of the questionnaire on the meaning of their symptom score and possible treatment options to discuss with the practitioner (see

Figures 1 and *2*). Practitioners were trained in the interpretation of symptom scores in the context of the patient's life situation, and in further assessment to inform their treatment decisions.

In routine clinical practice, as opposed to the trial situation, the GPs/NPs themselves would ask the patients to complete the PROM, either in the first consultation for depression or between consultations, but the trial situation was different, as patients had to consent to take part after being given sufficient time (at least 24 hours as specified by the Ethics Committee) to consider this, so the GP/NP could not give a PHQ-9 questionnaire at the first consultation. We took this difference into account when estimating the cost of the intervention.

Written feedback to patients

We provided patients in the intervention arm with written feedback on their PHQ-9 scores, including a 100-manikin representation of the proportion of people in the population with that level of depression (*Figure 1*), a 'traffic light' indication of the severity of their depression (*Figure 2*), and a brief indication of possible evidence-based treatments relevant to the severity, which they were asked to discuss with their GP/NP. PHQ-9 sum scores range from 0 to 27, with scores of 0–4, 5–9, 10–14 and ≥ 15 representing probable minimal or no depression, mild, moderate and severe depression symptom levels, respectively. These were fed back to patients in four corresponding probability categories: green, yellow, orange and red. An example of feedback to a patient scoring 14 on the PHQ-9 is shown in *Figures 1* and *2*.

This approach proved successful in Löwe *et al.*'s[33] DEPSCREEN-INFO study of providing written feedback after screening patients with cardiological problems for depression. Six months after screening, the patients in the feedback group showed significantly greater improvements in depression severity than, and were twice as likely to seek information about depression as, the control group.[33] Bernd Löwe was an international advisor on our proposal, and the infographic representations were used with his permission.
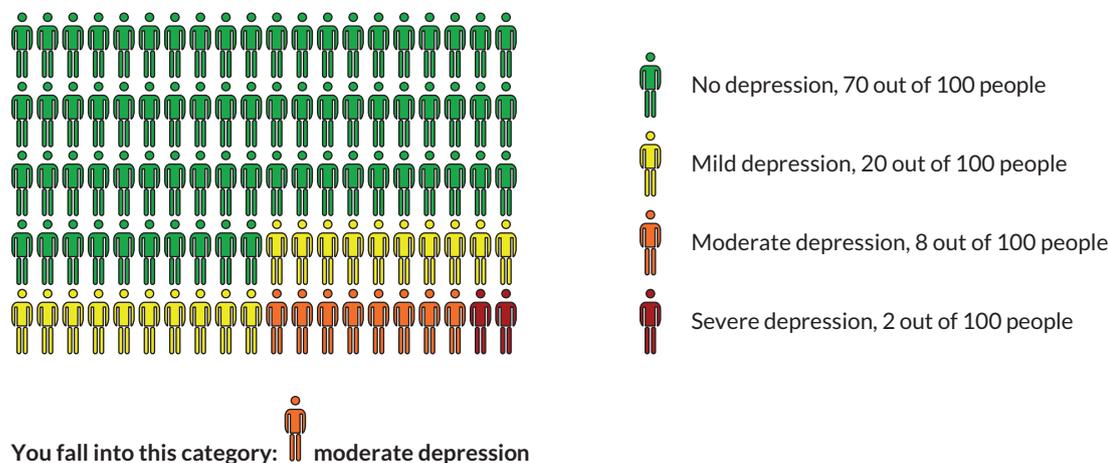
FIGURE 1 Distribution of depressive symptom severity in the general population.

FIGURE 2 Traffic light indication of severity of depression.

General practitioner/nurse practitioner training

Our feasibility study suggested that GPs' discussion of the PHQ-9 scores with patients, and use of them to inform treatment, was suboptimal, affecting both their own perception of the measure and patients' satisfaction with the care they received.[27] To try to change practitioner behaviour in the trial, we decided to implement up to 2 hours of structured training. By triangulating our qualitative feasibility findings with behavioural theory,[34] we determined the need for the training to focus primarily on GPs'/NPs' reflective motivation (e.g. beliefs about the usefulness of PROMs) and psychological capability (e.g. knowledge and understanding to apply PROMs effectively). These constructs are drawn from the 'COM-B' system of behaviour (referring to capability, opportunity, motivation and behaviour).[34] The COM-B system is used widely in behaviour change research and focuses on necessary antecedents for voluntary behaviour to occur.

Participating GPs/NPs were therefore given up to 2 hours of training, including information on the study and best practice in the use of the PHQ-9 in a PowerPoint presentation, followed up with questionnaires for the GPs/NPs to complete on the study processes and a set of four case vignettes asking questions about the management of depression and use of the PHQ-9. The training focused on evidence that patients do value using PROMs and can benefit from being more involved in their own care, even if the scores do not alter treatment, in order to get GPs/NPs to reflect on the value of the use of the measure. We addressed GP/NP concerns around the validity of the PHQ-9 by acknowledging individual differences in patient response set and advising them to combine more global open-ended questions with the questionnaire measure.

Training was originally planned to be face to face but in the event was provided online, through a link to a set of teaching slides with narration hosted on the SharePoint site at the University of Southampton and available to intervention arm practice GPs/NPs through a password provided individually to each practitioner. Participating GPs/NPs were asked to view the slides and then provide answers to two sets of questions, one on the use of the PHQ-9 in the study, and one on a set of depression case vignettes illustrating the use of the questionnaire in making treatment decisions. They were asked to feed back their answers to the two sets of questions to the study team to show that they had completed the training and understood it. *Appendix 1* shows the training feedback questionnaire test scores for participating practitioners.

Use of the PHQ-9 in practice was also modelled by one of the co-principal investigators, CD, with a simulated patient, in videos representing the first and second follow-up consultations for depression with a practitioner in the study. The videos can be found at https://youtu.be/dex-OOH3fUM (accessed November 2023) and https://youtu.be/-mgODzhGgj4 (accessed November 2023).

Procedures in intervention and control arms

After baseline assessment, the patient was asked to arrange a follow-up appointment with their GP/NP, either remotely or in person, as soon as possible and to take with them their completed PHQ-9 questionnaire plus written feedback so that they could discuss the score and the treatment suggestions with the GP/NP. To ensure that the PHQ-9 result and patient feedback reached the practice, we also e-mailed it using a secure NHS e-mail account.

Participating GPs/NPs were asked to take the PHQ-9 scores and patient advice into account when deciding about treatment(s) at their next consultation with the patient, following the treatment guidance given during training, taking the patient's response to a global open-ended inquiry into account, together with their level of functioning, social context and history.

The GP or NP was also asked to provide the patient with a fresh PHQ-9 at that second consultation that the patient could take away and complete immediately prior to a third, follow-up consultation 10–35 days later. At that third consultation, the GP/NP was asked to go through the follow-up PHQ-9

with the patient and take the change in score between consultations into account when deciding about possible changes to treatment(s).

In control arm practices, patients did not complete the PHQ-9. They were seen by the research team either in person or remotely as soon as possible after their first consultation for depressive symptoms, and they were asked to complete baseline research outcome measures, but they were not given feedback on the results. They were asked to arrange a follow-up appointment with the GP/NP, either remotely or in person, to match what happened in the intervention arm, but the GP/NP treating them did not receive training and was asked to provide their usual care.

Timing of starting treatment

Practitioners in both the intervention and control arms were advised that best practice in treating depression was not to start treatment at the consultation at which symptoms of a new episode were presented by the patient, unless they thought in their clinical judgement that it was absolutely indicated. This is because a significant proportion of patients will improve without treatment within 2–3 weeks, having had their problems acknowledged and having received general advice about the nature and course of depression. We were interested in this study with the use of the PHQ-9 in deciding on initial treatment, as well as follow-up monitoring, so we preferred treatment was not started before the baseline assessment in both arms, and before the first PHQ-9 questionnaire was administered by the researcher in the intervention arm. In the feasibility study this had been carried out on average 10 days (range 1–38 days) from receiving the patient's reply slip, and we therefore thought that in most cases it should be possible to complete baseline assessment within 2 weeks of the patient's first presentation.

It was possible, however, that patients recruited either opportunistically or via the weekly searches were started on treatment at the consultation when they first presented with a new episode if treatment could not be postponed in the judgement of the treating practitioner. We recorded whether or not treatment had already been started at the baseline assessment.

Primary outcome

The primary outcome was the patient's score on the BDI-II.[35] This is a 21-item self-report instrument that has been established as a valid and reliable instrument for depression screening in the general population[36] and is widely used in depression trials. It takes approximately 5 minutes to complete. Each item is scored from 0 to 3 and a total score of 0–13 is considered minimal depression, 14–19 is mild, 20–28 is moderate and 29–63 is severe.

Secondary outcomes

Social functioning

The Work and Social Adjustment Scale (WSAS) assesses problems in functioning with work, home management, social leisure activities, private leisure activities, and family and relationships, all on scales of 0–8.[37] It has been shown to be a sensitive, reliable and valid measure of impaired functioning and is used routinely in IAPT psychological therapy settings as well as in research studies in a variety of settings. A higher score represents better functioning.

Health-related quality of life

The EuroQol-5 Dimensions, five-level (EQ-5D-5L) measure of health-related quality of life[38] is the measure NICE favours in determining cost-effectiveness when developing its clinical guidelines. The EQ-5D-5L has five dimensions: mobility, self-care, usual activities, pain/discomfort and anxiety/depression. Each dimension is scored on five levels: no problems, slight problems, moderate problems, severe problems and extreme problems. Health states are converted into a single summary index

ranging from 0 (worst) to 1 (best) by applying weights to each level in each dimension derived from the valuation of EQ-5D-5L health states in adult general population samples.[39] The EQ-5D-5L measure was also used to generate utility scores to determine changes in quality-adjusted life-years (QALYs) for the health economics evaluation.

Costs

Costs were calculated from responses to a bespoke questionnaire based on the Client Service Receipt Inventory (CSRI)[40] but modified specifically for the study. A review of participating patients' digital medical records was also carried out by practice staff after the 26-week follow-up to augment questionnaire measurement of health and social service resource use using the modified CSRI.

Patient satisfaction with care

This was assessed using the 29-item Medical Informant Satisfaction Scale (MISS-29), which was developed in the USA to assess patient satisfaction with individual doctor–patient consultations and has been shown to be valid and reliable in UK primary care.[41] We adapted it to rate patient satisfaction at the 26-week follow-up, asking patients to look back over their consultations with GPs/NPs over the whole 26-week period. Again, a higher score indicates greater satisfaction.

Assessments at baseline and follow-up

Patients were recruited over a 39-month period and followed up for 26 weeks each, with assessments at baseline and at 12 and 26 weeks. Recruitment was planned to take only 26 months but, in the event, had to be prolonged by 13 months in order to recruit sufficient patients to answer the research question. Follow-up assessments took place at patients' general practices, at their homes, or remotely if they preferred, and remotely in every case when it became necessary due to the COVID-19 pandemic.

Baseline

At the baseline visit, the following measures were administered:

- the BDI-II[36] for current level of depression
- a bespoke questionnaire on sociodemographic details (gender, age, ethnicity, socioeconomic position, housing, education, marital status and dependants)
- a bespoke questionnaire on the duration of the current episode of mood disturbance, and any history of depression
- a bespoke questionnaire on antidepressant use, including name, dose and duration of prescriptions
- the Generalised Anxiety Disorder scale, 7-item (GAD-7), questionnaire for anxiety symptoms[42]
- the WSAS[37] for social functioning
- the EQ-5D-5L questionnaire for quality of life[38]
- a bespoke questionnaire on consultations, drug treatments, community care contacts and hospital care contacts over the 6 months prior to baseline assessment (to compare service use in intervention and control arms at baseline)
- a bespoke questionnaire asking about absence from work due to sickness in the previous 6 months, which was amended to also ask about loss of employment due to COVID-19.

The GAD-7 is a measure of generalised anxiety disorder symptoms.[42] The total score is calculated by assigning scores of 0, 1, 2 or 3 to the response categories of 'not at all', 'several days', 'more than half the days' and 'nearly every day', respectively, and adding the scores for the seven questions. Scores of 5, 10 and 15 are taken as the cut-off points for mild, moderate and severe anxiety, respectively. Using the threshold score of 10, the GAD-7 has a sensitivity of 89% and a specificity of 82% for generalised anxiety disorder.[42]

Subsequent assessments

At the 12-week follow-up, the following measures were administered:

- the BDI-II to measure changes in depressive symptoms (primary outcome)
- the WSAS to measure changes in social functioning
- the EQ-5D-5L to measure changes in quality of life
- the bespoke questionnaire on antidepressant use including name, dose and duration of prescriptions
- the bespoke questionnaire on sickness absence and absence due to COVID-19.

At the 26-week follow-up, the following measures were administered:

- the BDI-II to measure changes in depressive symptoms (secondary outcome)
- the WSAS to measure changes in social functioning
- the EQ-5D-5L to measure changes in quality of life
- the bespoke questionnaire on antidepressant use including name, dose and duration of prescriptions
- a modified version of the MISS-29 to measure patient satisfaction over the whole 26-week trial period
- the bespoke questionnaire on consultations, drug treatments, community care contacts and hospital care contacts over the 26-week trial period
- the bespoke questionnaire on antidepressant use, including name, dose and duration of prescriptions.

In addition to the patient questionnaire on consultations, drug treatments, community care contacts and hospital care contacts, data were extracted from participating patients' general practice computerised medical records by practice staff after the 26-week follow-up, using the same pro forma as the patient questionnaire. The medical records were the prime source of data on resource use to calculate NHS and Personal Social Services (PSS) costs for the health economics analyses.

Data were collected through face-to-face meetings or remotely, online or by post, but brief telephone follow-up was offered to obtain at least the primary outcome (BDI-II score) and quality of life (EQ-5D-5L) for the study's primary clinical and economic outcomes, if the researcher was unable to arrange to assess the patient fully online, by post or face to face.

Participants received a £10 high street shopping voucher at both the 12- and 26-week follow-ups to thank them for their participation in the study.

Assessing suicidal ideation

Patients who disclosed information during an interview (face to face, over the telephone or remotely) to the fieldworker indicating that they had attempted suicide, or that they had been thinking of ways to commit suicide, were considered to have suicidal ideation. In addition, anyone scoring more than zero on the self-harm question of either the BDI-II or the PHQ-9 were assumed to have possible suicidal ideation.

In each case, the P4 four-question suicide risk assessment screener[43] was implemented by the researcher, asking:

1. Have you ever attempted to harm yourself in the past?
2. Have you thought about how you might actually hurt yourself? (If yes, how?)
3. There's a big difference between having a thought and acting on a thought. How likely do you think it is that you will act on these thoughts about hurting yourself or ending your life some time over the next month? (Not at all likely/somewhat likely/very likely)
4. Is there anything that would prevent or keep you from harming yourself?

Their GP/NP practice was then informed immediately by telephone of the occurrence and the patient's responses to the four questions. The patient's GP was asked to review the patient as soon as possible, with a view to a possible urgent referral to mental health services in the case of a significant risk of suicide or self-harm [see *Appendix 2* for the standard operating procedure (SOP) we used for suicidal ideation].

Sample size calculation

We needed a sample large enough to detect a difference between the arms at follow-up in the minimal clinically important difference (MCID) score on the primary outcome, the BDI-II.

Button *et al.*[44] used data collected from 1039 patients from three RCTs on the management of depression and compared improvement on a 'global rating of change' question with changes in BDI-II scores to determine the MCID. They used general linear modelling to explore baseline dependency, assessing whether MCID is best measured in absolute terms (i.e. a specific change in score) or as percentage reduction in scores from baseline (i.e. as a ratio). Their modelling indicated that the MCID was best measured as a percentage reduction in score, and a 17.5% reduction from baseline was identified from receiver operator characteristics analyses as the optimal threshold above which individuals reported feeling 'better'.[44]

In the PROMDEP feasibility trial, we found the mean BDI-II score at baseline was 24.0 and the SD was 10.0.[27] At the 12-week follow-up, based on the results of the feasibility study, we anticipated a mean of 14.0 in the intervention arm and 17.0 in the control arm. This gave an anticipated mean difference of 3.0 on the BDI-II, which was an effect size of 0.3 SDs, and in keeping with the findings of Knaup *et al.*'s[11] systematic review for the expected effects of combined practitioner and patient feedback of PROMs. The difference of 3.0 points was 17.6% of the control arm's score of 17.0 at 12 weeks, and therefore just above the MCID established for the BDI-II.[44] The anticipated potential benefit was therefore relatively small, but likely to be clinically significant.

We aimed to recruit a mean of six patients per practice. We assumed an intracluster correlation coefficient of 0.03 (from the feasibility study). At the level of 5% significance, to have 90% power to detect a difference between 14.0 and 17.0 on the BDI-II we calculated that we needed 235 patients analysed per arm. Given a cluster size of six, the cluster design effect would have been 1.15, meaning we needed 270 per arm. We assumed a 20% loss to follow-up at 12 weeks, so the total sample size needed was 270 × 2/0.8 and our original target sample size was a total of 676 patients recruited, from 113 practices, from the three university recruitment centres (Southampton, UCL and Liverpool).

However, after 24 months of recruitment, when applying for additional funding for an extension of a further 12 months, we found a correlation coefficient *p* between baseline and follow-up values for the primary outcome (the BDI-II score at 12 weeks) of 0.6 (95% CI 0.5 to 0.7). Assuming conservatively that this correlation remained ≥ 0.5 until the end of follow-up, the necessary target sample size to give 90% power was therefore reduced by a deflation factor of $1-p^2$, that is, $1-0.5^2$, which meant that we needed only 222 patients analysed per arm. With the cluster design effect of 1.15 and assuming 20% loss to follow-up, this gave a revised target of 222 × 1.15 × 2/0.8 × (1−0.25) = 554 in total (revised 10 June 2021).

Randomisation

The randomisation of practices to the intervention or control arm was carried out by an independent statistician from the Southampton clinical trials unit, so that the study statistician, BS, could remain blind to allocation. Randomisation was by computerised sequence generation, using minimisation to balance

three factors between the two arms as much as possible: practice size (large or small), location (urban/suburban or rural) and university recruiting centre (Southampton or Liverpool or UCL).

Blinding

The blinding of patients, practitioners and researchers to the allocated arm of the trial was impossible, given the nature of the intervention and the clustered design. Therefore, self-report outcome measures were used to prevent observer rating bias by researchers. The statistical analyses were, however, carried out blind to allocation.

Statistical analyses

Full details of the analyses to be undertaken were set out in a statistical analysis plan approved by both the Trial Steering Committee and the Independent Data Monitoring Committee (see *Report Supplementary Material 1*). This was published on the ISRCTN register at https://doi.org/10.1186/ISRCTN17299295.

Primary and secondary outcomes

The primary outcome, the difference at 12 weeks between intervention and control arm patients in depressive symptoms as measured using the BDI-II, was analysed using a linear mixed model, adjusting for baseline BDI-II depression symptom score; baseline GAD-7 anxiety score; demographic factors (socioeconomic position, housing type, education, marital status, dependants, gender and age), duration of depression; history of depression; and clustering including a random effect for practice. The model used all the observed data and assumed that missing BDI-II scores were missing completely at random.

Analysis of secondary outcomes, including BDI-II scores at 26 weeks, social functioning on the WSAS, patient satisfaction on the MISS and EQ-5D-5L quality-of-life score, was also conducted using mixed linear regression for continuous outcomes and logistic regression for dichotomous outcomes, again adjusting for baseline depression, baseline anxiety, sociodemographic factors and practice as a random effect. Assumptions for linear regression models (linearity, normality, homoscedasticity) were checked using scatterplots of standardised residuals against fitted values, and Q–Q plots.

No a priori interim analyses were planned, and no prespecified stopping rules were established. There were no planned subgroup analyses at the beginning of the trial (but see below regarding a subsequent decision to conduct a subgroup analysis related to the COVID-19 pandemic).

Intention-to-treat analysis

The population for analysis included all randomised practices and all patients recruited within them regardless of treatment compliance. All summaries and analysis were on the intention-to-treat population. There were no pre-planned per-protocol analyses.

Withdrawal from trial

All data up until the point of patient withdrawal from the trial were used in analyses unless the patient withdrew consent and did not want their data already collected prior to withdrawal to be used for the trial. If a practice withdrew from the trial, no further patients were recruited from that practice, but all data on patients collected up until that point were used and any patients recruited continued to be followed up in accordance with the trial schedule.

Missing data

The primary analysis was of complete cases. If more than two items of the BDI-II or the GAD-7 had missing values, the total score was recorded as missing. If one or two items were missing, the scores

were imputed with the mean of the non-missing scores before summing. We examined the structure and pattern of missing data and undertook a sensitivity analysis as specified in the statistical analysis plan based on data imputed using a chained equations multiple imputation model. The chained equations imputation model included the outcome measure, baseline value of the outcome, randomisation group, clustering by practice and all covariates included in the analysis model.

Coronavirus disease 2019 pandemic

Due to the COVID-19 pandemic and subsequent lockdown period, we were aware that changes to the key outcomes unrelated to randomisation group may have occurred during the study period. We decided therefore to look at the scores in each arm in the pre-, peri- and post-COVID-19 periods in the whole study population in a post hoc subgroup analysis. We used descriptive statistics to explore any trends and aimed to control for any time-varying effect on outcomes in a sensitivity analysis.

Process of care

We examined differences between the intervention and control arms in patients' self-reported use of antidepressants at 12- and 26-week follow-up points, and medication and contacts with NHS and social services recorded in the general practice medical records over the whole 26 weeks, in particular mental health services including psychologists, psychiatrists, mental health nurses, counsellors, other therapists and social workers (again adjusting for baseline depression, baseline anxiety, sociodemographic factors and practice as a random effect).

Health economics analysis

A health economic evaluation was undertaken from an NHS and PSS perspective. The outcome was expressed as incremental cost per point improvement in the BDI-II clinical outcome (cost-effectiveness analysis) and incremental cost per QALY gained (cost–utility analysis). All items were costed using appropriate data (e.g. *Unit Costs of Health and Social Care*[45]). The primary analysis was at 26 weeks.

A generalised linear mixed model was used to estimate the mean differences in costs and QALYs (using the EQ-5D-5L to calculate utilities), again adjusting for baseline depression, baseline anxiety, sociodemographic factors and practice as a random effect. Where appropriate we estimated incremental cost-effectiveness ratios (ICERs) and used non-parametric bootstrapping to produce cost-effectiveness acceptability curves (CEACs).

Modelling of the likely benefit, if any, of using PROMs in practice included making assumptions about the extra time that would have to be taken for GPs/NPs (rather than the researcher) to administer the initial PROM in the non-trial situation. Major assumptions in the costing of the intervention for the QALYs analysis were tested through sensitivity analyses.

Qualitative process analysis

Interviews were planned with up to 15–20 practitioners and 15–20 patients in each arm to explore their reflections on the conduct of the trial, the use of the PHQ-9 as a PROM, and the potential for implementing its use in practice beyond the trial situation. The main qualitative method adopted was reflexive thematic analysis. Interview schedules included questions related to the normalisation process theory (NPT) framework on the possible implementation of the intervention in practice.[46] NPT focuses on understanding the mechanisms that promote, and the factors that inhibit, sense-making,

participation, action and monitoring by participants in implementation processes.[47] *Appendix 3* shows the main elements of the NPT framework.

Participants in the patient and health professional qualitative interviews gave verbal consent prior to the interview. Consent was audio-recorded prior to the interview and saved as a separate file from the interview to ensure anonymity. Subsequent written consent was obtained either in person, by post or by completing an electronic copy of the consent form on Microsoft Forms. Each patient received a £10 high street shopping voucher for taking part in the interviews.

Interviews with practitioners and patients were carried out as soon as possible after patient assessments at follow-up consultations to explore patient and practitioner recall of interactions within consultations and to identify variations in the use of PROMs and in usual practitioner care. All patients were recruited after they had completed the primary outcome measure at 12 weeks so that the discussion with the researcher at interview would not affect the BDI-II self-rating.

We deliberately sought to recruit a range of practitioners (including different genders and ethnicities, and from rural and urban settings) and patients [including different ages, genders, ethnicities and socioeconomic positions (employment status)] to capture diverse experiences and views.

Patient and public involvement

We recruited BP, convener of Southampton Depression Alliance, and another service user, MB, to join the study group through an advertisement put out by the NIHR South Central Research Design Service. BP and MB had advised us on the feasibility study previously. They read and commented on study design and on the participant information sheets and consent forms to ensure that these were easy to understand and read.

Summary of changes to the protocol

The first version of the protocol, version 1.0, was produced on 18 June 2018.

The first changes on 12 September 2018, resulting in version 1.1, resulted from recommendations from the Sponsor, the University of Southampton, to add to the protocol the IRAS number, the name of the funders and a table to log the protocol versions.

Version 1.2, dated 24 October 2028, provided for online training for GPs randomised to the intervention arm, rather than on-site face-to-face training. It also enabled follow-up questionnaires to be completed by patients by post, over the telephone or online, instead of face to face.

Version 1.3, dated 7 November 2018, included provision for patient participants to receive a £10 high street shopping voucher at the 12- and 26-week follow-up appointments, and another voucher for taking part in a qualitative interview.

Version 1.4, dated 4 January 2019, changed the factors by which randomisation was minimised to (1) small versus large practice size, (2) urban/suburban versus rural practice location and (3) recruiting centre (Southampton, Liverpool or University College London), instead of by (1) practice size, (2) location and (3) practice training versus non-training status.

Version 1.5, dated 4 May 2020, enabled participant consent to be completed verbally then followed up by post, and enabled baseline and follow-up assessments to take place remotely using Microsoft Teams or over the telephone, as a result of restrictions on meeting due to the COVID-19 pandemic.

Version 1.6, dated 19 January 2021, enabled participants to provide informed consent online, using Microsoft Forms, and changed the recruitment end date from 31 January 2021 to 31 March 2021.

Version 1.7, dated 1 March 2021, allowed for qualitative interviews with patients and practitioners to be conducted remotely over the telephone or on Microsoft Teams, due to the COVID-19 restrictions, and changed the recruitment end date from 31 March 2021 to 30 June 2021.

Version 1.8, dated 10 June 2021, changed the target sample size from 676 to 554 patients and the recruitment end date from 30 June 2021 to 31 January 2022.

Version 1.9, dated 4 November 2021, changed the recruitment end date from 31 January 2022 to 31 March 2022.

Chapter 3 Results

Recruitment of practices

We originally planned to recruit and induct 113 practices (33 or 34 from each of the three university sites, Southampton, Liverpool and UCL) between months 1 (November 2018) and 6 (April 2019). However, there was some initial lag in practice recruitment, as, first, an unavoidable delay in recruiting research staff meant that we did not have a full complement until the end of January 2019. Then we found that it took much longer than we had expected to engage with practices after they expressed an interest in participating and to induct and train the GPs, NPs and administrative staff. This was because practices were very busy, particularly during the COVID-19 pandemic, research was delegated to one or two people in the practice, and it sometimes took weeks to be able to meet with those people because of their other commitments.

We therefore modified the milestone of achieving the recruitment of 113 practices to the study from month 6 to month 10 (August 2019). In the event we exceeded our target, recruiting 121 practices by the end of August 2019. However, we decided to continue to recruit practices beyond that point, as the number of patients being recruited by each practice was not as large as we had hoped (see below), and some practices withdrew from the study having initially consented to participate. We eventually recruited a total of 189 practices by the end of December 2021.

However, 48 subsequently withdrew, 24 in each arm, usually due to other commitments, including increasing COVID-19-related work. Therefore, the final number of randomised and active practices was 141, 72 in the intervention arm and 69 in the control arm, which was still 28 above our original target of 113.

The graph of practice recruitment is shown in *Figure 3*.

Training of intervention arm practitioners

A total of 176 practitioners in the 72 intervention arm practices undertook the online training and completed the two sets of questions: one on the use of the PHQ-9 in the study, and the other on the
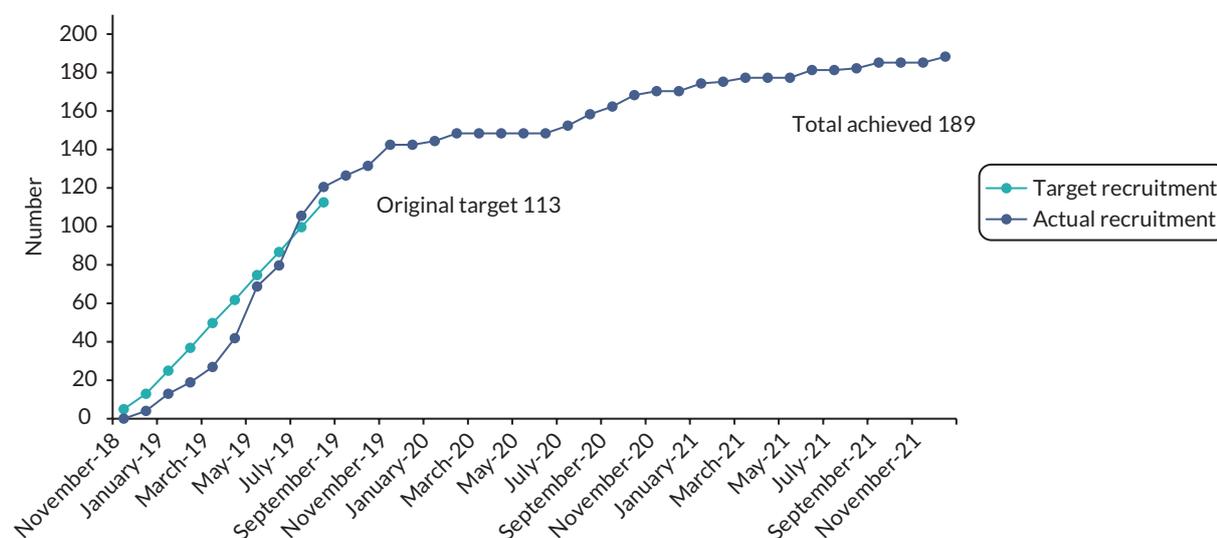
FIGURE 3 Practice recruitment.

four case vignettes describing the management of depression in relation to scores on the PHQ-9 and other factors. The mean score on the first quiz on use of the PHQ-9 was 8.7 out of 10 (range 5–10, SD 1.3). Twenty-one practitioners completed the quiz twice, and one did it three times, usually but not always improving their score on the subsequent attempts (see *Appendix 1*). The mean score on the case vignette was 5.9 out of 8 (range 2–8, SD 1.4). Twenty practitioners completed the quiz twice, and two did it three times, again usually but not always improving their score on the subsequent attempts (see *Appendix 1*).

Recruitment of patients

Our original target sample size was a total of 676 patients recruited from 113 practices across the three recruitment centres (Southampton, UCL and Liverpool). To meet this target, we planned to recruit 26 patients per month over 26 months between December 2018 and January 2021 inclusive. However, for the reasons listed above, which applied to patient recruitment as much as to practice recruitment, there was a 3-month delay in starting, until March 2019, and the rate per month did not reach the target of 26 patients in most months.

Recruitment was then suspended altogether during the first COVID-19 lockdown, for 4 months between March and June 2020 inclusive, because potential participants and researchers needed to self-isolate. After being given HRA and Sponsor permissions to restart patient recruitment in July 2020, the rate per month was even slower during the first year of the pandemic, apparently because practices were preoccupied with an additional related clinical workload that interfered with their ability to carry out research.

Extension of recruitment and revised target
We recruited 345 patients by January 2021, the original end date for recruitment, against the target of 676. We therefore applied for an extension to the study to allow another 12 months' recruitment, at the same time requesting a reduced sample size target. This was because during the course of the study we had found a correlation between baseline and follow-up values for the primary outcome (the BDI-II score at 12 weeks) of p = 0.6 (95% CI 0.5 to 0.7). Assuming conservatively that this correlation p remained ≥ 0.5 until the end of follow-up, the necessary target sample size to give 90% power was therefore reduced by a deflation factor of $1 - p^2$, that is, $1 - 0.5^2$, which meant we needed only 222 patients analysed per arm. With the original cluster design effect of 1.15, and still assuming 20% loss to follow-up, this gave a revised target of $222 \times 1.15 \times 2/0.8 \times (1 - 0.25)$ = 554 in total.

This revised target was agreed with our Independent Data Monitoring Committee, Trial Steering Committee and Sponsor, and the NIHR as funder, on 10 June 2021.

Recruitment met the revised monthly target of a mean of 16.5 patients from February until September 2021, when it slowed down again, apparently due to the need for practices to mount large COVID-19 vaccination initiatives, again reducing the time they could spend on research. We therefore continued recruitment beyond January 2021, finally reaching a total of 529 patients recruited over 39 months by the end of March 2021.

It was not possible to continue recruiting patients beyond March 2021 as the 6-month follow-up and analysis of results had to be completed before the end of the study funding at the end of October 2022. The graph of patient recruitment is shown in *Figure 4*.

Cluster size
Our original target was to recruit six patients per practice. Due to the difficulties practices encountered in finding time for research, we averaged 529/141 = 3.75 patients per practice. This smaller than anticipated cluster size was not included in the revised sample size calculation but would have reduced
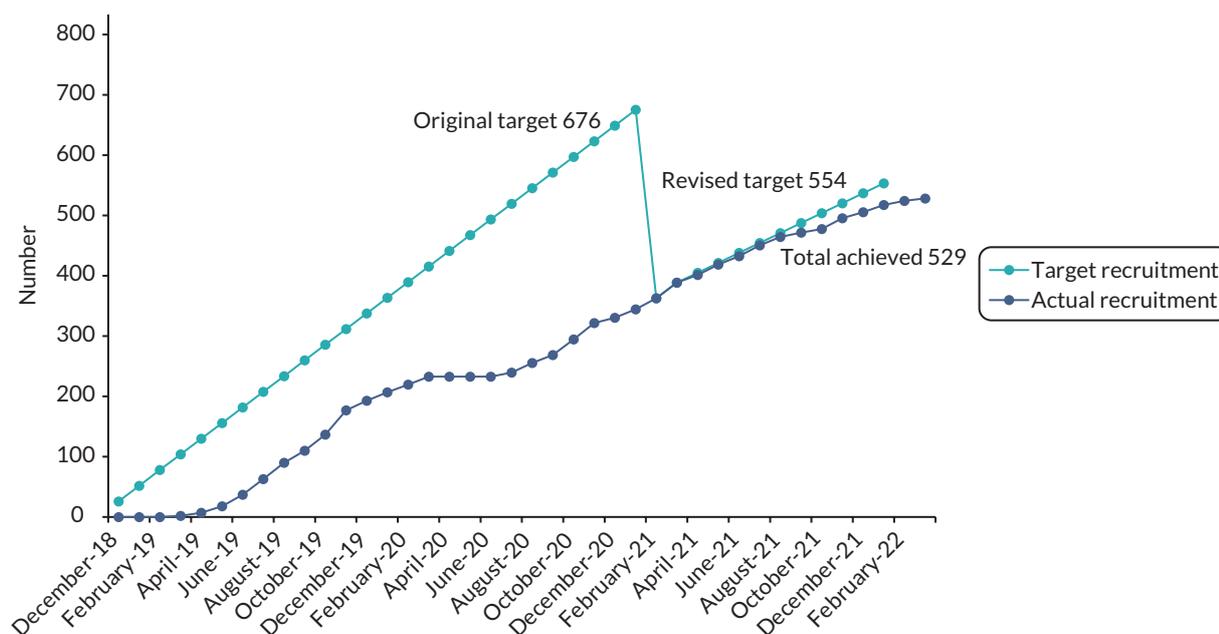
FIGURE 4 Recruitment of patients.

the inflation factor due to cluster randomisation and increased the power of the sample to address the research questions. The intracluster correlation coefficient was 0.06, which may be useful in informing sample size calculations for future trials in depression management in UK primary care.

Flow of participants through the trial

The Consolidated Standards of Reporting Trials (CONSORT) diagram (*Figure 5*) shows the flow of participants through the trial. In the 141 actively recruiting practices, a total of 11,468 patients were approached to take part: 5429 in the intervention arm and 6039 in the control arm. Of these, 539 patients (4.7%) were approached in consultations [364 (6.7%) in the intervention arm and 175 (2.9%) in the control arm] and 10,929 were approached through practice mail-outs to patients identified through regular searches of practice records for patients who had presented with new episodes of depression but had not been approached opportunistically in consultations (5065 intervention arm and 5864 control arm patients).

We actively encouraged recruitment for longer in the control arm practices because we had noticed a differential rate of recruitment between the arms early in the trial recruitment period, which was partly due to fewer patients being approached in consultations in the control practices, and we considered may have reflected lower motivation to recruit among the control arm practitioners.

Of the 11,468 patients approached, 1058 (9.2%) returned reply slips about the study: 574 (10.6% of those approached) in the intervention arm and 484 (8.0% of those approached) in the control arm. After the exclusion of patients who declined to participate, or proved ineligible at screening, or were not contactable, a total of 529 patients consented and were assessed at baseline: 302 (5.5% of those approached) in the intervention arm and 227 (3.8% of those approached) in the control arm. We considered this may have reflected lower motivation to take part among control arm patients. Therefore, the final ratio of intervention to control arm patients recruited was 1.3 to 1. Early in the study we had a ratio of 2 to 1, and that was why efforts to recruit for longer through both consultations and mail-outs were particularly encouraged in the control arm. This resulted in the ratio coming down significantly, but not to 1 to 1 as intended.
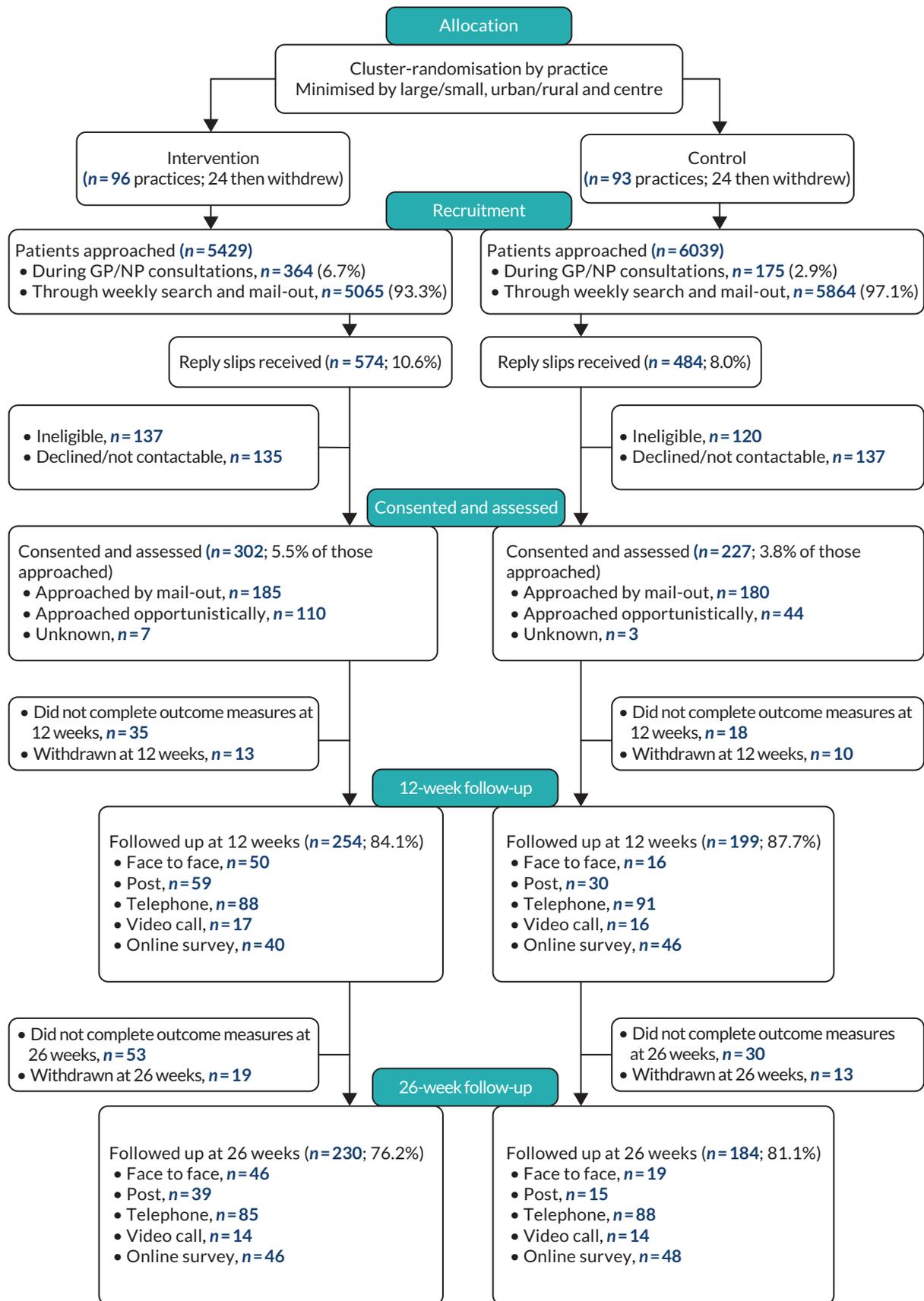
FIGURE 5 CONSORT diagram.

Follow-ups of patients at the 12- and 26-week points were carried out between May 2019 and September 2022 inclusive. Of the 529 patients recruited, 453 (85.6%) were followed up at 12 weeks. Full primary outcome data were available for 252 intervention arm (83.4%) and 195 control arm (85.9%) patients. At the 26-week point, 414 patients (78.3%) were followed up: 230 in the intervention arm (76.2%) and 184 in the control arm (81.1%).

Postal, face-to-face and remote follow-up

During the first COVID-19-related lockdown, NHS REC, HRA and Sponsor approvals were obtained quickly to allow the continued follow-up of already-recruited patients to take place remotely, in addition to postal or face-to-face follow-up, through telephone or video calls or online using the University of Southampton's *i-survey* website. Patient follow-up therefore continued during the study period regardless of the restrictions on contact due to the pandemic, enabling us to achieve the high follow-up rate of 85.6% at 12 weeks. Of the 453 follow-up assessments at 12 weeks, 66 (14.6%) were face to face, 89 (19.6%) were by post, 189 (41.7%) were by telephone, 33 (7.2%) were by video call and 86 (19.0%) were online. Of the 414 follow-up assessments at 26 weeks, 65 (15.7%) were face to face, 54 (13.0%) were by post, 173 (41.7%) were by telephone, 28 (6.8%) were by video call and 94 (22.7%) were online.

Baseline characteristics and comparability

The baseline characteristics of the participating practices were well balanced by arm. *Table 1* indicates that the minimisation by recruiting centre, size of practice and location was successful.

Table 2 shows that at baseline patients in the intervention arm had higher BDI-II depression scores, higher GAD-7 anxiety scores and lower EQ-5D-5L quality-of-life scores. Control arm patients were slightly more likely to have had two or more previous episodes of depression.

The sociodemographic characteristics were relatively well balanced by arm, except that control arm patients were more likely to have no dependants in the household.

TABLE 1 Cluster-level (participating general practice) characteristics at baseline

Characteristic	Intervention (*n*)	Control (*n*)	Total (*N*)
Centre			
Southampton	25	27	52
Liverpool	31	28	59
London	40	38	78
List size			
Small	34	32	66
Large	62	61	123
Location			
Urban/suburban	77	77	154
Rural	19	16	35

TABLE 2 Participating patient characteristics at baseline

Characteristic	Intervention (*N* = 302)	Control (*N* = 227)	Total (*N* = 529)
Mean baseline depression score on the BDI-II (SD)	24.1 (8.89)	22.4 (9.52)	23.4 (9.2)
Mean baseline anxiety score on the GAD-7 (SD)	12.8 (5.31)	11.8 (5.58)	12.4 (5.45)
Mean baseline quality-of-life score on the EQ-5D-5L (SD)	0.659 (0.232)	0.667 (0.226)	0.663 (0.230)
Duration of depression (years)			
Mean (SD)	3.4 (5.13)	2.6 (5.56)	3.1 (5.33)
Previous depression, *n* (%)			
None	87 (28.9)	46 (20.3)	133 (25.2)
Once before	79 (26.3)	62 (27.2)	141 (26.7)
Twice or more before	135 (44.9%	119 (52.4)	254 (48.1)
Female, *n* (%)	192 (63.6)	136 (59.9)	328 (62.0)
Mean age in years at baseline (SD)	45.2 (15.94)	45.0 (17.17)	45.1 (16.46)
Ethnicity, *n* (%)			
White	255 (84.7)	193 (85.0)	448 (84.9)
Black Caribbean	1 (0.3)	3 (1.3)	4 (0.8)
Black African	3 (1.0)	4 (1.8)	7 (1.3)
Black other	2 (0.7)	0 (0.0)	2 (0.4)
Indian	13 (4.3)	4 (1.8)	17 (3.2)
Pakistani	6 (2.0)	4 (1.8)	10 (1.9)
Bangladeshi	0 (0.0)	1 (0.4)	1 (0.2)
Chinese	4 (1.3)	3 (1.3)	7 (1.3)
Other Asian group	5 (1.7)	3 (1.3)	8 (1.5)
Other ethnic group	12 (4.0)	12 (5.3)	24 (4.6)
Socioeconomic position, *n* (%)			
Full-time work	140 (46.4)	113 (49.8)	253 (47.8)
Part-time work	55 (18.2)	28 (12.3)	83 (15.7)
Permanently sick/disabled	5 (1.7)	6 (2.6)	11 (2.1)
Unemployed	36 (11.9)	18 (7.9)	54 (10.2)
Retired	33 (10.9)	31 (13.7)	64 (12.1)
Student	8 (2.7)	12 (5.3)	20 (3.8)
Homemaker	5 (1.7)	4 (1.8)	9 (1.7)
Voluntary work	6 (2.0)	4 (1.8)	10 (1.9)
Other	14 (4.6)	11 (4.9)	25 (4.7)
Accommodation, *n* (%)			
Owner-occupied	142 (47.0)	106 (46.7)	248 (46.9)
Council/housing association	39 (12.9)	20 (8.8)	59 (11.2)
Private rental	71 (23.5)	57 (25.1)	128 (24.2%)

TABLE 2 Participating patient characteristics at baseline (*continued*)

Characteristic	Intervention (*N* = 302)	Control (*N* = 227)	Total (*N* = 529)
Job related	2 (0.7)	1 (0.4)	3 (0.6%)
Lives with parents	40 (13.3)	34 (15.0)	74 (14.0%)
Other	8 (2.7)	9 (4.0)	17 (3.2%)
Highest educational qualification, *n* (%)			
None	26 (8.7)	20 (8.9)	46 (8.8)
CSE/NVQ Level 1	22 (7.4)	3 (1.3)	25 (4.8)
GCSE/O Level	49 (16.4)	33 (14.7)	82 (15.7)
A Level/BTEC	54 (18.1)	41 (18.2)	95 (18.1)
HNC/HND/city and guilds	24 (8.0)	16 (7.1)	40 (7.6)
Degree/higher degree	111 (37.1)	90 (40.0)	201 (38.4)
Vocational qualification	8 (2.7)	14 (6.2)	22 (4.2)
Other	5 (1.7)	8 (3.6)	13 (2.5)
Marital status, *n* (%)			
Married	119 (39.4)	83 (36.6)	202 (38.2)
Cohabiting	26 (8.6)	26 (11.5)	52 (9.8)
Widowed	10 (3.3)	10 (4.4)	20 (3.9)
Separated	11 (3.6)	6 (2.6)	17 (3.2)
Divorced	25 (8.3)	13 (5.7)	38 (7.2)
Single	111 (36.8)	89 (39.2)	200 (37.8)
Number of dependants in the household, *n* (%)			
None	174 (58.2)	151 (67.1)	325 (62.0)
1	43 (14.4)	34 (15.1)	77 (14.7)
2	56 (18.7)	26 (11.6)	82 (15.7)
3	15 (5.0)	11 (4.9)	26 (2.3)
4	9 (3.0)	3 (1.3)	12 (2.3)
5	2 (0.7)	0 (0.0)	2 (0.4)

BTEC, Business and Technology Education Council; GCSE, General Certificate of Secondary Education; HNC, Higher National Certificate; HND, Higher National Diploma.

Primary outcome

The primary outcome of BDI-II scores improved in both arms at 12 weeks (*Table 3* and *Figure 6*). In the adjusted analysis the BDI-II score was slightly lower (by 0.46 points) in the intervention arm, but this was not statistically significant (95% CI –2.16 to 1.26; *p* = 0.60).

The analysis was adjusted for baseline BDI-II depression scores, baseline GAD-7 anxiety scores, sociodemographic factors (gender, age, socioeconomic position, housing, education, marital status and dependants), duration of depression, history of depression and clustering including a random effect for practice.

TABLE 3 Primary outcome (BDI-II score) at 12-week follow-up

Arm	Mean BDI score at baseline (SD)	Mean BDI score at 12 weeks (SD)	Mean adjusted difference (95% CI; *p*-value)[a]
Intervention	24.1 (8.89)	18.5 (10.17)	−0.46 (−2.16 to 1.26; *p* = 0.602)
	N = 302	N = 252	
Control	22.4 (9.52)	16.9 (10.3)	REF
	N = 227	N = 195	

REF, reference value.
a Adjusted for baseline depression, baseline anxiety, sociodemographics and practice as a random effect.
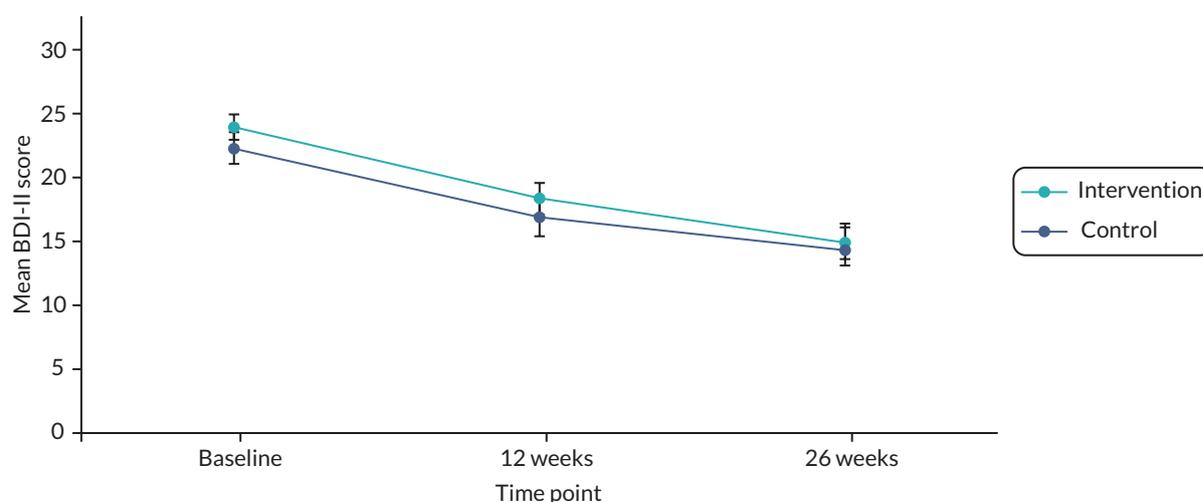
FIGURE 6 Mean BDI-II scores in intervention and control arms at 12 and 26 weeks.

Missing data

As a sensitivity analysis, we undertook the analysis of the primary outcome measure based on data imputed using a multiple imputation model including the outcome measure, baseline value of the outcome, randomisation group, clustering by practice and all covariates included in the analysis model. The inferences at 12 and 26 weeks were unchanged. The adjusted mean difference at 12 weeks was −0.18 (95% CI −1.82 to 1.45; *p* = 0.83) and at 26 weeks was −0.93 (−2.69 to 0.83; *p* = 0.30).

Secondary outcomes

The adjusted analysis of the BDI-II at 26 weeks found that both arms improved further. The score was again slightly lower in the intervention arm, but this was not statistically significant (mean adjusted difference −1.63, 95% CI −3.48 to 0.21; *p* = 0.082). The 95% CI includes a difference favouring the intervention arm by 3 points on the BDI-II, so we cannot completely exclude a clinically important difference in the outcome of depression at 26 weeks.

A similar pattern was seen for social functioning on the WSAS at 12 and 26 weeks, with scores improving between baseline and 12 weeks and improving further between the 12- and 26-week follow-ups. Again, there was no statistically significant difference between the trial arms on the WSAS. Looking back over the 26-week study period, the total scores on the MISS measure of satisfaction with care

were very similar between the arms, and the same was found for all the MISS subscales (see *Table 4*). All of the differences found in social functioning and patient satisfaction were in the direction of favouring the intervention arm, but none was statistically significant.

Post hoc analysis of 50% improvement and remission by arm

We conducted a post hoc analysis of categorical improvements in the BDI-II depression score at 26 weeks to further investigate possible differences in the outcome for depression given the relatively wide CIs at that point, and because we found a significant difference in the anxiety/depression scale of the EQ-5D-5L at 26 weeks (see *Chapter 4*). We compared the proportions of patients in each arm with BDI-II scores that improved by ≥ 50%, and the proportions of those who scored > 13 at baseline (the threshold for 'caseness' on the BDI-II) and subsequently remitted to a score of ≤ 13 by 26 weeks.

TABLE 4 Secondary outcomes at baseline and follow-up

Measure	Mean score at baseline (SD)	Mean score at follow-up (SD)	Mean adjusted difference (95% CI; *p*-value)[a]
BDI-II at 26 weeks			
Intervention (*n* = 226)	24.1 (8.96)	15.1 (10.84)	−1.63 (−3.48 to 0.21; *p* = 0.082)
Control (*n* = 184)	22.4 (9.52)	14.7 (10.65)	REF
WSAS at 12 weeks			
Intervention (*n* = 237)	17.3 (9.94)	14.7 (9.54)	0.48 (−1.03 to 2.00; *p* = 0.531)
Control (*n* = 195)	16.6 (10.06)	13.2 (9.90)	REF
WSAS at 26 weeks			
Intervention (*n* = 212)	17.3 (9.94)	11.6 (9.59)	1.34 (−3.20 to 0.53; *p* = 0.160)
Control (*n* = 183)	16.6 (10.06)	12.0 (9.99)	REF
MISS – total			
Intervention (*n* = 217)	N/A	121.8 (27.37)	5.39 (−1.39 to 12.16; *p* = 0.119)
Control (*n* = 176)	N/A	116.0 (26.75)	REF
MISS – distress relief			
Intervention	N/A	44.2 (11.21)	1.78 (−0.97 to 4.53; *p* = 0.205)
Control	N/A	42.1 (11.26)	REF
MISS – communication comfort			
Intervention	N/A	17.0 (4.03)	0.78 (−0.17 to 1.72; *p* = 0.107)
Control	N/A	16.3 (3.97)	REF
MISS – rapport			
Intervention	N/A	48.4 (10.61)	2.33 (−0.27 to 4.94; *p* = 0.079)
Control	N/A	46.4 (10.22)	REF
MISS – compliance intent			
Intervention	N/A	16.6 (4.01)	0.43 (−0.47 to 1.33; *p* = 0.355)
Control	N/A	15.9 (4.19)	REF

N/A, not applicable; REF, reference value.
a Adjusted for baseline depression, baseline anxiety, sociodemographics and practice as a random effect.

Table 5 shows the results of these post hoc analyses. The proportions improving by ≥ 50% did not differ significantly between the arms (45.1% vs. 37.3%, respectively), but the proportion of patients remitting in the intervention arm was significantly greater (49.8% vs. 39.9%; adjusted odds ratio 2.18, 95% CI 1.12 to 4.24; *p* = 0.02).

Recorded use of the Patient Health Questionnaire-9 questionnaire

Information on the recorded use of the PHQ-9 questionnaire in patients' medical records was collected for 261 intervention arm patients (86.4%) and 201 control arm patients (88.5%). In the intervention arm 190 patients (72.8%) had one or more PHQ-9 results recorded, ranging up to six results per patient. In the control arm 35 patients (17.4%) had one or more PHQ-9 results recorded, again ranging up to six results per patient. There was no apparent relationship between the number of PHQ-9 results recorded and the BDI-II scores at the 12-week follow-up.

Use of antidepressants

Table 6 shows the proportions of patients who reported taking antidepressants at baseline and at the 12-week and 26-week follow-ups. It is noteworthy that more than half of the patients recruited in both arms had already started taking antidepressants between their first presenting to the GP/NP and the researcher at the baseline assessment administering the first PHQ-9 questionnaire. This was permitted in the trial if practitioners thought it necessary for patients to commence treatment immediately, although we emphasised to practitioners that best practice is not to start treatment at the first consultation when patients present with new symptoms of depression, as a proportion of those patients will improve significantly within a week or two with support and active monitoring alone. More patients in the intervention arm reported using antidepressants at both the 12- and 26-week follow-up assessments, but again these differences were not statistically significant.

TABLE 5 Rates of 50% improvement in depression scores and of remission at 26 weeks

	Number (%) improved (by ≥ 50% of baseline score on the BDI-II) (*N* = 411)	Number (%) remitting (BDI-II score fell from > 13 to ≤ 13) (*N* = 349)
Intervention	102/226 (45.1)	100/201 (49.8)
Control	69/185 (37.3)	59/148 (39.9)
Adjusted odds ratio	1.53 (0.92 to 2.56; *p* = 0.101)	2.18 (1.12 to 4.24; *p* = 0.021)

TABLE 6 Self-reported antidepressant use at baseline and follow-up

	Number (%) using antidepressants at baseline	Number (%) using antidepressants at 12 weeks	Number (%) using antidepressants at 26 weeks
Intervention	166 (55.0)	150 (59.5)	123 (54.4)
Control	127 (56.0)	102 (52.3)	94 (50.8)
Odds ratio of difference[a] (95% CI; *p*-value)	N/A	1.82 (0.93 to 3.53; *p* = 0.083)	1.00 (0.58 to 1.74; *p* = 0.999)

a Adjusted for baseline depression, baseline anxiety, baseline antidepressant use, sociodemographics and practice as a random effect.

Medical records data were retrieved by general practice staff for 258 intervention arm (85.4%) and 201 control arm patients (88.5%) for the 26-week follow-up period. According to the medical records data, antidepressants were prescribed for 174 (67.4%) intervention arm and 112 (55.7%) control arm patients during the 26-week follow-up period, but the difference between the arms was not statistically significant, with an odds ratio of 1.83 (95% CI 0.96 to 3.48; p = 0.07) (adjusted for baseline depression, baseline anxiety, baseline antidepressant use, sociodemographics and practice as a random effect).

These proportions are greater than the proportions in the self-report data at 12 and 26 weeks, which might have been because the 12- and 26-week data were snapshots at those moments. It might also have been because prescriptions had been given and recorded in the notes but were not used by patients, or patients might not have been aware that the medications they were taking were antidepressants, or they might have forgotten they had taken them.

Adverse events

Two serious adverse events were reported. One patient in the control arm reported suicidal ideas; this patient was assessed by the researcher and trial principal investigator and found to be at higher risk, and the GP was informed immediately. The patient was referred by the GP to a community mental health team for assessment and withdrawn from the study. One patient in the intervention arm was hospitalised with COVID-19 and ketoacidosis, which was deemed severe and led to an outcome of long COVID and diabetes, but the event was not related to the trial.

Four non-serious adverse events were reported (three in the intervention arm and one in the control arm). In each case, the GP was informed and the patient remained in the study. One intervention arm patient had an attack of feeling paralysed on waking from sleep, which passed spontaneously. Another intervention arm patient reported drowsiness thought to be unrelated to their treatment. The third intervention arm patient reported doubling up on sleeping tablets on one occasion. One control arm patient reported injury from a car accident requiring hospital treatment, but this was unrelated to the trial.

The suicidal ideation SOP was triggered 318 times, 180 times for intervention arm patients and 138 times for control arm patients, which was in proportion to the number of patients in each arm. In each case, the P4 four-question suicide risk assessment screener[43] was implemented and the GP was informed. Altogether, 267 occurrences (intervention arm, n = 146; control arm, n = 121) were rated 'minimal risk', 38 (intervention arm, n = 25; control arm, n = 13) were rated 'lower risk' and 13 (intervention arm, n = 9; control arm, n = 4) were rated 'higher risk'. In four cases – two intervention and two control – the participants were withdrawn from the study, including the one (control arm) patient who was referred to the community mental health team.

Post hoc subgroup analysis of effects of coronavirus disease 2019 pandemic

Due to the COVID-19 pandemic and subsequent lockdown periods, we were aware it was possible that secular changes to the key outcomes had occurred during the study period that were unrelated to randomisation arm. We decided therefore to look at the scores in each arm in the pre, peri- and post-COVID-19 periods in the whole study population. We defined the 'pre' period as before 23 March 2020 and the 'post' period from 19 July 2021.

The results are shown in *Table 7*. There was no evidence that the timing of the BDI-II measurements relative to the COVID-19 pandemic period influenced the scores in a clinically meaningful way. Analysis of these data was likely be underpowered, but an exploratory analysis controlling for the timing of measurement suggested no significant impact on the trial inferences, with a between-group difference on the BDI-II at 12 weeks of –0.73 (95% CI –2.47 to 1.01; p = 0.411).

TABLE 7 Depressive symptom scores on the BDI-II and the COVID-19 pandemic

	Pre-COVID		Peri-COVID		Post-COVID	
	Control	**Intervention**	**Control**	**Intervention**	**Control**	**Intervention**
Mean (SD) baseline score	22.3 (9.35)	25.0 (9.45)	22.0 (8.39)	23.4 (8.48)	23.6 (11.70)	22.8 (7.57)
Mean (SD) 12-week score	16.6 (9.31)	20.0 (10.60)	17.1 (9.75)	18.6 (9.71)	16.9 (9.32)	14.6 (9.13)
Mean (SD) 26-week score	14.4 (9.92)	15.5 (10.88)	15.1 (11.03)	16.4 (11.81)	14.5 (10.72)	12.2 (8.41)

Chapter 4 Economic evaluation

Introduction

The economic evaluation was a key component of the study and involved the following stages:

1. Measurement of service use over the 6-month period prior to baseline assessment using the bespoke self-report questionnaire based on the CSRI.[40]
2. Comparison of service use over the 6 months prior to baseline in the intervention and control arms to check for differences at baseline, using the questionnaire data.
3. Measurement of service use over the 6-month trial period from baseline using data extracted from participating patients' computerised general practice medical records after their 26-week follow-up assessments.
4. Calculation of the costs of service use over the 6-month trial period, using the medical records data.
5. Comparison of service use and costs in the intervention and control arms over the 6-month trial period, again using the medical records data.
6. Calculation of the cost-effectiveness and cost–utility of the PHQ-9 intervention compared with usual care in the control arm. The primary analysis was over 26 weeks, undertaken from an NHS and PSS perspective.

Services recorded included those provided in the primary care setting (face-to-face GP and nurse consultations, GP and nurse telephone contacts, and GP and nurse e-mail or e-consult contacts), secondary care mental and physical health services (inpatient, outpatient, day patient, accident and emergency), community health services (e.g. health visitors, district nurses, counselling or psychological therapists) and social care services (e.g. social workers, housing workers). The questionnaire and medical records data extraction form were identical in structure and recorded whether or not patients had used specific services, how many contacts they had received and, where relevant, the average duration of service contact (i.e. across all contacts the individual made with each service). The names of medications were recorded along with the dose, frequency and duration of use.

The data extracted from patients' computerised general practice medical records were used in the primary analysis. The patient questionnaire data collected at the 26-week follow-up will be compared with the data collected from the medical records to look at the differences between them, but this was not a prime objective of the study.

The unit costs of health service use were derived from *Unit Costs of Health and Social Care*[45] for primary and community care; the *British National Formulary*[48] for costs of drug treatments; and the national NHS reference cost schedules[49] for secondary care costs.

Outcome measures

The outcomes were expressed as incremental cost per 1-point improvement in the BDI-II clinical outcome (cost-effectiveness analysis), and incremental cost per QALY gained (cost–utility analysis) using the EQ-5D-5L to calculate patient utilities.[39]

Analytical methods

A generalised linear mixed model was used to estimate the mean differences in costs and QALYs, adjusting for baseline characteristics including baseline BDI-II score, baseline anxiety, sociodemographic

factors and practice as a random effect. Bootstrapping methods were employed to estimate the incremental costs per BDI-II score and per QALY gained, together with their associated 95% CIs. CEACs were produced based on 1000 bootstrapping samples with replacement.

Self-reported resource use prior to baseline by arm

Table 8 shows a comparison between the intervention and control arms of self-reported NHS and social services resource use in the 6 months leading up to the baseline assessment according to the modified CSRI resource use self-report questionnaire. Intervention arm patients reported more face-to-face contacts with GPs, but fewer telephone, online and e-mail GP contacts, than control arm patients. One control arm patient reported prior contact with a psychiatrist, compared to none in the intervention arm. Fewer hospital outpatient visits, but slightly more inpatient stays, were reported in the intervention arm than in the control arm. Three control arm patients reported receiving home care visits at baseline, compared with one patient in the intervention arm.

TABLE 8 Self-reported resource use over 6 months prior to baseline assessment

Resource use	Intervention arm (N = 302)			Control arm (N = 227)		
	Number (%) of patients using resource	Mean number of times used	SD	Number (%) of patients using resource	Mean number of times used	SD
Prescriptions for medication	186 (61.6)	3	2.6	135 (59.5)	3	2.6
GP face-to-face contact	194 (64.2)	3	3.2	126 (55.5)	2.6	2.2
GP telephone contact	184 (60.9)	2.3	1.9	164 (72.2)	2.8	2.4
GP online or e-mail contact	23 (7.6)	1.9	1.9	24 (10.6)	2.1	2.3
GP video call	1 (0.3)	1	–	5 (2.2)	1.6	1.3
GP out-of-hours contact	10 (3.3)	1.5	0.7	14 (6.2)	1.4	0.8
Practice nurse face-to-face contact	82 (27.2)	1.6	1.3	67 (29.5)	1.7	1.5
Practice nurse telephone contact	6 (2.0)	6.3	6.9	20 (8.8)	7.3	7.1
Practice nurse online or e-mail contact	0	–	–	2 (0.9)	1	0
Practice nurse video call	0	–	–	1 (0.4)		–
Practice nurse out-of-hours contact	2 (0.6)	1.5	0.7	5 (2.2)	3	3.9
District nurse contact	4 (1.3)	2	1.4	0	–	–
Community mental health nurse contact	7 (2.3)	1.7	0.8	11 (4.8)	1.9	1.2
Other nurse contact	10 (3.3)	1.9	1.6	7 (3.1)	3	5.3
Health visitor contact	4 (1.3)	2.8	1.3	3 (1.3)	1.7	1.2
Counsellor contact	30 (9.9)	1.8	1	36 (15.9)	2.6	2.3
Other therapist contact	29 (9.6)	3	3	26 (11.5)	10.1	34.7
Alternative therapist contact	18 (6.0)	3.1	4.6	12 (5.3)	2.8	3.2
Psychologist contact	7 (2.3)	1.7	1.1	7 (3.1)	1.9	1.9
Psychiatrist contact	0	–	–	1 (0.4)	1	–
Community-based doctor contact	3 (1.0)	1.3	0.6	7 (3.1)	1	0
Occupational therapy contact	7 (2.3)	1.4	0.5	5 (2.2)	1.8	0.8
Social worker contact	2 (0.6)	3	2.8	1 (0.4)	2	–

TABLE 8 Self-reported resource use over 6 months prior to baseline assessment (*continued*)

Resource use	Intervention arm (*N* = 302)			Control arm (*N* = 227)		
	Number (%) of patients using resource	Mean number of times used	SD	Number (%) of patients using resource	Mean number of times used	SD
Home care contact	1 (0.3)	12	–	3 (1.3)	14	1.4
Community support worker contact	1 (0.3)	1	–	2 (0.9)	2.5	2.1
Hospital outpatient visit	72 (23.8)	1.9	1.5	70 (30.8)	2.5	3
Day-care attendance	7 (2.3)	1.6	0.8	7 (3.1)	1.4	1.1
Accident and emergency visit	33 (10.9)	2.8	9.2	17 (7.5)	1.4	1
Hospital inpatient stay	22 (7.3)	1.1	0.3	12 (5.3)	1.1	0.3
Other hospital service	2 (0.6)	2	1.4	6 (2.6)	2.8	3.5

Note
Data from patient self-report questionnaires.

Recorded resource use over 6 months' follow-up by arm

Table 9 shows a comparison of the resource use between the arms for 258 (85.4%) intervention arm and 201 (88.5%) control arm patients, based on data extracted from the patients' general practice medical records by practice staff after the end of the 26-week follow-up.

TABLE 9 Recorded service resource use over 6 months' follow-up

Resource use	Intervention arm (*N* = 258)			Control arm (*N* = 201)		
	Number (%) of patients using resource	Mean number of times used	SD	Number (%) of patients using resource	Mean number of times used	SD
Prescriptions for medication	230 (91.5)	4	3.2	172 (85.6)	4.2	3.9
GP face-to-face contact	169 (65.5)	3.7	3	121 (60.2)	3	2.1
GP telephone contact	188 (72.9)	4	3.1	159 (79.1)	4.3	3.8
GP online or e-mail contact	44 (17.1)	3.7	5.3	29 (14.4)	2.8	3.5
GP video call	6 (2.3)	1.5	0.8	0	–	–
GP out-of-hours contact	27 (10.5)	1.5	0.8	26 (12.9)	1.6	1.8
Practice nurse face-to-face contact	97 (37.6)	2.4	2.7	78 (38.8)	2.2	2
Practice nurse telephone contact	21 (8.1)	1.9	1.2	36 (17.9)	1.6	1
Practice nurse online or e-mail contact	1 (0.4)	1	–	2 (1.0)	2	0
District nurse contact	1 (0.4)	6	–	1 (0.5)	1	–
Community mental health nurse contact	10 (3.9)	1.9	1.7	10 (5.0)	3.5	2.8
Other nurse contact	37 (14.3)	2.2	1.7	18 (9.0)	2.2	2.3
Health visitor contact	1 (0.4)	4	–	1 (0.5)	1	–
Counsellor contact	37 (14.3)	1.8	2.3	25 (12.4)	3.1	5.3

continued

TABLE 9 Recorded service resource use over 6 months' follow-up (*continued*)

Resource use	Intervention arm (*N* = 258)			Control arm (*N* = 201)		
	Number (%) of patients using resource	Mean number of times used	SD	Number (%) of patients using resource	Mean number of times used	SD
Other therapist contact	38 (14.7)	1.8	2.3	31 (15.4)	2.3	3.6
Alternative therapist contact	7 (2.7)	1	–	4 (2.0)	1.3	0.5
Psychologist contact	14 (5.4)	2.1	2.7	13 (6.5)	2.1	2.1
Psychiatrist contact	5 (1.9)	1.6	0.9	6 (3.0)	1.3	0.8
Community-based doctor contact	6 (2.3)	1.7	0.8	5 (2.5)	1	0
Occupational therapy contact	1 (0.4)	1	–	2 (1.0)	1	0
Social worker contact	1 (0.4)	2	–	3 (1.5)	1.7	0.6
Home care contact	0	–	–	2 (1.0)	33	43.8
Community support worker contact	2 (0.8)	1	–	3 (1.5)	11	12.3
Hospital outpatient visit	94 (36.4)	2.1	1.5	78 (38.8)	2.6	3
Day-care attendance	12 (4.7)	1.2	0.4	7 (3.5)	1	0
Accident and emergency visit	44 (17.1)	1.3	0.6	34 (16.9)	1.3	0.6
Hospital inpatient stay	15 (5.8)	1.1	0.3	14 (7.0)	1.2	0.4
Other hospital service	27 (10.5)	1.9	2.2	27 (13.4)	1.3	0.6

Note
Data from general practice computerised medical records.

The proportion of patients in the intervention arm receiving any medications was 6% higher than in the control arm, a difference in keeping with the findings for antidepressant use reported in *Chapter 3*. More intervention arm patients had face-to-face contacts with GPs but fewer had telephone contacts, similar to the pattern found in the baseline self-report data in *Table 8*.

A slightly greater proportion of control arm patients had recorded hospital outpatient visits, inpatient stays and other hospital services. Two patients in the control arm received home care visits during the 26 weeks, with a mean number of 33 visits each in the 6 months, compared with none in the intervention arm.

Recorded contacts with mental health and social services
Overall, 90 out of 258 intervention arm patients (34.6%) and 68 out of 201 control arm patients (33.8%) had contacts with mental health and social services recorded in their medical records during the 26-week follow-up period (including contacts with community mental health nurse, counsellor, other therapist, psychologist, psychiatrist and social worker). The difference between the arms was not statistically significant (adjusted odds ratio 1.37, 95% CI 0.71 to 2.63; *p* = 0.342).

Costs

Costing the intervention
Modelling the likely cost of using the PHQ-9 as a PROM in routine clinical practice included making assumptions about the extra time needed for GPs/NPs to administer the initial questionnaire themselves (rather than the researcher doing it) in the non-trial situation, as well as for administering the

follow-up PHQ-9. In addition to the administration time, the cost would include the time spent going over the results of the PHQ-9 with the patient and discussing the possible implications of the score for the management of their depression.

We assumed that 10 minutes' extra GP time would be needed in both the initial and the follow-up consultations to administer the questionnaire and go over the results with the patient, effectively making each of these a double appointment, as GP appointments currently usually last 10 minutes. This means that a total of 20 minutes' extra GP time would be needed per patient in the intervention arm.

We estimated that 5 minutes' extra GP time would be needed to administer the questionnaire in practice. This was a conservative estimate as Spitzer *et al.*, the developers of the PHQ-9, observed that physicians took ≤ 3 minutes in 85% of cases.[50] In addition, we assumed that 5 minutes' extra time would be needed to go over each of the results with the patient, discussing individual symptoms as well as the overall score and the implications for treatment.

Other health economic appraisals of the use of the PHQ-9 have estimated that similar amounts of time would be needed. A study[51] modelling the likely cost-effectiveness of PHQ screening and collaborative care for depression in New York City assumed that 3 minutes of physician time would be needed to go over the findings of the PHQ-9 questionnaire, in addition to 6 minutes of nurse time spent administering it. Another study[52] modelling the cost–utility of screening for depression in primary care again assumed nurse time of 6 minutes and physician time of only 1 minute to view the PHQ-9 results, but it did not include physician time to go over the results with the patient.

A proportion of the (maximum) cost of training on the PHQ-9 of 2 hours of GP time was also included in the cost of the intervention, discounted over 100 patients, as we assumed that the training would last for at least the assessment of that number of patients before it might need to be refreshed. This gave a total cost of approximately £33 per patient whose depression was monitored with the PHQ-9.

Costs over 6 months' follow-up by arm

Table 10 shows a comparison of the estimated costs of resources used between the intervention and control arms over the 6 months' follow-up.

The costs of GP care (including face to face, telephone, online or e-mail, and video calls) were slightly higher in the intervention arm. Hospital costs (including outpatient, inpatient and other hospital services) were higher in the control arm. The mean cost per patient of home care was particularly high, although only two patients in the control arm used the home care service.

The total mean per-patient cost of resources used over 6 months was £1124 (SD £1371) in the intervention arm, compared with £1292 (SD £1214) in the control arm, a relatively small and non-significant unadjusted mean saving of £168 per patient.

Utility scores

The EQ-5D-5L was used to measure quality of life at the baseline assessment and at the 12- and 26-week follow-up points. The EQ-5D-5L is the measure NICE favours in determining cost-effectiveness when developing its clinical guidelines. The EQ-5D-5L has five dimensions, mobility, self-care, usual activities, pain/discomfort and anxiety/depression, each scored on five levels.

Health states are converted into a single summary index by applying weights to each level in each dimension derived from the valuation of EQ-5D-5L health states in adult general population samples.[39] Crosswalk methods were applied to derive utility scores using the algorithm for the EQ-5D-5L.[39]

TABLE 10 Costs of resource use over 6 months' follow-up

Resource	Intervention arm (N = 258)			Control arm (N = 201)		
	Number (%) of patients using resource	Mean cost (£)	SD	Number (%) of patients using resource	Mean cost (£)	SD
Prescriptions for medication	230 (91.5)	25	28	172 (85.6)	28	34
GP face-to-face contact	169 (65.5)	646	527	121 (60.2)	518	390
GP telephone contact	188 (72.9)	568	450	159 (79.1)	623	612
GP online or e-mail contact	44 (17.1)	315	448	29 (14.4)	229	295
GP video call	6 (2.3)	127	71	0	–	–
GP out-of-hours contact	27 (10.5)	240	138	26 (12.9)	282	315
Practice nurse face-to-face contact	97 (37.6)	23	31	78 (38.8)	23	26
Practice nurse telephone contact	21 (8.1)	12	8	36 (17.9)	11	9
Practice nurse online or e-mail contact	1 (0.4)	4	–	2 (1.0)	9	0
District nurse contact	1 (0.4)	84	–	1 (0.5)	14	–
Community mental health nurse contact	10 (3.9)	35	31	10 (5.0)	64	53
Other nurse contact	37 (14.3)	21	18	18 (9.0)	26	32
Health visitor contact	1 (0.4)	31	–	1 (0.5)	8	–
Counsellor contact	37 (14.3)	50	54	25 (12.4)	92	165
Other therapist contact	38 (14.7)	41	38	31 (15.4)	66	110
Alternative therapist contact	7 (2.7)	8	2	4 (2.0)	10	4
Psychologist contact	14 (5.4)	82	114	13 (6.5)	83	66
Psychiatrist contact	5 (1.9)	90	50	6 (3.0)	75	46
Community based doctor contact	6 (2.3)	36	18	5 (2.5)	22	0
Occupational therapy contact	1 (0.4)	51	–	2 (1.0)	51	0
Social worker contact	1 (0.4)	23	–	3 (1.5)	59	35
Home care contact	0	–	–	2 (1.0)	1568	2127
Community support worker contact	2 (0.8)	23	–	3 (1.5)	253	283
Hospital outpatient visit	94 (36.4)	283	199	78 (38.8)	362	405
Day care attendance	12 (4.7)	949	316	7 (3.5)	813	0
Accident and emergency visit	44 (17.1)	174	80	34 (16.9)	173	78
Hospital inpatient stay	15 (5.8)	2066	2351	14 (7.0)	999	1075
Other hospital service	27 (10.5)	259	300	27 (13.4)	183	85
Intervention (use of PHQ-9 PROM)	258 (100)	33	0	0	–	–
Mean total cost of services per patient (£)	**258**	**1124**	**1371**	**201**	**1292**	**1214**

Table 11 and *Figure 7* show the utility scores at the baseline assessment and at each follow-up point. Quality of life improved in both arms between baseline and 12-week follow-up. It then improved further in the intervention arm but went down slightly in the control arm. The difference between arms was not statistically significant at 12 weeks (estimated difference in utility score –0.002; $p = 0.94$). However, the difference was statistically significant at the 26-week follow-up, favouring the intervention (estimated

TABLE 11 EuroQol-5 Dimensions, five-level utility scores and QALYs over 26 weeks' follow-up

Arm	Time point	Recorded number	Mean EQ-5D-5L scores or QALYs gained	SD
Intervention (N = 302)	Baseline	302	0.659	0.232
	12 weeks	256	0.694	0.236
	26 weeks	221	0.718	0.249
	QALY	215	0.346	0.104
Control (N = 227)	Baseline	226[a]	0.667	0.226
	12 weeks	197	0.708	0.213
	26 weeks	183	0.696	0.225
	QALY	177	0.344	0.098

a One patient withdrew immediately after enrolment and did not complete an EQ-5D-5L at baseline.
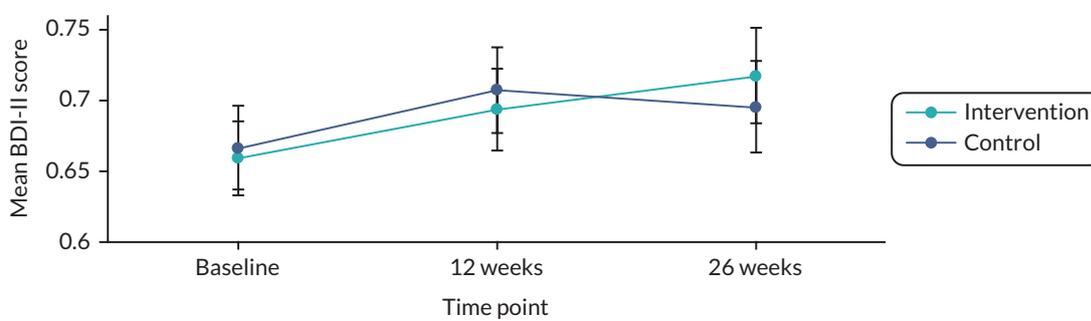
FIGURE 7 Mean EQ-5D-5L utility scores at baseline and at 12- and 26-week follow-ups.

difference 0.053, 95% CI 0.013 to 0.093; p = 0.01). The analysis was adjusted for baseline EQ-5D-5L, history of depression, baseline anxiety, sociodemographics and practice as a random effect.

Changes in the five dimensions of the EuroQol-5 Dimensions, five-level
Table 12 shows the changes in the proportions of patient responses for the five dimensions of the EQ-5D-5L from baseline to the 12- and 26-week follow-up assessments. Each dimension has five levels from 1 to 5, representing no problems, slight problems, moderate problems, severe problems and extreme problems.

Patients in the two arms were similar at baseline in mobility, self-care, and pain and discomfort, and remained so throughout the 26 weeks' follow-up. More patients in the intervention arm were at levels 4 and 5 (severe or extreme problems) for anxiety and depression at baseline. At 26-week follow-up there were slightly more patients at level 1 (the lowest level, indicating no problems) for the usual activity dimension in the intervention arm (55.2% vs. 48.1% in the control arm). The biggest difference between the arms at 26 weeks, however, was in the proportions reporting no problems in the anxiety and depression dimension (22.6% in the intervention arm vs. 13.5% in the control arm). The improvement in the anxiety and depression dimension therefore seems to have contributed most to the overall greater improvement in the mean score for quality of life on the EQ-5D-5L in the intervention arm.

TABLE 12 Changes in the five EQ-5D-5L dimensions at 12- and 26-week follow-ups

EQ 5D-5L dimension	Level	Baseline, n (%)		12 weeks, n (%)		26 weeks, n (%)	
		Intervention (N = 302)	Control (N = 226)[a]	Intervention (N = 256)	Control (N = 198)	Intervention (N = 221)	Control (N = 185)
Mobility	1	237 (78.5)	170 (75.2)	188 (73.4)	151 (76.3)	155 (70.1)	125 (67.6)
	2	33 (10.9)	28 (12.4)	29 (11.3)	29 (14.6)	35 (15.8)	38 (20.5)
	3	20 (6.6)	19 (8.4)	25 (9.8)	13 (6.6)	22 (10.0)	14 (7.6)
	4	10 (3.3)	5 (2.2)	12 (4.9)	4 (2.0)	6 (2.7)	6 (3.2)
	5	2 (0.7)	4 (1.8)	2 (0.8)	1 (0.5)	3 (1.4)	1 (0.5)
Self-care	1	257 (85.1)	189 (83.6)	213 (83.2)	162 (81.8)	182 (82.4)	152 (82.2)
	2	30 (9.9)	22 (9.7)	24 (9.4)	24 (12.1)	22 (10.0)	19 (10.3)
	3	11 (3.6)	13 (5.8)	13 (5.1)	8 (4.0)	13 (6.0)	8 (4.3)
	4	3 (1.0)	2 (0.9)	5 (2.0)	3 (1.5)	4 (1.8)	4 (2.2)
	5	1 (0.3)	0	1 (0.4)	0	0	1 (0.5)
Usual activity	1	104 (34.4)	87 (38.5)	111 (43.4)	90 (45.5)	122 (55.2)	89 (48.1)
	2	88 (29.1)	76 (33.6)	89 (34.8)	64 (32.3)	54 (24.4)	55 (29.7)
	3	83 (27.5)	43 (19.0)	37 (14.5)	27 (13.6)	34 (15.4)	29 (15.7)
	4	19 (6.3)	14 (6.2)	19 (7.4)	11 (5.6)	9 (4.1)	10 (5.4)
	5	8 (2.6)	6 (2.7)	0	5 (2.5)	2 (0.9)	1 (0.5)
Pain and discomfort	1	145 (48.0)	103 (45.6)	117 (45.7)	87 (43.9)	102 (46.2)	80 (43.2)
	2	88 (29.1)	63 (27.9)	78 (30.5)	71 (35.9)	64 (29.0)	61 (33.0)
	3	42 (13.9)	42 (18.6)	41 (16.0)	28 (14.1)	39 (17.6)	22 (12.0)
	4	21 (7.0)	17 (7.5)	17 (6.6)	10 (5.1)	12 (5.4)	20 (10.8)
	5	6 (2.0)	1 (0.4)	3 (1.2)	2 (1.0)	4 (1.8)	1 (0.5)
Anxiety and depression	1	24 (7.9)	17 (7.5)	38 (14.8)	27 (13.6)	50 (22.6)	25 (13.5)
	2	73 (24.2)	66 (29.2)	83 (32.4)	83 (41.9)	94 (42.5)	85 (45.9)
	3	134 (44.4)	99 (43.8)	95 (37.1)	65 (32.8)	52 (23.5)	53 (28.6)
	4	54 (17.9)	33 (14.6)	31 (12.1)	19 (9.6)	19 (8.6)	19 (10.3)
	5	17 (5.6)	11 (4.9)	9 (3.5)	4 (2.0)	6 (2.7)	3 (1.6)

a One patient withdrew immediately after enrolment and did not complete the EQ-5D-5L at baseline.

Quality-adjusted life-years

Quality-adjusted life-years were calculated using the area under the curve approach. The baseline utility score was added to the score at 12 weeks and this total was divided by 2, based on the assumption of a linear change over the 12-week period. This figure was then multiplied by 0.25, as only one-quarter of a QALY could be gained over the 12-week period. The QALY gain in the 12- to 26-week period was calculated in a similar way. Gains in QALYs over the entire 26-week follow-up period were calculated by adding these two QALY gains.

Table 11 shows that the mean QALY gain between baseline and the 26-week follow-up was 0.346 (SD 0.104) for the intervention arm and 0.344 (0.098) for the control arm. The difference was not statistically significant (estimated difference 0.008; *p* = 0.26). The analysis again was adjusted for baseline EQ-5D-5L, baseline anxiety, history of depression, sociodemographics and practice as a random effect.

Cost-effectiveness analysis

Data on the costs of services used were linked with the BDI-II scores to assess the possible cost-effectiveness of the intervention. In *Chapter 3* we showed that the improvement in depressive symptoms at the 12-week follow-up in the intervention arm on the BDI-II (the primary outcome) was very similar to that in the control arm.

However, given that the costs of service resource use in the intervention arm were lower than the costs in the control arm for the same improvement in outcome, it was necessary to compute ICERs to assist decision-makers in assessing whether adding monitoring with the PHQ-9 to usual GP/NP care for depression represents value for money.

Given that the cost data are skewed, which is frequently the case and can cause a violation of the assumptions of standard significance tests, bootstrapped estimating (multiple resampling of pairs of values for patients within treatment arms) was carried out so that estimated mean costs could still be compared while imposing no prior assumptions regarding the data distribution.

Table 13 shows the costs in the two arms together with the BDI-II scores at 12 weeks and the incremental cost per point change in the BDI-II score. The mean costs were estimated using bootstrap methods with 1000 resamples, which is why they differ slightly from the raw data in *Table 10*.

The ICER of £129 is the adjusted mean saving per point improvement in the BDI-II in the intervention arm compared with usual care in the control arm (the negative value for the incremental change is in favour of the intervention). The 95% CI is again wide and includes zero.

Cost–utility analysis

Similarly, data on the costs of services used were linked with the values found for QALYs gained to assess the possible cost–utility of the intervention. *Table 14* shows the costs in the two arms together with the QALYs gained over 26 weeks and the incremental cost per QALY gained. The mean costs were again estimated using bootstrap methods with 1000 resamples, and the QALYs were estimated based on imputed quality-of-life data.

TABLE 13 Incremental cost per point change on the BDI-II score at 12 weeks (ICER)

Arm	Cost (£), mean (95% CI)	Incremental cost (£), mean (95% CI)	BDI-II score at 12 weeks, mean (95% CI)	Incremental BDI-II score,[a] mean (95% CI)	ICER (£ per point on the BDI-II score), mean (95% CI)
Intervention	1131 (1010 to 1269)	−163 (−349 to 28)	18.5 (17.2 to 19.7)	−0.47 (−1.77 to 0.83)	129 (−1185 to 1939)
Control	1294 (1160 to 1437)		16.9 (15.5 to 18.5)		

a Adjusted by baseline BDI-II score, baseline anxiety, sociodemographic factors and practice as a random effect.

TABLE 14 Incremental cost per QALY gained over 26 weeks (ICER)

Arm	Cost (£), mean (95% CI)	Incremental cost (£), mean (95% CI)	QALYs, mean (95% CI)	Incremental QALY, mean (95% CI)	ICER (£ per QALY), mean (95% CI)
Intervention	1131 (1010 to 1269)	−163 (−349 to 28)	0.347 (0.337 to 0.356)	0.0013 (−0.0157 to 0.0182)	−5216 (−109,336 to 95,761)
Control	1294 (1160 to 1437)		0.346 (0.336 to 0.356)		

The ICER of −£5216 is the mean saving per QALY gained in the intervention arm compared with usual care in the control arm. (Here, the positive value for the incremental change is in favour of the intervention.) Again, the 95% CI is very wide and includes zero (see *Table 14*).

Cost-effectiveness plane

The above calculation is based on the mean costs and differences in QALYs gained and therefore does not take into account uncertainty around these estimates. To address such uncertainty, a cost-effectiveness plane was produced to show the probability that the intervention arm would have higher or lower costs and better or worse outcomes than usual care in the control arm. *Figure 8* shows the scatterplot of the comparison of intervention and control arms from the bootstrapped analyses using 1000 resamples of pairs of values, including the 95% confidence ellipse.
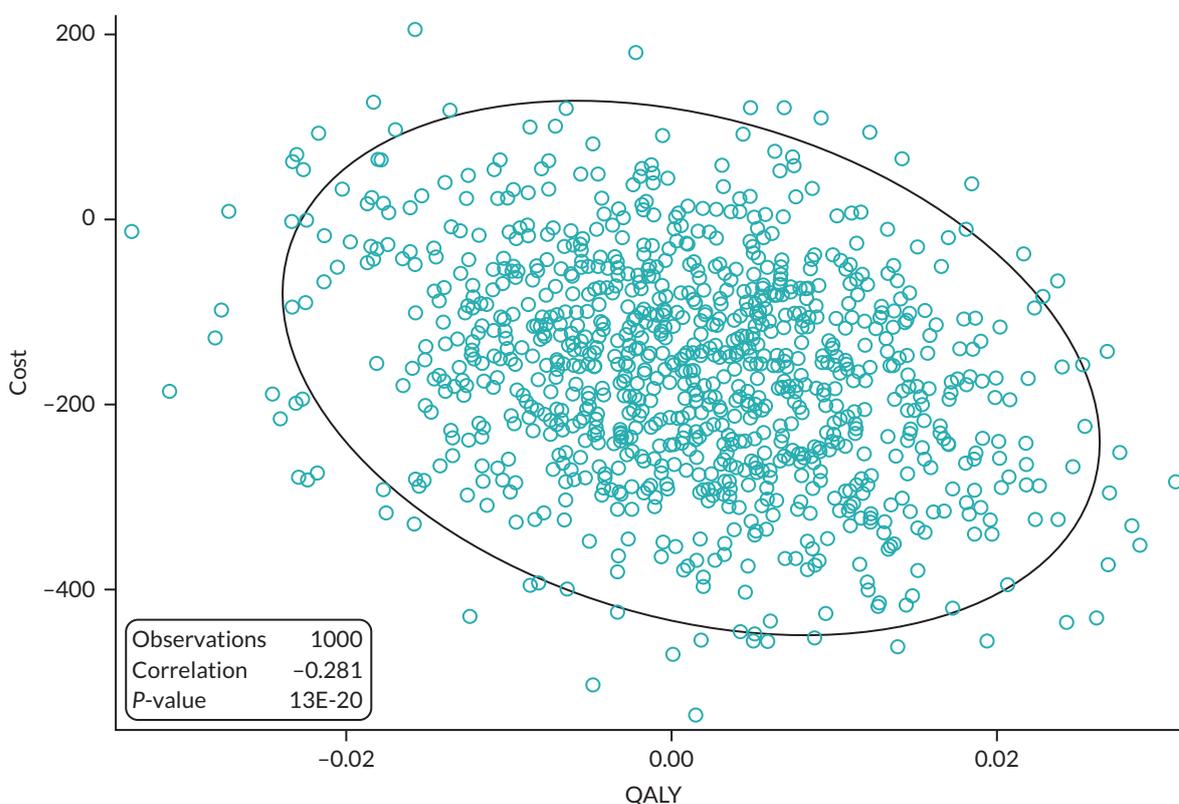
FIGURE 8 Cost-effectiveness plane of mean cost and mean QALY gain over 6 months. Scatterplot with 95% prediction ellipse.

Cost-effectiveness acceptability curve

To further explore the uncertainty around the cost–utility estimate for the intervention, a CEAC was produced to model the likelihood of the intervention being cost-effective at varying values of societal willingness to pay placed on a QALY gained, compared with usual care in the control arm.

Figure 9 shows the cost-effectiveness curve for the intervention based on the QALY gains found over 6 months. The probability that the intervention would be cost-effective compared with usual care in the control arm was 77% and 72%, respectively, at the lower and higher thresholds of societal willingness to pay adopted by NICE, of £20,000 and £30,000 per QALY gained, used for judging the relative cost-effectiveness of interventions.

Sensitivity analysis

We assumed above that the total time taken to administer the PHQ-9 and discuss the results with the patient in routine practice outside the trial situation would be approximately 20 minutes (an extra 10 minutes during each of the first and second consultations). We also included the training cost of up to 2 hours of GP time, discounted over 100 patients. The influence of these assumptions on cost-effectiveness was tested through a sensitivity analysis assuming that only 5 minutes of extra time would be needed at each of the initial and follow-up consultations if the patient had already completed the questionnaire prior to each consultation, and assuming no extra cost attached to the intervention for the training, which in practice could take place during routine GP vocational training.

The ICER computed using this revised costing of the intervention was a mean saving of £6243 per QALY gain over 6 months (95% CI –£122,625 to £100,543), instead of £5216 (95% CI –£109,336 to £95,761) per QALY. The CEAC for the sensitivity analysis gave probabilities of the intervention being cost-effective at the £20,000 and £30,000 per QALY thresholds of 79% and 74%, respectively. So, the probability of cost-effectiveness was not particularly sensitive to the estimated time spent by the practitioner in going over the results of the PHQ-9 with the patient in the two consultations.
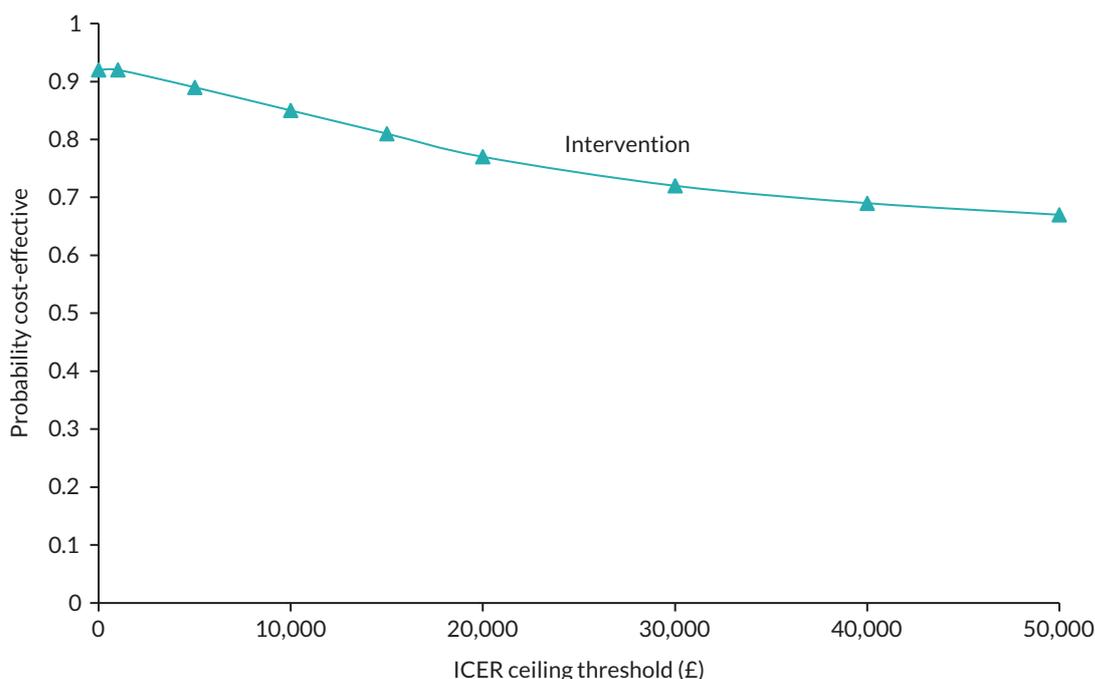
FIGURE 9 Cost-effectiveness acceptability curve comparing the intervention with usual care.

Chapter 5 Qualitative process evaluation

Objectives and methods

The objectives of the qualitative process evaluation in the trial were to identify, characterise and explain the perspectives of patient and practitioner participants on the conduct of the trial and the use of the PHQ-9 as a PROM. If patients' outcomes were improved, this would enable the construction of a taxonomy of factors affecting the potential for the use of PROMs to be normalised in everyday practice, outside the trial situation, using NPT as a framework.[47] A summary of the key domains of the NPT framework is provided in *Appendix 3*.

Interviews were conducted by telephone or online call. Semistructured interview guides were used for the patient interviews and practitioner interviews, which were developed collaboratively in the research team, in consultation with our PPI colleagues. Interviews were audio-recorded and transcribed verbatim, with potentially identifying information removed from transcripts to ensure participant confidentiality and anonymity. If a participant became distressed by discussion of personal experiences of depression, the interviewer would offer to suspend the interview and return another time. If a risk of self-harm was suspected, the interviewer followed the suicidal ideation SOP (see *Appendix 2*), including sharing relevant information with patients' GPs if necessary.

Analysis

We conducted reflexive thematic analysis.[53] We sought immersion in the data by reading and re-reading all transcripts and reflecting on interviews and discussing potential themes in research team meetings. Two academic psychologist researchers, BCFC and RDH, independently coded a set of transcripts and collaboratively developed an initial coding frame. This framework was then used to code subsequent transcripts and was iteratively extended and revised as new codes were identified. Coding was inductive and derived from the data. However, we also applied the NPT lens for implementation evaluation in health settings[47] when approaching the data. We drew on insights from the wide range of studies that have employed NPT, giving a basic structure to the topic guide written in advance of the interviews and refined iteratively as analysis proceeded.

We worked prospectively and inductively to ensure that we identified, characterised and understood disconfirming evidence and processes that were not accounted for within NPT. This was to ensure that we considered context–mechanism–outcome domains in the data that were potentially vital in translating research findings to service provision. BCFC developed the final set of themes from the coded data, which were refined with feedback from RDH, the rest of the research team and our PPI colleagues.

Quality

We adhered to frameworks for conducting and writing up high-quality qualitative research.[53] To maximise the validity of the results, multiple researchers were involved in data collection, coding and analysis. At different stages of the analysis, we presented preliminary findings to the research team and discussed the face validity of themes.

To capture diverse experiences and views of participants to ensure that our findings were transferable to practitioners and patients in the UK, we attempted to recruit a group of participants who represented different characteristics and identities. This also enhanced the potential transferability of our findings.

We also sought to increase the trustworthiness of our interpretations and the transparency of the lenses that guided our research team's approach to the data. We evidence our analyses through anonymised quotations from participants below.

Reflexivity

Reflexive strategies were built into the research process from design through to collection, analysis and reporting. The key researchers kept reflexive journals. The wider research team members who commented on the emerging analysis were a diverse group of academic and clinical researchers at different career stages, including research assistants, research fellows and professors. Study team members ranged in gender, culture and ethnicity, including black, Asian and white British; had expertise across multiple specialties, including general practice, health psychology and psychiatry; and brought a range of research interests and experiences into the study, including qualitative methods, development and evaluation of complex interventions, implementation science and primary care service development. This array of heterogeneous identities and perspectives enhanced our knowledge and insight in this study, but we maintained curiosity and reflected on our preconceived notions, allowing diverse ideas to be developed inductively from the data.

Results

Patient interviews took place between October 2019 and July 2022 and lasted between 13 minutes and 1 hour 13 minutes (average 34 minutes). GP/NP interviews took place between August 2019 and July 2022 and lasted between 15 minutes and 1 hour 57 minutes (average 38 minutes). *Tables 15* and *16* show the numbers and characteristics of patients and practitioners interviewed.

Five inductive themes were identified after the analysis of all interview data, which are discussed below. Anonymised participant identification codes are used throughout to maintain participant confidentiality.

Themes, subthemes and illustrative quotations

1. Improved understanding of depression
   Recognising symptoms
   Monitoring over time
   Motivation and hope
2. Usability of the PHQ-9
   Being pigeonholed
   Cognitive aspects
   Accessibility
3. Impact of using the PHQ-9 on the consultation
   Driver of discussion
   Patient–GP relationship
   Person-centred care
4. Impact of the PHQ-9 on the practitioner care of depression
   Evidence to inform treatment and management
   Objectivity versus subjectivity
5. Organisational barriers to and facilitators of using the PHQ-9 in practice
   Time restraints
   Technological integration
   Frameworks and guidelines

TABLE 15 Participant characteristics: patient interviews

Characteristic	Intervention	Control
N = 29	*N = 18*	*N = 11*
Gender		
Female	10	8
Male	8	3
Age, mean years (SD)	36.7 (12.7)	46.8 (19.7)
Ethnicity		
White	15	10
Chinese	1	0
Indian	1	0
African/Irish	1	0
Persian	0	1
Marital status		
Married/cohabiting	7	7
Single	11	3
Divorced	0	1
Education level		
None	2	0
CSE/NVQ Level 1	1	0
GCSE/O Level/NVQ Level 2	3	1
HNC/HND/city and guilds/teaching qualification/NVQ Level 4	2	1
Degree/higher degree/NVQ Level 5	10	7
Vocational qualification	0	2
Employment status		
Full-time work/self-employed	11	5
Part-time work	2	5
Homemaker	1	0
Retired	1	1
Student	3	0
IMD[a] (median)	6.5	9

CSE, Certificate of Secondary Education; GCSE, General Certificate of Secondary Education; HNC, Higher National Certificate; HND, Higher National Diploma; NVQ, National Vocational Qualification.
a Index of Multiple Deprivation score for practice area.

Improved understanding of depression

Recognising symptoms

Patients and practitioners described how using the PHQ-9 could help patients recognise that the range of different symptoms they were experiencing, in the context often of adverse life events and

TABLE 16 Participant characteristics: practitioner interviews

Characteristic	Intervention	Control
N = 15	*N* = 11	*N* = 4
Gender		
Female	4	3
Male	7	1
Location		
Urban	8	3
Rural	3	1
Practice size		
Large	7	3
Small	4	1
IMD[a] (median)	7	7.5

a Index of Multiple Deprivation score for practice area.

difficulties, could be understood better as occurring together as part of the recognised syndrome of depression. The PHQ-9 was described as helpful in identifying the severity categories, particularly for patients having a first episode of depression. Patients' accounts and researchers' diaries suggested that seeing each item and the infographic helped to validate experiences of depressive symptoms and provide a way to understand them together as depression:

> **Patient PT03037-03:** *I didn't realise half of the symptoms and I'm guessing other people don't either, unless you know or you have an inkling that you are depressed, you probably aren't going to – look into it. It's probably not something you want to look into either; so having that there is really useful.*
>
> **Patient PT02024-03:** *As I was reading [the items] I was thinking to myself – yes, that is me, that is me and I kind of understood it bit more.*

Both groups described how the PHQ-9 provided a global assessment and explicit marker of the severity of patients' difficulties, but also provided insight into the areas with which individual patients needed help:

> **Practitioner GP01056-01:** *It's like another little bit of a jigsaw really … You've got someone who is coming in and you're like, you're in the severe zone, I know that … Gosh, you really are struggling.*
>
> **Patient PT03037-03:** *It made me realise I'm apparently more depressed than I thought I was. There are a few points here that I didn't realise meant you're depressed, like concentration.*

Some patients found that the PHQ-9 score and infographic were complementary and helped them to understand how they were doing. These could be reassuring if the severity level was not too high. However, they could also be distressing if there was dissonance, making patients feel judged by the severity category when the PHQ-9 outcome did not align with their own expectations. For some individuals this was a sharp 'wake-up call' to seek support but also gave them solace that they were not alone in their struggles:

Patient PT01023-02:   *It reassured me that I wasn't going mad, that it is an illness, and it does affect a lot of other people as well.*

Patient PT01026-02:   *I guess when you're feeling really rubbish and then you look at it and you're like actually, well, I haven't got all of these things on here, so it's not as bad as it could be [laughter].*

However, others did not necessarily feel that the PHQ-9 score accurately reflected how they were feeling. Some found the score difficult to accept, and the infographic could invoke feelings of fear and anxiety:

Patient PT03059-01:   *I wouldn't necessarily describe myself as seriously depressed but maybe I am … That was difficult to process … Let's say on your way to work, you see 100 people. Oh, I'm one of two people.*

Monitoring over time

The PHQ-9 was reported to be a useful tool for GPs to monitor their patients' depressive symptoms over time. Follow-up scores provided an opportunity for reflection and to identify improvements or deteriorations and relative progress:

Practitioner GP02013-01:   *I used to just say it's working or it's not. So if the PHQ-9 score was decreasing, then I would congratulate the patient and say – maybe, you know, you're not feeling that to an extent, but look at the number that you scored some time ago because they've forgotten what the number was, but they've come with this new number, which could be 13 and it would be 24 in the past. So I say, well, that's great progress.*

Symptoms were being monitored, and there was relief if things were getting better, although sometimes the scores were felt not to be nuanced enough:

Patient PT01002-04:   *There have been times where it has gone back up a little bit but then it comes back down again. It is one of those things – sometimes it's a mixture. If it goes back up slightly you do sometimes feel a little bit disappointed but it's definitely helpful to see what areas you need to focus on and where to get better.*

Practitioner GP03044-01:   *I could see patients improving and patients didn't realise they were improving until they saw the scores and that was quite a pleasant discovery for them and that motivated them … Especially when they're depressed, they can't really see a way out and can't see the progress that they're making, but it was a nice way for the patient to see that they've actually made progress.*

Patient PT01002-04:   *Looking at the questionnaire my score had gone down … I left [the appointment] with that sense of relief that stuff was helping and I was starting to feel a bit better.*

Many patients reported that recognising changes in their depressive symptoms was difficult without the 'physical' scores on the questionnaire. The PHQ-9 also enabled patients to progressively 'map out' necessary steps and areas of improvement:

Patient PT03037-03:   *It's mainly mapping to see which bits needed to be improved and which bits could be not focused on for a little while, what to work on or how to make yourself feel better.*

Motivation and hope

Using the PHQ-9 initially and realising the severity of their depression motivated some patients to actively seek support and address their problems.

> **Patient PT01002-04I:** *I don't think it made me feel worse. If anything, it put things into perspective as to how bad I was feeling and how much – I was grateful that I was able to see that because I knew then how much I needed that help and how important it was.*
>
> **Patient PT01096-04:** *It was a bit of a shock at first … but it certainly did help, I have to say. It is fairly bad, so you've just got to try and get on with it and get better.*

Over time, seeing improvements motivated patients to continue their efforts to try to get better and emphasised that there was worth in their perseverance. It also gave them hope that progress was taking place.

> **Patient PT01002-04:** *When I see it go down, you're relieved. You're like, oh, it's working; it's worth getting help.*

Usability of the Patient Health Questionnaire-9

Being pigeonholed

GPs and patients described the response options on the PHQ-9 as pushing patients to be pigeonholed into arbitrary categories. Patients perceived this as difficult to do, inaccurate in reflecting their actual problems, and an oversimplification of their complex experiences:

> **Patient PT01002-06:** *You just get judged by these 9 questions, to assess how severe you are and how ill you are and that becomes […] people are worried about becoming numbers and becoming just data. […]. It can almost feel like an efficient method to judge something that's almost impossible to understand fully, which I guess it is.*

Some GPs reported that the items were not clear, which resulted in the loss of nuanced understanding of their patients' difficulties; for example, daily changes were not identified or items unhelpfully captured both ends of a spectrum at once. This affected the quality of symptom-specific care, which required GPs to gather more information beyond the pre-ordained symptom categories and scales, which could work to conflate or even 'lump together' different types of concern:

> **Practitioner GP01002-01:** *I just got a score for the symptoms and need to discuss: Do they comfort eat? Do they skip meals? All these things are not separated. The PHQ-9 asks are you overeating or are you skipping meals? Well, what is it? It's not clear, it could be either … Maybe they need a referral to an eating disorder clinic or maybe they need to see a dietitian.*

Some patients echoed this but also said that the PHQ-9 only considers the frequency of symptoms over the period of 2 weeks. This was important for patients for whom frequency of symptoms did not necessarily change but the impact and severity of symptoms did. Not being able to reflect tiny but meaningful triumphs was disappointing and ultimately worked to obfuscate rather than illuminate changes that were meaningful to the patient and potentially helpful for the GP:

> **Patient PT01096-05:** *It only had three points and I felt like it wasn't enough for me to say, 'I am improving, but actually I'm not improving enough to go to the next point on the scale' … The frequency of symptoms remained the same, but the intensity has decreased … I was disappointed that I couldn't reflect my progress.*

Cognitive aspects

General practitioners and patients found the PHQ-9 easy to fill in and the visuals of the infographic understandable. The brevity of the questionnaire was reported to be a relief, especially for those patients whose scores indicated moderate to severe depressive symptoms and for whom concentration may be a problem:

> **Patient PT01096-04:** *If someone would have just said to me, 'this is your score, this is what it means,' I would have been a little bit confused but having something to look at and analyse, for me it gave me a lot better of an understanding.*
>
> **Patient PT03037-03:** *Nine questions is easy compared to other questionnaires I had to do ... My head was in a very odd place, it felt easy to focus on a small document ... Especially if you do come under like severe or moderate, you probably don't want to read a large bulk of information straightaway.*
>
> **Practitioner GP03044-01:** *I think it was quite quick and easy and patients didn't struggle to fill it in; it was quite you know, patient-centred. They understood the terminology, there weren't any queries about, you know, clarifying anything; I think it was quite good.*

However, a minority of patients found various aspects difficult. One patient found themselves overthinking their response to each item when completing the PHQ-9 independently, which made the process time-consuming. Another also reported that their answers differed hugely depending on their mood at the time of completing the PHQ-9. Reflecting on their week was sometimes helpful.

> **Patient PT03059-01:** *If I was filling it in thinking, I've had such a shit week, blah ... and then I looked back through my diary and hang on a minute. I really enjoyed that day and that day and they were all within the week ... How I'm feeling on the particular day that I'm filling it in is reflected on the whole week, rather than just that day ... Looking in my diary I did feel okay that day, why have I said that it was awful?*

Accessibility

Both patients and GPs had mixed perceptions of the accessibility of the PHQ-9. Some preferred to be able to complete it independently at home online or over the telephone, whereas others wanted to complete it on paper. GPs found that some patients needed help from their doctor to complete it:

> **Practitioner GP01002-01:** *The PHQ-9 was developed as a self-assessment but some patients, when they're feeling depressed, they find it difficult to focus on what the questions mean. You have to interpret things, so I often did it for [them].*
>
> **Patient PT03037-03:** *It was different doing it on my own because the first time I did it with [researcher] and I guess doing on my own ... I would say – it's really hard because I know when you answer things, you're supposed to just kind of like tick whatever one you feel like initially or – like – just – whichever one you thought – not put too much thought into it, but when you're on your own, it's harder to do that. I spent a bit longer – like – thinking about each point and being – like – and also focusing on the last two weeks rather than how I'm feeling at that moment, and I think it's easier when you're – when someone else is doing it with you.*

Some GPs also highlighted potential populations in whom accessibility may be reduced, such as older adults who may be less 'tech-savvy', people with sight impairments or patients without fluent English:

> **Practitioner GP02013-01:** *We should try [the PHQ-9] in different languages as well ... If they can't read or they can read but they can't understand English very well – it's like merely ticking the boxes. We don't know whether that's accurate.*

Some GPs suggested that patients should complete the questionnaire in the waiting room before their appointment, and that staff in reception could support those who needed help filling it in. However, issues around confidentiality needed to be considered. Relying on patients to independently complete the PHQ-9 at home was not considered 'bulletproof' as some patients may struggle to complete it as asked:

> **Practitioner GP01056-01:** *I would probably be more inclined to give someone it to take away or to fill out in the waiting room and hand in rather than doing it there and then because, you know, it's time-consuming.*
>
> **Practitioner GP02013-01:** *The receptionist used to be well trained to just hand over the PHQ 9 while they were waiting to be called ... We've got a text messaging system on our EMIS, so if we have a link to the PHQ-9, we can just text them the PHQ-9 link and they can fill it out. Or if it comes in the form of an app, let's say a PHQ-9 app, that can be sent through so we can ask them to do the score once a month and they can send the whole stuff to, you know, text it to the surgery or e-mail it to the surgery and we could just – it gets automatically downloaded on the patient notes.*
>
> **Practitioner GP01026-01:** *It's fine if someone's going to your house and doing the PHQ-9 with you or in the surgery there and then ... Someone who's depressed and can't motivate themselves to get out of bed ... They may not get that motivation to do the form.*

Impact of using the Patient Health Questionnaire-9 on the consultation

Driver of discussion

The PHQ-9 could act as a guide for practitioners to discuss patients' difficulties, needs and care. It facilitated productive conversations as both parties had more information on the problem as assessed using the PHQ-9. It also removed perceived pressure on some patients to formulate what they wished to discuss in the consultation:

> **Patient PT01002-04:** *It sometimes is really difficult to explain how you're feeling. Being able to put it down on paper like that both yourself and your GP can see how you have been feeling and what areas are a bit worse than others. I think it's very helpful.*

It could also help those who found it hard to talk about their depression for various reasons, including stigma, or found it hard to articulate their difficulties and experiences and felt that the GP practice was not a legitimate place to talk about mental health:

> **Patient PT03059-01:** *So it has been helpful in that way to – be able to quiz him from a medical perspective about certain things, to try to understand why – some of the stuff might be going on where, in the past, I would never ever have gone to the doctor really to talk about depression because you feel like you're wasting their time, where there are other places like talking therapies that you can actually talk to a therapist about it.*

As each item highlighted a specific symptom, some GPs felt that the tool helped them clarify with patients each of their difficulties and its individual impact. Many GPs found the final item, which asks patients about whether they have thought about self-harm and suicide, a helpful reminder to conduct necessary risk assessments, as this may not happen routinely in a less structured consultation. This worked as a safety net for the patient, it could be helpful for onward referral to mental health services, and it could also ensure that GPs covered key risks to minimise legal and medical consequences:

> **Practitioner GP01002-01:** *The good thing is the safety aspect. You are forced to ask the question about the self-harm ... I think a big plus of this tool, you can't just do the*

> *score without asking the question and medically and legally, of course it's*
> *indefensible if things happen … In a consultation where you have other*
> *things to discuss with them, it can be easy to forget.*

Patients sometimes reported feeling that using the PHQ-9 was beneficial in bringing structure to the consultation and as a way to compare how they were doing from one consultation to the next:

Interviewer:	*in terms of the conversation, what were you mainly discussing in that [second] appointment?*
Patient PT03037-03:	*I think it was the second PHQ-9; yes, because we talked how about they were the same score. Then she wanted to know other stuff about what had been happening over the last few months and before that and I know that we talked about things that I hadn't talked about in the first one […] It was mainly the PHQ-9 score and seeing – like – which – because she had two of them in front of her so she could see which questions had changed and she kind of just went through which ones had gone up and which ones had gone down and stuff like that.*

The patient–GP relationship

The PHQ-9 could facilitate GPs' engagement with and support for patients through validation, praise, monitoring and evidence of improvement:

Practitioner GP02013-01:	*If the PHQ 9 score was decreasing, then I would congratulate the patient and say – you're not feeling that to an extent but look at the number that you scored some time ago, but they've [come] with this new number, which could be [13] and it would be 24 in the past. I say, that's great progress.*

Many GPs felt that the PHQ-9 was appropriate only for use with patients with whom they had an established relationship. If stepping in for a colleague with a patient they had never seen, they might be reluctant to use the PHQ-9, as no comparisons could be made with that person's previous level of depression.

Practitioner GP01002-01:	*If I see the patient only once, in between the doctor who normally sees them, I would not feel inclined to use this tool, because what am I going to compare it against, and what am I going to do with this information?*

Involvement in the trial encouraged practices to offer patients follow-up consultations with the same GP, which patients found could make it easier to develop a trusting relationship and open up about their depression. Continuity of care was voiced as paramount by patients, who felt that it made a huge difference as they could build rapport with their doctor, feel comfortable, and not need to start again from scratch and repeatedly present their problem or concern to another practitioner:

Patient PT03059-01:	*It was good that I could then go and see my doctor quickly after [completing the first PHQ-9], because then it was nice to be able to talk it through with him. He was able to explain a few things in more depth and talk to me about what I wanted to focus on and try and work on and things. […] I mean I felt more – I guess – safe in a way because I knew then that I'd be seeing him in another four weeks and there is something reassuring about that, knowing that you've got something booked in, especially when – now I've been trying to book an appointment just for a general appointment – oh, we don't have anything, we don't have anything, we don't have anything; that's been going on for three months now and I'm just like – oh my God. It's so ridiculous, like seriously – I*

>   *don't understand. So yes, in that sense I feel like I've got a bit of a safety net in*
>   *being able to see him once a month.*
> **Patient PT01002-04:**   *It's someone who you're already comfortable with. You've already opened up*
>   *and [the GP] knows what's going on so you don't have to talk about some of*
>   *the more difficult things again. It's just a carry on the conversation from the last*
>   *appointment sort of thing. I definitely preferred having the same doctor.*
> **Patient PT03044-04:**   *It was good knowing that there's somebody there to monitor that progress. I*
>   *think that's reassuring.*

Person-centred care

Patients very much valued having trust in the GP and feeling that the GP cared, listened to them and was interested in them, and being viewed holistically as a person in context rather than reduced to simply a person with depression. Patients described feeling that the GP cared when they seemed excited or disappointed about the impact of treatments. Patients also valued flexibility, including discussions around what they wanted, scheduling follow-ups depending on their needs and choice, and flexible treatment plans:

> **Patient PT02139-04:**   *I know it's silly because it's not, really, it's just procedure, I know that, but it's*
>   *just the way that they called and had time for you [...] we're trying to ring the*
>   *GP, we're almost impossible to speak with them. They were always busy, and*
>   *the fact that they'd taken the time out to speak with you. I think he'd rung and*
>   *spoke with me for half-an-hour, which is a lot, when you consider how busy*
>   *they are. So it was nice to have that, and it made you feel like you did matter. I*
>   *know that's daft. It is daft, it was just a phone call, but it's amazing how that can*
>   *change your mood.*
> **Patient PT03036-04:**   *She's really great, asking questions; it felt like she really made the time. She*
>   *asked me certain questions about – like – my day-to-day and my – well first*
>   *just how am I feeling, am I feeling any better than the last appointment or not?*
>   *How do I feel the meds are doing for me and just – am I – I don't know, just like*
>   *day-to-day stuff.*

The PHQ-9 helped GPs and patients to understand individual symptoms and the impact these had on different patients:

> **Practitioner GP01002-01:**   *You actually ask about individual symptoms, which maybe are not widely*
>   *recognised as symptoms of depression ... I think it's a more holistic*
>   *assessment than you would do if you're left to your own devices and you*
>   *ask just three random questions, which might not actually show you the*
>   *whole extent of how things affect you.*

Having their complex and individual experiences understood made patients feel listened to and appeared to instil a sense of agency as they were enabled to play an active part in the decision-making about their symptoms and appropriate care:

> **Patient PT01094-04:**   *I felt quite positive that he had listened to me and my views on it. Whether he*
>   *did or not, I don't know, but I felt like he did.*

However, some GPs thought that the PHQ-9 training did not necessarily reflect the idiosyncratic approach that they should take in clinical practice. They said that the PHQ-9 should be a stepping stone to deeper conversations about patients' problems, which the training did not fully elucidate:

> **Practitioner GP02024-01:**   *The training was a bit too black and white for my taste ... Normally the*
>   *answer was, talk to the patient and listen to them a bit more and see what*

52

> *they wanted, and that was never an option. It was always you had to do this, if antidepressants were appropriate, you'd offer them, and they may or may not take them … I wouldn't be pushing any particular course of action so that training wasn't that realistic.*

Some patients reported feeling that they had more of a say in their own treatment and support and feeling empowered, which increased their hope as a result of more collaborative decision-making about treatment:

> **Patient PT01096-04:** *So the score I got, if I remember rightly, he said to me that he probably would recommend tablets, but he said if I wanted to do, on the side, therapy, then I could, but I decided to see how the tablets would go. He said normally he would recommend the tablets, but he then went through all the side effects with me, and he let me make my own decision. He just recommended and went through all the risks and side-effects they can have and things like that.*
>
> **Interviewer:** *What do you think about the patient's reaction when you gave them their feedback on their scores?*
>
> **Practitioner GP01012-01:** *I think they were mostly absolutely fine and accepted that, you know, it was – putting it in – in less of a judgemental way; so it gave them a bit of ownership of it. So, yes, I think they liked it.*

Impact of using the Patient Health Questionnaire-9 on the practitioner care of depression

Evidence to inform treatment and management

Most GPs and patients found the PHQ-9 informative for treatment and management plans. They felt that the PHQ-9 solidified decisions about potential treatments to suggest; for example, lower PHQ-9 scores might indicate lifestyle improvements and medication might be more appropriate for those with higher PHQ-9 scores. The strength of GPs' endorsement of treatment suggestions was also influenced by the PHQ-9 score.

> **Practitioner GP01056-01:** *I perhaps might be more persuasive towards someone who's dark red saying I actually think we do need to start you on treatment … If someone's yellow, I might sit more on the fence and just say – this is an option, how do you feel? Would you prefer to have treatment or not have treatment? … I think it does help you knowing where someone is on the severity score in terms of how you emphasise the different treatment.*

The utility of completing the questionnaire with the doctor was described by some patients in terms of focusing the discussion on the severity of the depression and the corresponding treatments:

> **Patient PT01096-04:** *I remember we both had the form up on there and we were going through it, and he was explaining to me the diagram and things like that, and we were talking about the category that I fell into, and different ways to help it and things like that. So that's, again, when he said about the two different types of treatment, which was tablets or therapy. Yes, so it was used to see where I was at, at that current point in time, basically.*

However, practitioners often felt that a score on the questionnaire was not needed as they already knew the patient and could judge what management/treatment they might need, although having the score could be helpful with patients they did not know:

> **Practitioner GP03044-01:** *I think I'll probably preserve it for patients that actually need a more formal management plan.*
>
> **Interviewer:** *And how would you identify those patients, as compared to those who don't – who won't use the PHQ-9?*
>
> **Practitioner GP03044-01:** *Just because I know the patients quite well and I think the ones who will open up to me and I've got a good rapport and history with them, I can probably judge, you know, what I need to do with them, faster. Maybe somebody like new patients, patients who don't come in [that well], ones I'm not sure about, would be a good place to use that.*

The PHQ-9 score was also described as indicative of whether or not the treatment plan was working. If scores did not change or increased, this made GPs and patients consider stepping up treatment:

> **Patient PT01002-04:** *Looking at the scores and stuff, [the GP] would make that decision on whether he thought some of the dosage for things needed to be increased or stay the same. We were really looking at the score and seeing where were at and how much it's gone down.*
>
> **Practitioner GP03044-01:** *You could actually see what the changes were and in patients where there were no changes, if they were worse, then we were going to implement. It helped me decide treatment, you know, plan … It was a nice way for me to see that the medication is or is not working and then for me to refer onwards if need be or start medication.*

On the other hand, a few GPs reported that the PHQ-9, although helpful to an extent, was not necessary or influential in treatment decisions:

> **Practitioner GP02024-01:** *The treatment options are unlikely to be that different. Maybe you might be encouraging drugs more if they are more severe, but you don't necessarily need the PHQ 9 to do that.*

Very occasionally, practitioners thought that patients might respond to the questionnaire in a particular way to effect a change in their care (e.g. one patient was described as 'lying' about their symptoms in order to be discharged from counselling they were not finding helpful).

Objectivity versus subjectivity

In the presence of uncertainty about patients' difficulties and needs, some GPs believed that the PHQ-9 could provide objective and clarifying quantitative information that could be helpful, whether in going along with their clinical judgement or contrasting with it. Some saw it as helpful because patients may not always be present with symptoms in the same way, and symptoms may not be overtly visible during consultations:

> **Practitioner GP01026-01:** *Does it make any difference to what I'm going to do? Ninety-nine per cent, no. As I say, I've used it literally twice in the last couple of weeks, purely because I was thinking, this patient's coming in and they're depressed, but I'm not really sure they are. [The PHQ-9] did reaffirm already what I had thought in my head, that they were mild. That's the only time I've found it beneficial.*
>
> **Practitioner GP03044-01:** *We stopped using [the PHQ-9] but it's something we should think about using, especially with something as grey as depression, because people have different emotions. So people could be really, really tearful but not have really bad depression and there could be other people who look quite,*

> *you know, which happened with a few patients. They looked quite together with it but actually who were, like severely depressed. So, yes; it's a helpful, objective tool.*

However, several GPs, especially those with more years of experience, preferred to use clinical judgement by itself to identify symptoms and make decisions about patients' care. They did not like the rigidity of the categories and the associated suggested treatments, which they referred to as 'tick-box medicine'. These GPs expressed resistance to the continued use of the PHQ-9. However, some suggested that it could be used as a guide for younger GPs with less experience:

Practitioner GP01026-01: *I think if you're relatively OK at your job, I think you should be able to assess that anyway, without having to do a tick box exercise on a piece of paper; it's probably not my kind of thing. So I would not be keen for them to come back in. As I say, when I was talking to the GP colleagues about the study, there was lots of moans and groans, 'Oh, are we going to have to do PHQ-9 again?' – which gives everybody's tale of what they think about it, basically.*

Practitioner GP01002-01: *Doctors as a rule don't like tick-box medicine, unless they are starting out as new doctors and haven't found their feet yet and are glad about any pointers that they can get hold of to do a proper job. As you're becoming more senior, doctors tend to do less of those things, because they feel they are being patronised and being told what to do and they didn't become a doctor to be told by an administrator they need to use this or that tool … They'd rather do what they think is right rather than somebody else telling them to do so.*

Many among this group of GPs were open to implementing the PHQ-9 if either there were demonstrable data on its effectiveness on patient outcomes or patients stated a preference for its use:

Practitioner GP01026-01: *It'll be interesting to know from the results of the study, whether they have found that it helped [patients] with their symptoms … It'll be interesting to know from the patient's point of view … It may be turned on its head, with all the patients saying, it really is. Then I'm happy to change my mind.*

Organisational barriers to and facilitators of using the Patient Health Questionnaire-9 in practice

Time restraints
Some GPs stated that they did not have as much time as they would like during their 10-minute consultations to discuss depression with patients and often felt unable to spend time administering or discussing the PHQ-9 on top of that:

Practitioner GP01026-01: *They'll be given a 10-minute appointment, which is the usual in GP land. We don't tend to have appointments longer than that which is a bit of a shame when assessing depression, because it clearly takes longer than that, let alone then having to do a PHQ-9. I think most of [the GPs] just found it a complete pain.*

However, because the PHQ-9 scores immediately generate specific treatment suggestions, this could sometimes balance out the additional time it took to use the questionnaire during consultations:

Practitioner GP01056-01: *It perhaps helps formulate that decision more quickly … If someone's coming to their appointment with [the PHQ-9] already done, you've*

> *got that information in front of you. I think it makes your consultation more efficient.*

Patients could also on occasion value shorter, more focused follow-up consultations:

Patient PT01094-04: *(Talking about a follow-up call) It was quite short, which surprised me, I think. When I had spoken to my own GP, I had gone to great lengths to explain about everything that was going on and the family situation. Whereas, when I had the phone call with the second GP, it was much more, 'Right, you've been told. Right, here's the questions. Give me numbers. Give me answers to this' Then he gave me a score at the end. Then he was like, 'Right. So you've been on the medication now for a month. What do you think? What are your thoughts? How do you feel? Yes, I agree with what you're saying. I think that's a good route to go. Let's do that' End of call. So the whole thing was quite short, but it got the job done.*

Technological integration

Practitioners mentioned that the PHQ-9 was being integrated into GP computer systems to streamline its use in e-consultations ahead of contact with the patient, and some valued being able to text it to patients during remote telephone or video consultations using an integrated messaging system on their computer:

Practitioner GP03044-01: *It's quite nice if it was integrated into our clinical system, our IT system; so it would be faster to generate – to tick with the patient there, go through it and then they get the results.*

Practitioner GP02013-01: *We can just text them the PHQ-9 link and they can fill it out. We can ask them to do the score once a month and they can text it or e-mail it to the surgery. It gets automatically downloaded on the patient notes.*

Patient PT03059-01: *It's so hard, especially in London, to get appointments with doctors. Having my GP apps, things where you can fill it in, and it goes directly to your doctor. If they're concerned, they send you a message saying, would you like to come in and have an appointment?*

Patient PT02024-03: *I just showed him the form and he said he already had some of the results, they'd been e-mailed through to him and just showed him the graph and I told him I didn't realise that I was that low.*

Frameworks and guidelines

Several GPs said that they had used the PHQ-9 only when it was a requirement of the GP contract QOF and would go back to using it only if it were a requirement again and GPs were paid to use it:

Practitioner GP02013-01: *When the PHQ-9 was part of QOF, we religiously gave PHQ-9 to all the patients … The thing now is because general practice is so pressured with time, unless there's a monetary incentive to any GP, they wouldn't do it. I've always done the PHQ-9 before, I still am and I will be but – I'm doing less of the PHQ-9 than before because it's not incentivised monetarily.*

Practitioner GP01056-01: *Going back it used to be part of a QOF target [...] So we used to use it all the time as a monitor. It's difficult because I think it does give you a number that can be quite useful to see whether people are getting worse getting better, but equally, I think if you – a lot of that you'd get a feel for anyway with the questions and I think the real challenge for us is we are so time pressured, that I think adding in – into a routine consultation – adding in*

	another assessment tool, I think time is the biggest barrier to using that more frequently.
Practitioner GP03044-01:	*I think I had used it before, because it was part of QOF and so we were using it before then and then we often use it as sort of an abbreviated version when we're doing a referral for psychology services as well, it's on the referral form as well. But then I think when it didn't become – a requirement for QOF anymore, we stopped using it. But really using it for the study actually – I think is something I'll continue to use.*

Many GPs felt that if the PHQ-9 were recommended in NICE guidelines, they would feel more encouraged to implement it in practice. To facilitate this, they wanted a clear evidence base that it was effective and should be used routinely in primary care. Some commented on the need for clearer guidance for GPs on what to do depending on patient scores:

Practitioner GP01002-01:	*If it was part of NICE guidelines, that would be different … We have to adhere to them and justify why we didn't … I think you find most people won't have a better system to assess patients, certainly not an evidence-based system. That probably would be the single best way to get it into the general GP population and use it in practice.*

Chapter 6 Discussion and conclusions

Summary of findings

Clinical outcomes

We found no significant difference between the intervention and control arms in the primary outcome, depressive symptoms on the BDI-II questionnaire, at 12-week follow-up. There was also no significant difference in depression symptoms on the BDI-II at 26-week follow-up, although it was not possible to rule out a benefit in terms of depression outcome at 26 weeks as the 95% CIs for the difference did include a possible clinically significant difference favouring the intervention. We found some evidence of benefit in a categorical analysis of remission of depression to a BDI-II score of < 13 at 26 weeks, but this result should be treated with caution as the analysis was post hoc.

No significant differences between the intervention and control arms were found in the secondary outcomes of social functioning on the WSAS, or satisfaction with medical care received on all the MISS scales, although the differences found were in the direction of favouring the intervention.

More patients in the intervention arm than in the control arm were given prescriptions for antidepressants over the 26 weeks' follow-up, but again this difference was not statistically significant. There was also no difference found in the use of specialist mental health services, with around one in three patients in both arms having at least one contact with a psychologist, counsellor, community mental health nurse, psychiatrist, social worker or other therapist.

Health economic outcomes

Despite the lack of significant differences in mean depression scores, EQ-5D-5L quality-of-life scores were significantly higher at 26 weeks. The difference in QALYs gained over 26 weeks was very small, however, while costs were lower in the intervention arm, but not significantly. Cost-effectiveness and cost–utility analyses therefore suggested that the intervention was dominant over usual care, but with considerable uncertainty around the point estimates. The CEAC showed that the probability of the intervention being cost-effective compared with usual care, at the lower and higher thresholds for societal willingness to pay adopted by NICE, of £20,000 and £30,000 per QALY, was 77% and 72%, respectively.

Qualitative interview findings

In the qualitative interviews, practitioners and patients described the PHQ-9's various benefits, including providing information on the range of symptoms and severity categories of depression; highlighting particular symptoms, including suicidal thoughts; identifying changes in mood over time; and informing treatment plans. However, a number of practitioners stated that their own clinical judgement was more important in making management decisions.

Some patients said that the PHQ-9 oversimplified their complex experiences of depression, and some practitioners did not like the rigidity of the severity categories and the associated suggested treatments, which they referred to as 'tick-box medicine'. Several practitioners expressed resistance to using the PHQ-9 for these reasons, although some suggested that it could be used to guide less experienced practitioners.

Barriers to using the questionnaire in routine practice included the extra time taken in the consultation, which could be reduced if administering the PHQ-9 were made automatic through technological integration in practice–patient communication and computerised record systems. Practitioners wanted an evidence base of its effectiveness, clear guidance on what to do depending on patient scores and remuneration for the extra time taken.

Interpretation of the study findings in the light of previous research

The findings add to the evidence from the previous trials in the USA and Europe, cited in the introduction, that showed no consistent benefit in outcome for people with depression of using PROMs in primary care follow-up monitoring. Only one previous primary care trial found a reduction in depressive symptoms but no changes in the process of care that might have explained the benefit found.[19,20] Two other studies found changes in the process of care but no improvement in outcome.[17,22] The PROMDEP trial has improved the quality of the evidence base for monitoring depression in primary care, compared with previous studies, which were generally rated as being at considerable risk of bias.[13]

The lack of significant differences in depression scores begs the question why quality-of-life scores were significantly better at 26 weeks. The improved quality-of-life scores were apparently because a greater proportion of patients in the intervention arm reported no problems with anxiety/depression. We did not measure anxiety symptoms specifically at follow-up, but the qualitative interviews suggested that intervention arm patients valued seeing their scores change over time, and so the mechanism for this finding may be that they were reassured that their depression was improving when they were fed back their follow-up PHQ-9 scores and, as a result, felt less anxious.

During this trial, another primary-care-based study of PROMs for depression was published. The PReDicT randomised trial tested the use of a predictive algorithm to guide the antidepressant treatment of patients seen largely in primary care in five European countries, compared with unguided treatment, recruiting 913 patients and following them up for 48 weeks.[54] The algorithm was based on an affective processing task (rating emotions from facial expressions) and changes in item scores on the Quick Inventory of Depressive Symptoms (QIDS-SR-16) questionnaire[55] over 2 weeks from the baseline assessment. The prediction that a patient was not responding prompted a change of medication in 65% of cases, mostly an increase in dose rather than a switch to a different antidepressant or augmentation with a second treatment. The primary outcome of depressive symptoms at week 8 did not differ significantly between the arms, although significant benefits were found in the intervention arm of reduced anxiety at week 8 and improved functional outcome at week 24. The authors concluded that personalising antidepressant treatment based on early changes in symptoms may improve outcomes in depressed patients, although it should be noted that the reported positive findings were for only 2 out of 10 secondary outcomes measured. It is possible that the limited extent of the changes in treatment resulting from the predictive algorithm limited the impact of the intervention. Health-related quality of life was not apparently measured in the PReDicT trial,[54] and the cost-effectiveness of the intervention has not yet been reported.

More consistent evidence of benefit from the routine outcome monitoring of depression with PROMs has been found in psychological therapy settings, where this has been shown to improve depression outcomes for 'not on track' clients[14] and save resources by shortening therapy for 'on track' clients.[15] This may be in part because in psychological therapy services all clients present with mental health problems and so the administration of PROMs to all is justified and can be made routine and embedded in the service, unlike in primary care, where only a proportion of patients presenting have mental health problems. In the UK the IAPT service delivery model requires the routine collection of clinical, social and employment outcomes as part of a national outcome monitoring system, and the performance of psychological well-being practitioners is measured through these outcomes. The IAPT national outcomes monitoring system, including session-by-session symptom measures, is built into the psychological well-being practitioner training curriculum.[56] The implications of PROM results for therapy are also discussed with clinical supervisors between sessions with clients. As a result, practitioners' experience and expertise in outcome monitoring is much greater in the IAPT programme, as it tends to be in other psychological therapy settings.

The PROMs used in monitoring progress in psychological therapy settings are also more extensive than the PHQ-9 alone. IAPT services use the PHQ-9, GAD-7 and WSAS together at every assessment and

therapy session, that is, a total of 21 items. Other psychological therapy and mental health services that have reported benefits from outcome monitoring have used the 45-item Outcome Questionnaire (OQ-45), which has an algorithm to interpret the implications of clients' scores for continuing or changing therapies.[12] Outcome monitoring in IAPT services has been shown to be improved when PROM scores feed into an algorithm that interprets the results for the therapists and produces clear recommendations for continuing, changing or terminating treatments.[14] The benefits found in the PReDicT primary care trial may similarly have been due to the more extensive PROMs and the algorithm used to deliver specific recommendations for making changes to antidepressant treatment, which was effective in two-thirds of cases (mainly in increasing antidepressant doses).[54]

Finally, the difference may be because psychological therapists have greater access in their services to a range of treatments with good evidence of effectiveness. Primary care treatments are more limited, usually comprising an offer of antidepressant treatment to the majority of patients and referral to exercise classes, psychological therapies or mental health services to a minority. Antidepressants are less effective than psychological therapies for treating the level of depression commonly found in primary care,[7] which could account for the difference found in the effectiveness of outcome monitoring in the psychological therapy setting in terms of outcome of depression.

The qualitative findings are generally in line with those of previous studies of symptom questionnaires for monitoring depression in primary care, which suggest that patients value these but practitioners have a more mixed view, with some reluctant to use such questionnaires unless they are remunerated for their time doing so.[24,30] The finding that some practitioners perceived the PHQ-9 as a 'tick-box exercise' that took up extra time in consultations and was not helpful in deciding on treatment choices echoes the findings of a GP focus group study carried out during the period when using symptom measures was incentivised in the QOF.[57] Practitioners reported that the PHQ-9 generated work, forcing them to adapt consultations so that they could tick boxes, while reducing the time available to offer appropriate care when questions about symptoms triggered patient distress.[57] Previous qualitative work has also suggested that GPs regard symptom severity questionnaires as intrusive in the consultation and a threat to their identities as professionals with expertise, which they see as integral to the process of diagnosis.[24,58] Some have doubted the validity of the PHQ-9 as an accurate measure of a person's severity of depression.[24]

The limitations of the PHQ-9 as a PROM were acknowledged with the participating GPs during the trial training. We pointed out that, although previous research has shown that the PHQ-9 is valid against longer, interview-based assessments of depression,[50] this applies only at the group or mean level; individual patients' scores vary widely around the mean, and there are always false positives and false negatives. Consequently, a mismatch occurs quite commonly between patients' PHQ-9 scores and their own global rating of how they feel.[59] Third, the threshold score of ≥ 10 out of 27 for considering drug or psychological treatment may be too low, as more people are labelled as depressed at that threshold on the PHQ-9 than on the Hospital Anxiety and Depression Scale questionnaire.[23] Lastly, we emphasised that a symptom frequency score in itself cannot suffice to indicate that treatment is needed, and that the effects of patients' symptoms on their daily functioning at home, at work and in their relationships are key.[32]

Notwithstanding these limitations with respect to its validity, patients may still value the PHQ-9's use in follow-up monitoring, as it indicates to them what is happening to their individual scores over time, which can be reassuring if those scores are improving.

The extra time taken to use the PHQ-9 in consultations could be significantly reduced by automating the remote administration of the questionnaire beforehand through sending it by text message or e-mail or by including its administration in online consultations (e-consults). Some practice computer systems already incorporate the PHQ-9 (and the GAD-7 for anxiety) automatically into e-consult questions for patients who indicate that they may have mental health problems. The developers of the PHQ-9

envisaged the instrument saving clinicians time enquiring about the presence and severity of each of the nine DSM-IV symptoms in assessing outcomes, especially if it were administered ahead of the consultation over the telephone or even by 'interactive voice recording'.[18] Telephone administration has been shown to be reliable[60] and could be done by administrative staff whose time is less expensive than that of the practitioners. Completion of the PHQ-9 before the consultation may in any case be more valid than administration by the practitioner, as it has been shown that a practitioner can influence patients' responses to the questionnaire by, for example, not giving them the full range of options for answering each item.[61]

Strengths and limitations

A major strength of the study is that its design was informed by a feasibility trial, which led to the choice of the cluster-randomised design; this was necessary to avoid contamination between the arms in the application of the intervention and to optimise adherence to the study procedures within practices. However, a limitation of the cluster-randomised design was the risk that participating practitioners might exhibit selection bias when deciding whether or not to approach individual patients attending consultations for depression about taking part in the study. That was why we identified potential participants through frequent practice records searches in addition to opportunistic recruitment by practitioners during their consultations.

While the majority of patients (69%) were recruited through the practice record searches rather than opportunistically, opportunistic recruitment was more frequent in the intervention arm (110 patients, compared with only 44 in the control arm). The cluster-randomised design meant that practitioners and patients knew which arm they would be allocated to before they were asked to consent to take part, and there may have been lower motivation to take part among the patients, due to possible disappointment at not being in the intervention arm and, therefore, being offered only usual care.

This may explain why we experienced a slower recruitment rate in the control practices, resulting in differential recruitment of patients to the two arms. The final ratio of intervention to control arm patients was 1.3 to 1, having been 2 to 1 early in the study. The ratio was brought down by persistent efforts on the part of the research team to encourage longer patient recruitment in control arm practices through both opportunistic recruitment in consultations and practice records searches and mail-outs of invitations, but we did not manage to eliminate the disparity entirely.

Selection bias may explain why patients in the intervention arm had higher depression and anxiety scores and lower quality-of-life scores at baseline. These differences would have made it harder for the intervention to have made a difference, although the two arms of the study were well balanced in terms of patient demographic factors, and all of the analyses were adjusted for these baseline differences.

The feasibility study also informed the choice of the PHQ-9 as the practitioners' preferred PROM to be tested, as it was familiar, but its findings nevertheless identified a need for training practitioners in its use. Participating practitioners were, therefore, trained in both the use of the questionnaire and the choice of treatments related to the severity score, while taking into account other contextual factors. Their training was also assessed by asking them sets of questions on the two topics. Their responses to the questions scored generally highly, sometimes only after a second attempt at answering, and occasionally after a third attempt. The amount of training was limited to 2 hours, but this was considered a feasible amount to offer at scale should the intervention prove successful and need to be disseminated throughout NHS primary care practices.

The recruitment of practices and patients was very challenging due to practices' other commitments, particularly during the COVID-19 pandemic, when they had extra work isolating patients, donning and doffing personal protective equipment, switching over to remote consultations, staffing COVID-19

'hot' centres, and mounting the national vaccination programmes. However, although recruitment took longer than envisaged, we did manage to recruit significantly more practices than our original target by persistent approaches through 13 CRNs in England and Wales.

We did not quite achieve the revised sample size target of 554 patients, falling short by 25, but the follow-up rate of 85.6% was better than the 80% predicted and so we did gather primary outcome data on a sufficient number of participants to answer the main research question with precision. That was reflected in the relatively tight 95% CIs around the estimates of depressive symptoms at baseline and 12-week follow-up of around only 1–2 points on the BDI-II, which has a maximum score of 63 points. It is not likely, therefore, that a possible benefit at 12 weeks was missed because of a lack of power (a type II error) and so the results for the primary outcome may be regarded as robustly negative. However, it was not possible to rule out a clinically significant benefit on the BDI-II score at 26 weeks given that the upper limit of the 95% CI for the difference in favour of the intervention included the minimal important clinical difference of 3 points.

We measured a range of relevant secondary outcomes including work and social functioning, satisfaction with care, and adverse events, as well as health-related quality of life and costs for the health economics evaluation. We had safeguarding procedures in place for patients expressing thoughts of self-harm or suicide, which did not differ in relative frequency between the trial arms. We measured changes in the process of care, including antidepressant use, in the two arms, and contacts with mental health services, and we included a qualitative process analysis to inform the findings, illuminating the potential processes involved from patients' and practitioners' viewpoints.

It was a pragmatic trial, with patients' care remaining with their GP or NP as it would outside the trial. It was not possible to blind practitioners, patients or researchers to trial arm allocation given the pragmatic nature of the trial and the cluster design, but we used self-report outcome measures to avoid possible observer bias, and all of the statistical analyses were carried out blind to allocation to arm.

Intervention delivery was challenging and not entirely carried out as it would be envisaged working in routine practice. The GP/NP could not give the PHQ-9 questionnaire at the initial consultations with patients presenting with new episodes of depression, because, ethically, patients had to be given information about the study and sufficient time (at least 24 hours) to consider taking part before being asked to sign the consent form. To avoid asking the GPs/NPs to bring the patient back quickly to carry out the first PHQ-9, which would generate significant extra work, the researchers administered the first PHQ-9s at baseline assessments on behalf of the GPs/NPs. We endeavoured to carry out the baseline assessments and administer the first PHQ-9s as soon as possible after patients first presented with symptoms, but this was sometimes delayed by 2–3 weeks. In the meantime, treatment with antidepressants had already been started by the GP/NP in around half of the intervention arm patients. The first PHQ-9 score could therefore not be taken into account when first choosing possible treatments for those patients.

Although we emphasised to the practitioners that best practice is not to start treatment immediately when patients present with new symptoms of depression, they were permitted to institute treatment at the first consultation if they deemed it necessary. A similar proportion of patients in the control arm were also already taking antidepressants by the time of the baseline assessment, so it appears that being in the PHQ-9 arm did not affect the proportion of patients starting treatment at the consultation when they first presented with a new episode of depression.

In the intervention arm, 73% of patients had at least one PHQ-9 result recorded in their medical records, which indicates that the GPs/NPs did not routinely record patients' scores, as we know that 100% of intervention arm patients had at least the first PHQ-9 administered by a researcher at baseline. We cannot be sure that all intervention arm patients had a second, follow-up PHQ-9 administered in the practice, and we do not know how often the participating practitioners checked the first or second

PHQ-9 scores and took them into account when deciding on treatment, although these problems were not apparent from the qualitative patient interviews. Effectively we were testing instituting a policy of outcome monitoring using the PHQ-9, which we knew would not necessarily always be carried out per protocol, as would likely be the case to an even greater extent in routine practice outside the trial situation.

A smaller number (17%) of patients in the control arm also had at least one PHQ-9 recorded, despite the fact control arm practitioners were asked not to use it. These may have been administered outside the practices in IAPT services, or by temporary practitioners in the practices. However, this was a relatively low level of use in the control arm, so there was good differentiation between the arms, and all analyses were conducted on an intention-to-treat basis, including all patients in the arm to which their practice had been randomised, regardless of whether PHQ-9 use was recorded or not.

Generalisability

There were relatively few exclusion criteria that would tend to increase the heterogeneity of the patient sample and the generalisability of the findings. All adults presenting with new episodes of depression were eligible, with no upper age limit and no exclusions for co-existing physical health problems, which are very commonly found among people with depression. In addition to a significant risk of suicide, the only exclusion criterion was having one or more of three comorbid mental health problems, specifically dementia, psychosis or substance misuse, which was necessary to avoid complicating the diagnosis and affecting the range of treatments that could be offered by the participating practitioners.

The medical records computer codes used to identify potential patients included symptom codes (such as 'low mood') as well as specific diagnoses (such as 'depressive disorder') to avoid missing patients not given a specific diagnosis at presentation. This was more inclusive and would have tended to reduce potential selection bias arising from whether or not practitioners chose to attach specific diagnostic labels to patients in a non-systematic way.

However, there was a relatively large drop-off from the 11,468 patients approached to take part down to the 529 who eventually consented and were enrolled in the study. This was mostly due to patients not returning reply slips to the study team to find out more about what enrolment would mean, with only 11% of those approached in the intervention arm and 8% in the control arm returning them. This difference in the rate of return was one factor leading to differential recruitment in the two arms. After the exclusion of patients declining to participate, proving ineligible at screening or being uncontactable, only 6% of those approached in the intervention arm and 4% in the control arm were recruited.

We can compare our trial sample with a population of 1658 people presenting with depression and assessed with the PHQ-9 in routine clinical practice, identified among 38 general practices in an observational study of the use of the PHQ-9 in 2006–7, during the period when its use with all eligible patients was incentivised through the QOF.[23] In terms of gender, the proportion of male patients in this trial, at 38%, is similar to the proportion of 36% found among with those with depression diagnosed and assessed using the PHQ-9 in the observational study. In terms of age, the proportion aged ≥ 65 years in this study, at 13.5%, is reasonably similar to the proportion of 15.9% found in the observational study.

In terms of history of depression, however, it is noteworthy that the proportion of people in this study with no previous episodes, at 25.2%, is much lower than the corresponding proportion of 62.4% found in the observational study of using the PHQ-9 with all patients presenting with depression.[23] The trial patients were much more likely to have recurrent depression, including 26.7% with one previous episode and 48.1% with two or more previous episodes. This means that the trial patients were more experienced in receiving treatment for depression and could therefore have been more resistant than treatment-naive patients would be to changes in drug treatment, or new referrals for psychological

therapy or to mental health services. This might help to explain why no statistically significant difference was found between the two arms of the trial in the levels of antidepressant drug treatment and secondary care mental health service contacts.

Implications for health care

The popularity with patients, improved quality of life and reasonably high probability of the PHQ-9 being cost-effective found in the PROMDEP study support current recommendations to use PROMs for monitoring depression in primary care settings in the UK, the USA and Europe. Organisations recommending the use of PROMs include NICE,[62] the US Federal Health Resources and Services Administration,[63] the US Departments of Veterans Affairs and of Defense,[64] Kaiser Permanente Health Maintenance Organisation[65] and the Nederlands Huisartsen Genootschap (Dutch Society of GPs).[66]

On the other hand, the findings of no significant differences in the management or outcome of depression suggest that the use of PROMs should be discretionary, with those patients who value seeing changes in their symptom scores over time, rather than being made mandatory with every patient, as they are in the NHS Talking Therapies programme.[56]

Recommendations for research

Future research in primary care should look at using PROMs for depression monitoring that cover more symptoms, including anxiety, and measures of functioning; that can be completed remotely via the internet or telephone by patients before and between consultations; that benefit from automated computerised analysis and feedback of the results to both practitioners and patients; and that produce algorithms delivering specific recommendations for changes in treatment arising from the PROM results.

Patient-reported outcome measures should include anxiety symptoms, as reduced anxiety might have been the main contributor to improved quality of life in the PROMDEP trial, and the PReDicT randomised trial showed some benefit from PROMs in terms of reduced anxiety.[54] Depression in primary care is frequently mixed with anxiety[1] and the recent NIHR PANDA trial of treating depression in primary care with the antidepressant sertraline found greater and earlier benefit in terms of reducing anxiety symptoms than depression.[67] Anxiety is an important outcome which has been relatively neglected in trials of monitoring depression.[13,68]

Patient populations should include people with multiple physical conditions who have a high prevalence of depression and greater overall morbidity, mortality and healthcare costs. Studies should also be conducted with adolescent patients and older patients who have been neglected in research on PROMs so far.[13,68]

Trials should measure depressive and anxiety symptoms including total symptoms, proportions of participants with categorical remission and categorical improvement in depression, quality of life, social functioning, adverse effects including drug side effects, satisfaction with care, use of services and costs for cost-effectiveness estimation. Studies should also follow up patients for longer than 6 months.

Trials should consider including subgroup comparisons and qualitative interviews to address whether PROMs provide particular benefits for certain patients, for example, those who do not readily report symptoms or articulate well how they have been progressing when asked an open question about how they are feeling; and patients who are treatment-naïve and less informed about what treatments might be indicated.

Equality, diversity and inclusion

Participant representation

We took active steps to optimise participation of all people presenting to their general practices with new episodes of depression through getting practices to search their medical records for new episodes, as well as asking practitioners to recruit people during consultations. As a result, the participant population was particularly inclusive in terms of gender and ethnicity and reasonably inclusive in terms of age.

Gender

As indicated above, the proportions of women and men included in the trial were similar to those among all patients presenting to general practices and assessed with the PHQ-9 in observational research of routine practice,[23] including 62% women and 38% men. This reflects a ratio of around 2 : 1 women to men usually found in studies of people with depression.[1] It may represent a differential susceptibility to depression related to gender, or a differential willingness to seek help, or both.

Ethnicity

Table 17 shows the percentages by self-declared ethnic group of the study population compared with the population of England and Wales in the 2011 census.[69] Overall there was a reasonably good representation of people from ethnic minority groups compared with national figures, which probably resulted from our recruitment efforts conducted through 13 CRNs in England and Wales and including general practices ranging from inner city and suburban to rural practice areas, across the spectrum of social deprivation.

Age

In terms of age, the proportion of people in our sample aged ≥ 65 years, at 13.5%, is lower than the proportion in England and Wales of 16.5% found in the 2011 census.[69] This may reflect the fact that follow-up contact with those returning reply slips was by telephone, text and e-mail as well as post where possible, which might have been less likely to reach older patients less used to using mobile phones and the internet.

Research team representation

Study team members ranged in gender, culture and ethnicity, including black and Asian ethnicities which are generally under-represented in healthcare research, as well as white British. Our research team was made up of a diverse group of academic and clinical researchers at different career stages, including

TABLE 17 Ethnic groups of trial population compared with the 2011 national census

Ethnicity	Study population, (%)	England and Wales, (%)
White	84.9	86.0
Black Caribbean	0.8	1.1
Black African	1.3	1.8
Black other	0.4	0.5
Indian	3.2	2.5
Pakistani	1.9	2.0
Bangladeshi	0.2	0.8
Chinese	1.3	0.7
Other Asian group	1.5	1.5
Other ethnic group	4.6	3.1

research assistants, research fellows and professors. Development opportunities were provided for more junior members of the team, some of whom left during the study to enrol for PhDs and clinical doctorates in psychology.

Patient and public involvement

We recruited two service users, MB and BP, to join the study team, very early on, at the point of designing the study and applying for funding. MB and BP had helped us previously with the PROMDEP feasibility study, and their involvement at the design stage ensured the relevance of our study aims to patients' perspectives on the problem to be addressed.

BP was convener of a self-help group of people with depression run by Southampton Depression Alliance (DA), which has since been merged with Mind. Through the DA group, BP helped TK to ask a group of six people with experience of depression and depression treatments to look at a range of depression PROMs, which influenced the choice of the PHQ-9 for the PROMDEP trial.

BP and MB were both very active members of the study group, attending study team meetings and commenting on relevant documents through e-mail throughout the 4 years of the study. One or other of them attended all but two study team meetings, so that we almost always had their support and input. They were paid £18.75 per hour for their time, in line with INVOLVE recommendations in 2018, which included time spent at meetings and commenting on documents, plus travel and any other out-of-pocket expenses.

Our PPI colleagues helped ensure that easily understood patient information was provided, and participation was voluntary, through reading and commenting on participant information sheets and consent forms. They also commented on the feedback given to patients in the intervention arm on the meaning of their PHQ-9 scores and possible treatments related to the level of severity of their depression. They also reviewed the semistructured interview guides used for the patient and practitioner interviews for the qualitative process analysis, and provided PPI feedback on the meaning for them of the emerging qualitative findings. BP wrote in support of the 12-month extension request. We gave our PPI colleagues regular feedback on our interactions with them and asked for theirs, both of which were very positive.

They also helped revise our plain English summary of the trial results. BP ran it through the First Word online readability test at https://thefirstword.co.uk/readabilitytest/ (accessed November 2023) which, after revision, gave it a Flesch–Kincaid reading ease score in the 'plain English' category.

BP has also agreed to help publicise the results of the trial locally through his Depression Alliance self-help group and nationally through Mind.

Conclusions

We found no benefit from using the PHQ-9 in relation to the primary outcome of depression on the BDI-II at 12-week follow-up. There were also no significant differences found between the arms in the secondary outcomes of BDI-II scores at 26 weeks, work and social functioning, patient satisfaction, medication use or contacts with mental health services, although all the differences found in these measures were in the direction of favouring the intervention arm.

However, we did find a significant benefit in terms of improved quality of life at 26 weeks, for lower mean service costs. We also found evidence of benefit in a categorical analysis comparing rates of remission of depression at 26 weeks, although this result should be treated with caution as it was a post

hoc analysis. CEACs showed the probability of the intervention being cost-effective, at the lower and higher thresholds adopted by NICE of £20,000 and £30,000 per QALY, was 77% and 72%, respectively.

We found that patients valued using the PHQ-9 to identify changes in their scores. The mechanism by which feedback of scores might improve patients' quality of life, despite not changing the management of their depression, might be through increasing their awareness of improvement in their symptoms over time, supporting personal reflection on their progress to recovery.

The findings support NICE and other guideline organisations in recommending routine outcome monitoring using PROMs in primary care. Further research should be conducted in primary care evaluating longer PROMs which include anxiety symptoms, administered remotely between primary care consultations, with automated algorithm-driven interpretation delivering clear recommendations for possible changes in treatment.

Dissemination

The feasibility trial was published in the journal *BMJ Open* in 2017.[27] The study protocol was published in *Trials* in 2020.[70] A review article based on the background to the trial, plus an updated literature search, was published in the *BMJ* in 2020.[68]

The findings will be further disseminated to patients, clinicians, academics and the media as follows:

1. Patients. With assistance from our PPI representatives, we will disseminate study reports in plain English to people with depression through the Depression Alliance, Mind, and other patient groups.
2. Clinicians. We will publish the findings in short articles in GP trade journals (Pulse, Doctor, and the Practitioner) as well as peer-reviewed academic journals such as the *Lancet*, *British Medical Journal*, *Annals of Family Medicine*, *British Journal of General Practice* and *Family Practice*. In addition, we will submit abstracts to the Royal College of General Practitioners' annual conference, aiming to publicise the findings through oral or poster presentations.
3. Academics. In addition to publishing the findings in academic journal papers as above, we will submit abstracts to the Society for Academic Primary Care, and North American Primary Care Research Group, aiming to publicise the findings through oral or poster presentations.
4. Media. We will send press releases to local and national papers and media organisations.

Acknowledgements

Contributions of authors

Participants

Most of all, thanks go to the participating practitioners and patients of the following general practices: Abbeywell Surgery, Abercromby Family Practice, Acre Surgery, Acrefield Surgery, Akerman Medical Practice, Aksyr Medical Practice, Banbury Cross Health Centre, Barton Surgery, Bay Medical Group, Bellingham Green Surgery, Blackburn Road Medical Centre, Bridge Lane Group Practice, Bridgewater Surgeries, Brigstock Family Practice, Brook Green Surgery, Brownlow Group Practice, Cambrian Surgery, Cartmel Surgery, Castlegate and Derwent Surgery, Cathedral Medical Group, Chalkhill Family Practice, Chawton Park Surgery, Cheam Family Practice, Clapham Park Group Practice, Claremont Medical Centre, Clarence Medical Centre, Clevedon Medical Centre, Colliers Wood Surgery, Country Park Practice, Cowgill Surgery, Darwen Healthcare, Emsworth Surgery, Eric Moore Partnership, Eynsham Medical Centre, Fairview Medical Centre, Francis Grove Surgery, Gibson Lane Practice, Gladstone Medical Centre, Gosford Hill Medical Centre, Greenhead Family Doctors, Grove Park Terrace Surgery, Haigh Hall Medical Centre, Hampstead Group Practice, Hampton Hill Medical Centre, Hartwood Healthcare, Harvey Group Practice, Heston Health Centre (Livingcare Heston), Heswall and Pensby Practice, Highcliffe Medical Centre, Hillcrest Surgery, Homewell Practice, Honley Surgery, Horton Park Medical Practice, Hurley and Riverside Practices, Kings Medical Centre, Kirkburton Health Centre, Layton Medical Centre, Liphook and Liss Surgery, Longrove Surgery, Mann Cottage Surgery, Marine Lake Medical Practice, Meddygfa Bronyffynnon Surgery, Milman Road Health Centre, Mitcham Family Practice, Mulberry and St Denys, Newton Surgery, Nightingale Valley Practice, North House Surgery, North Kensington Medical Centre, North Street Surgery, Novum Health Partnership, Oakenhurst Medical Practice, Oaks Healthcare, Old Fire Station Surgery, Ongar Health Centre, Park and St Francis Surgery, Parliament Hill Medical Centre, Paxton Medical Group, Peel House Medical Practice, Pendle View Medical, Peterloo Medical Centre, Phoenix Surgery, Pioneer Medical Group, Plas Ffynnon Medical Centre, Plas Y Bryn Medical Centre, Preston Road Surgery, Prince of Wales Medical Centre, Queen Square Medical Practice, Queens Road Partnership, Ringwood Medical Centre, Rowden Surgery, Salisbury Medical Practice, Sedbergh Medical Practice, Selsey Medical, Shifa Surgery, Shipley Medical Practice, Shreeji Medical Centre, South Oxford Health Centre, South Oxhey Surgery, St Andrews Medical Practice, St Bartholomew's Medical Centre, St Georges Medical Centre Wirral, St George's Medical Centre Barnet, Station House Surgery, Streatham Common Practice, Summertown Health Centre, Sunnybank Medical Centre, Swanage Medical Practice, Swanwood Partnership, The Boathouse Surgery, The Bosmere Medical Practice, The Elms Medical Centre, The Exchange Surgery, The Freshford Practice (Freshwell Health Centre), The Haven Surgery, The Jenner Practice, The Mayfield Surgery, The Old Court House Surgery, The Park Surgery, The Pendle Medical Partnership, The Village Practice, The Willows Medical Centre, Thornton and Denholme Medical Practice, Three Chequers Medical Practice, Trafalgar Medical Group, Twickenham Park Medical, Two Rivers Medical Partnership, Vauxhall Primary Health Care, Village Surgery, Wakeman's Hill Practice, Wareham Surgery, West Meon Surgery, West Timperley, Westlands Medical Centre, Westwood Surgery, White Horse Medical Practice, Wokingham Medical Centre, Woodbridge Hill Surgery, Woodlands Practice, Woolstone Medical Centre and Worden Medical Practice.

Department of Health and Social Care disclaimer

This publication presents independent research commissioned by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by the interviewees in this publication

Ethics statement

The study was approved by the NHS Research Ethics Committee West of Scotland REC 5 on 21 September 2018 (reference 18/WS/0144). A number of subsequent communications were sent to both the NRES and the Health Research Authority either seeking approval for substantial amendments or informing committees of minor changes. Clinical trial authorisation was given by the Medicines and Healthcare products Regulatory Agency. The trial sponsor was the University of Southampton. In obtaining and documenting informed consent verbally, followed up in writing, the researchers complied with applicable regulatory requirements and adhered to the principles of Good Clinical Practice. Discussion of objectives, risks and inconveniences of the study and the conditions under which it was to be conducted were provided to participants by appropriately trained and delegated staff with knowledge in obtaining informed consent and with reference to the patient information leaflet.

Information governance statement

The University of Southampton was the Sponsor for the study. The University is committed to handling all personal information in line with the UK Data Protection Act (2018) and the General Data Protection Regulation (EU GDPR) 2016/679. Under the Data Protection legislation, the University of Southampton was the Data Controller, and you can find out more about the University handles personal data, including how the owners of personal data can exercise their individual rights, through the Data Protection Officer, who is the Head of Research Integrity and Governance at the University (e-mail: rgoinfo@soton.ac.uk; telephone: 02380 595058).

Data-sharing statement

The authors retain an exclusive use on the data generated from this study until 31 December 2024. Once this has expired, applications for the sharing of quantitative data can be made to the corresponding author, which will be reviewed by the lead investigators and study management team. Due to the difficulty inherent in providing anonymity and confidentiality for patient and practitioner interviews, the qualitative data are not available for sharing.

Publication

Kendrick T, Dowrick C, Lewis G, Moore M, Leydon G, Geraghty AWA, *et al*. Depression follow-up monitoring with the PHQ-9: open cluster-randomised controlled trial [Published online ahead of print February 26 2024]. *Br J Gen Pract* 2024. https://doi.org/10.3399/BJGP.2023.0539

References

1. McManus S, Bebbington P, Jenkins R, Brugha T, editors. *Mental Health and Wellbeing in England: Adult Psychiatric Morbidity Survey 2014*. Leeds: NHS Digital; 2016. URL: https://digital.nhs.uk/data-and-information/publications/statistical/adult-psychiatric-morbidity-survey/adult-psychiatric-morbidity-survey-survey-of-mental-health-and-wellbeing-england-2014 (accessed November 2023).

2. Patel K, Robertson E, Kwong ASF, Griffith GJ, Willan K, Green MJ, *et al.* Psychological distress before and during the COVID-19 pandemic among adults in the United Kingdom based on coordinated analyses of 11 longitudinal studies. *JAMA Netw Open* 2022;**5**:e227629. https://doi.org/10.1001/jamanetworkopen.2022.7629

3. McCrone P, Dhanasiri S, Patel A, Knapp M. *Paying the Price: The Cost of Mental Health Care in England to 2026.* London: King's Fund; 2008.

4. NHS Business Services Authority (NHSBSA). *Medicines Used in Mental Health: England – Quarterly Summary Statistics April to June 2021*. Newcastle: NHSBSA; 2021.

5. NHS Digital. *Psychological Therapies, Annual Report on the Use of IAPT Services, 2020–21*. Leeds: NHS Digital; 2021. URL: https://digital.nhs.uk/data-and-information/publications/statistical/psychological-therapies-annual-reports-on-the-use-of-iapt-services/annual-report-2020-21 (accessed November 2023).

6. Jorm AF, Patten SB, Brugha TS, Mojtabai R. Has increased provision of treatment reduced the prevalence of common mental disorders? Review of the evidence from four countries. *World Psychiatry* 2017;**16**:90–9. https://doi.org/10.1002/wps.20388

7. National Institute for Health and Care Excellence. *Depression in Adults: Treatment and Management* (Update). 29 June 2022. URL: www.nice.org.uk/guidance/ng222 (accessed November 2023).

8. Kendrick T, King F, Albertella L, Smith PW. GP treatment decisions for patients with depression: an observational study. *BJGP* 2005;**55**:280–6.

9. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. *Lancet* 2009;**374**:609–19. https://doi.org/10.1016/S0140-6736(09)60879-5

10. Black N. Patient-reported outcome measures could help transform healthcare. *BMJ* 2013;**346**:f167. https://doi.org/10.1136/bmj.f167

11. Knaup C, Koesters M, Schoefer D, Becker T, Puschner B. Effect of feedback of treatment outcome in specialist mental healthcare: meta-analysis. *Br J Psychiatry* 2009;**195**:15–22.

12. Shimokawa K, Lambert MJ, Smart DW. Enhancing treatment outcome of patients at risk of treatment failure: meta-analytic and mega-analytic review of a psychotherapy quality assurance system. *J Consult Clin Psychol* 2010;**78**:298–311. https://doi.org/10.1037/a0019247

13. Kendrick T, El-Gohary M, Stuart B, Gilbody S, Churchill R, Aiken L, *et al.* Routine use of patient-reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. *Cochrane Database Syst Rev* 2016;**2016**:CD011119. https://doi.org/10.1002/14651858.CD011119.pub2

14. Delgadillo J, de Jong K, Lucock M, Lutz W, Rubel J, Gilbody S, *et al.* Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial. *Lancet Psychiatry* 2018;**5**:564–72. https://doi.org/10.1016/S2215-0366(18)30162-7

15. Delgadillo J, Overend K, Lucock M, Groom M, Kirby N, McMillan D, *et al.* Improving the efficiency of psychological treatment using outcome feedback technology. *Behav Res Ther* 2017;**99**:89–97. https://doi.org/10.1016/j.brat.2017.09.011

16. Derogatis LR, Lipman RS, Rickels K, Uhlenhuth EH, Covi L. The Hopkins Symptom Checklist (HSCL): a self-report symptom inventory. *Behav Sci* 1974;**19**:1–15.

17. Mathias SD, Fifer SK, Mazonson PD, Lubeck DP, Buesching DP, Patrick DL. Necessary but not sufficient: the effect of screening and feedback on outcomes of primary care patients with untreated anxiety. *J Gen Intern Med* 1994;**9**:606–15. https://doi.org/10.1007/BF02600303

18. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression measure. *J Gen Intern Med* 2001;**16**:606–13.

19. Yeung AS, Jing Y, Brenneman SK, Chang TE, Baer L, Hebden T, *et al.* Clinical Outcomes in Measurement-based Treatment (COMET): a trial of depression monitoring and feedback to primary care physicians. *Depress Anxiety* 2012;**29**:865–73. https://doi.org/10.1002/da.21983

20. Chang TE, Jing Y, Yeung AS, Brenneman SK, Kalsekar I, Hebden T, *et al.* Effect of communicating depression severity on physician prescribing patterns: findings from the Clinical Outcomes in Measurement-based Treatment (COMET) trial. *Gen Hosp Psychiatry* 2012;**34**:105–12.

21. Montgomery SA, Asberg M. A new depression scale designed to be sensitive to change. *Br J Psychiatry* 1979;**134**:382–9.

22. Wikberg C, Westman J, Petersson EL, Larsson MEH, André M, Eggertsen R, *et al.* Use of a self-rating scale to monitor depression severity in recurrent GP consultations in primary care: does it really make a difference? A randomised controlled study. *BMC Fam Pract* 2017;**18**:6. https://doi.org/10.1186/s12875-016-0578-9

23. 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;**338**:b750.

24. Dowrick C, Leydon GM, McBride A, Howe A, Burgess H, Clarke P, *et al.* Patients' and doctors' views on depression severity questionnaires incentivised in UK quality and outcomes framework: qualitative study. *BMJ* 2009;**338**:b663.

25. Moore M, Ali S, Stuart B, Leydon GM, Ovens J, Goodall C, *et al.* Depression management in primary care: an observational study of management changes related to PHQ-9 score for depression monitoring. *BJGP* 2012;**62**:e451–7. https://doi.org/10.3399/bjgp12X649151

26. Shaw EJ, Sutcliffe D, Lacey T, Stokes T. Assessing depression severity using the UK Quality and Outcomes Framework depression indicators: a systematic review. *BJGP* 2013;**63**:e309–17. https://doi.org/10.3399/bjgp13X667169

27. Kendrick T, Stuart B, Leydon GM, Geraghty AWA, Yao L, Ryves R, *et al.* Patient-reported outcome measures for monitoring primary care patients with depression: PROMDEP feasibility randomised trial. *BMJ Open* 2017;**7**:e015266. https://doi.org/10.1136/bmjopen-2016-015266

28. Roth AJ, Kornblith AB, Batel-Copel L, Peabody E, Scher HI, Holland JC. Rapid screening for psychologic distress in men with prostate carcinoma. *Cancer* 1998;**82**:1904–8. https://doi.org/10.1002/(SICI)1097-0142(19980515)82:10<1904::AID-CNCR13>3.0.CO;2-X

29. Ashworth M, Kordowicz M, Schofield S. 'PSYCHLOPS' (Psychological Outcome Profiles): an outcome measure. *Integr Sci Pract* 2012;**2**:36–9.

30. Pettersson A, Björkelund C, Petersson EL. To score or not to score: a qualitative study on GPs views on the use of instruments for depression. *Fam Pract* 2014;**31**:215–21.

31. Kendrick T, Stuart B, Newell C, Geraghty AWA, Moore M. Did NICE guidelines and the Quality Outcomes Framework change GP antidepressant prescribing in England? Observational study with time trend analyses 2003–2013. *J Affect Disord* 2015;**186**:171–7. https://doi.org/10.1016/j.jad.2015.06.052

32. Spitzer RL, Williams JBW, Kroenke K. *Instructions for Patient Health Questionnaire (PHQ) Measure*. URL: https://www.phqscreeners.com/ (accessed November 2023).

33. Löwe B, Blankenberg S, Wegscheider K, König H-H, Walter D, Murray AM, *et al.* Depression screening with patient-targeted feedback in cardiology: DEPSCREEN-INFO randomised clinical trial. *Br J Psychiatry* 2017;**210**:132–9. https://doi.org/10.1192/bjp.bp.116.184168

34. Michie S, Van Stralen M, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011;**6**:42. https://doi.org/10.1186/1748-5908-6-42s

35. Beck AT, Steer RA, Brown GK. *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation; 1996.

36. Vanheule S, Desmet M, Groenvynck H, Rosseel Y, Fontaine J. The factor structure of the Beck Depression Inventory–II: an evaluation. *Assessment* 2008;**15**:177–87. https://doi.org/10.1177/1073191107311261

37. Mundt JC, Marks IM, Shear MK, Greist JH. The work and social adjustment scale: a simple measure of impairment in functioning. *Br J Psychiatry* 2002;**180**:461–4.

38. EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. *Health Policy* 1990;**16**:199–208.

39. EuroQol Research Foundation. *EQ-5D*. Rotterdam: EuroQol Office. URL: https://euroqol.org/eq-5d-instruments/eq-5d-5l-about/ (accessed November 2023).

40. Beecham J, Knapp M. Costing psychiatric interventions. In Thornicroft G, Brewin C, Wing J, editors. *Measuring Mental Health Needs*. Gaskell: London; 1992. pp. 163–83.

41. Meakin R, Weinman J. The 'Medical Informant Satisfaction Scale' (MISS-21) adapted for British general practice. *Fam Pract* 2002;**19**:257–63.

42. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;**166**:1092–7.

43. Dube P, Kurt K, Bair MJ, Theobald D, Williams LS. The p4 screener: evaluation of a brief measure for assessing potential suicide risk in 2 randomized effectiveness trials of primary care and oncology patients. *Prim Care Companion J Clin Psychiatry* 2010;**12**:m00978. https://doi.org/10.4088/PCC.10m00978blu

44. Button KS, Kounali D, Thomas L, Wiles NJ, Peters TJ, Welton NJ, *et al.* Minimal clinically important difference on the Beck Depression Inventory–II: according to the patient's perspective. *Psychol Med* 2015;**45**:3269–79. https://doi.org/10.1017/S0033291715001270

45. Curtis L, Burns A. *Unit Costs of Health and Social Care 2017*. Canterbury: Personal Social Services Research Unit, University of Kent. https://doi.org/10.22024/UniKent/01.02/65559

46. May CR. *Normalization Process Theory*. URL: https://normalization-process-theory.northumbria.ac.uk (accessed November 2023).

47. May CR, Albers B, Bracher M, Finch TL, Gilbert A, Girling M, *et al.* Translational framework for implementation evaluation and research: a normalisation process theory coding manual for qualitative research and instrument development. *Implement Sci* 2022;**17**:1–15. https://doi.org/10.1186/s13012-022-01191-x

48. Joint Formulary Committee. *British National Formulary* (online). London: BMJ Group and Pharmaceutical Press. URL: https://bnf.nice.org.uk/ (accessed November 2023).

49. NHS England. *National Cost Collection for the NHS*. URL: www.england.nhs.uk/costing-in-the-nhs/national-cost-collection/ (accessed November 2023).

50. Spitzer RL, Kroenke K, Williams JBW; Patient Health Questionnaire Study Group. Validity and utility of a self-report version of PRIME-MD: the PHQ Primary Care Study. *JAMA* 1999;**282**:1737–44.

51. Jiao B, Rosen Z, Bellanger M, Belkin G, Muennig P. The cost-effectiveness of PHQ screening and collaborative care for depression in New York City. *PLOS ONE* 2017;**12**:e0184210. https://doi.org/10.1371/journal.pone.0184210

52. Valenstein M, Vijan S, Zeber JE, Boehm K, Buttar A. The cost-utility of screening for depression in primary care. *Ann Intern Med* 2001;**134**:345–60.

53. Braun V, Clarke V. One size fits all? What counts as quality practice in (reflexive) thematic analysis? *Qual Res Psychol* 2021;**18**:328–52. https://doi.org/10.1080/14780887.2020.1769238

54. Browning M, Bilderbeck AC, Dias R, Dourish CT, Kingslake J, Deckert J, *et al*. The clinical effectiveness of using a predictive algorithm to guide antidepressant treatment in primary care (PReDicT): an open-label, randomised controlled trial. *Neuropsychopharmacology* 2021;**46**:1307–14. https://doi.org/10.1038/s41386-021-00981-z

55. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, *et al*. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003;**54**:573–83.

56. Health Education England. *National Curriculum for Psychological Wellbeing Practitioner (PWP) Programmes*. 4th edn. London: Health Education England; 2022.

57. 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**:e279–89. https://doi.org/10.3399/bjgp11X572472

58. Leydon GM, Dowrick CF, McBride AS, Burgess HJ, Howe AC, Clarke PD, *et al*.; QOF Depression Study Team. Questionnaire severity measures for depression: a threat to the doctor–patient relationship? *Br J Gen Pract* 2011;**61**:117–23. https://doi.org/10.3399/bjgp11X556236

59. Robinson J, Khan N, Fusco L, Malpass A, Duffy L, Lewis G, *et al*. Why are there discrepancies between depressed patients' Global Rating of Change and scores on the Patient Health Questionnaire depression module? A qualitative study of primary care in England. *BMJ Open* 2017;**7**:e014519. https://doi.org/10.1136/10.1136/bmjopen-2016-014519

60. Pinto-Meza A, Serrano-Blanco A, Peñarrubia MT, Blanco E, Haro JM. Assessing depression in primary care with the PHQ-9: can it be carried out over the telephone? *J Gen Intern Med* 2005;**20**:738–42. https://doi.org/10.1111/j.1525-1497.2005.0144.x

61. Ford J, Thomas F, Byng R, McCabe R. Use of the Patient Health Questionnaire (PHQ-9) in practice: interactions between patients and physicians. *Qual Health Res* 2020;**30**:2146–59. https://doi.org/10.1177/1049732320924625

62. National Institute for Health and Care Excellence. *Quality Statement 1: Assessment*. URL: www.nice.org.uk/guidance/qs8/chapter/Quality-statement-1-Assessment (accessed November 2023).

63. US Federal Health Resources and Services Administration. Uniform Data System Clinical Quality Measures 2020. URL: https://bphc.hrsa.gov/sites/default/files/bphc/data-reporting/2020-clinical-measures-handout.pdf (accessed November 2023).

64. US Department of Veterans Affairs/Department of Defense. *VA DoD Clinical Practice Guidelines*. URL: www.healthquality.va.gov/guidelines/MH/mdd/ (accessed November 2023).

65. Kaiser Permanente Health Maintenance Organization. *Mental Health Monitoring Tool*. URL: https://wa-provider.kaiserpermanente.org/static/pdf/provider/patient-ed/screenings/bhi-monitoring-tool.pdf (accessed November 2023).

66. Nederlands Huisartsen Genootschap (Dutch Society of General Practitioners). *Depressie*. URL: https://richtlijnen.nhg.org/standaarden/depressie#volledige-tekst-3-beoordelen-van-de-ernst-van-de-depressieve-stoornis (accessed November 2023).

67. Duffy L, Lewis G, Ades A, Araya R, Bone J, Brabyn S, *et al*. Antidepressant treatment with sertraline for adults with depressive symptoms in primary care: the PANDA research programme including RCT. *Programme Grants Appl Res* 2019;**7**(10).

68. Kendrick T, Maund E. Do PROMS improve outcomes in patients with depression in primary care? *BMJ* 2020;**370**:m3313. https://doi.org/10.1136/bmj.m3313

69. Office for National Statistics. *Office for National Statistics 2011 Census: Age Groups*. URL: www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/demographics/age-groups/latest (accessed November 2023).

70. Kendrick T, Moore M, Leydon G, Stuart B, Geraghty AWA, Yao G, *et al*. Patient-reported outcome measures for monitoring primary care patients with depression (PROMDEP): study protocol for a randomised controlled trial. *Trials* 2020;**21**:441. https://doi.org/10.1186/s13063-020-04344-9

Appendix 1 Patient Health Questionnaire-9 training feedback questionnaire scores

1. **In the PROMDEP study, the first PHQ-9 questionnaire will be administered to the patient by a member of the research team.**

		Frequency	Per cent
Valid	Incorrect	8	4.5
	Correct	168	95.5
	Total	176	100.0

2. **The PHQ-9 is a quick and adequate substitute for a global enquiry into how the patient is feeling.**

		Frequency	Per cent
Valid	Incorrect	28	15.9
	Correct	148	84.1
	Total	176	100.0

3. **A score of 10–14 on the PHQ-9 suggests moderate depression, and that active treatment might be indicated.**

		Frequency	Per cent
Valid	Incorrect	7	4.0
	Correct	169	96.0
	Total	176	100.0

4. **A score of 5–9 on the PHQ-9 excludes any need for active treatment.**

		Frequency	Per cent
Valid	Incorrect	27	15.3
	Correct	149	84.7
	Total	176	100.0

5. **Very few patients have a mismatch between their PHQ-9 scores and their own global rating of how they feel.**

		Frequency	Per cent
Valid	Incorrect	39	22.2
	Correct	137	77.8
	Total	176	100.0

6. **A score of 15 or more on the PHQ-9 suggests moderately severe depression, and that combined treatment with an antidepressant plus psychological treatment might be indicated.**

		Frequency	Per cent
Valid	Incorrect	22	12.5
	Correct	154	87.5
	Total	176	100.0

7. **If the PHQ-9 score is less than 10, but the patient says they are depressed, this may be because their particular symptoms are not covered well by the PHQ-9.**

		Frequency	Per cent
Valid	Incorrect	8	4.5
	Correct	168	95.5
	Total	176	100.0

8. **In the PROMDEP study, the second follow-up PHQ-9 will be administered to the patient by a member of the research team.**

		Frequency	Per cent
Valid	Incorrect	31	17.6
	Correct	145	82.4
	Total	176	100.0

9. **A reduction of three points in the PHQ-9 follow-up score on treatment is adequate to show treatment is working.**

		Frequency	Per cent
Valid	Incorrect	56	31.8
	Correct	120	68.2
	Total	176	100.0

10. **If the PHQ-9 score at follow-up is still high, but the patient says they are less depressed, this may be because one or two troublesome symptoms are having a disproportionate effect on the total score.**

		Frequency	Per cent
Valid	Incorrect	2	1.1
	Correct	174	98.9
	Total	176	100.0

Descriptive statistics

	N	Minimum	Maximum	Mean	Std. deviation
Total score	176	5	10	8.70	1.306

CASE HISTORIES

The following case histories are reprinted with permission from: Linda Gask, Tony Kendrick, Robert Peveler, Carolyn A. Chew-Graham (Eds.), Primary Care Mental Health, 2nd Edition © The Royal College of Psychiatrists 2009, 2018, published by Cambridge University Press (9781911623021).

The most appropriate responses are indicated by 'TRUE' and score 2 points, but other debatable, potentially appropriate, responses are acknowledged, and score 1 point each.

Case 1

<table>
<tr>
<td>
John is 68 years old and attends your surgery for a check-up and a repeat prescription of aspirin, digoxin, furosemide and finasteride. He apologises profusely if he is wasting your time but wants to ask you about some new symptoms while he's here. He complains of feeling tired for some weeks, all day long, in spite of sleeping more than usual. He is becoming forgetful and thinks he is getting senile. He has stopped following the horseracing on TV as he cannot concentrate anymore and he is eating less and has lost a bit of weight. Although he is not a man who gets depression, he admits there really is no enjoyment at all in his life. He has been lonely since his wife died 2 years ago. He goes down to the pub but hardly knows anyone there anymore.

On pressing him gently, he admits he feels he has nothing to live for, but has no plans to harm himself. He does think he could 'do with a bit of a tonic'. His PHQ-9 score is 9/27 (mild depression).
</td>
<td>

Which action would be most appropriate at this point?

1. Reassure John that he does not have depression requiring treatment.

FALSE Score 0. Not at this point – John has at least five symptoms of major depressive disorder including possible suicidal thoughts and he needs further assessment – the low PHQ-9 score is a false negative possibly related to his unwillingness to admit to symptoms.

2. Advise John to access computerised CBT on-line.

FALSE Score 0. Probably not – unsupported self-help is not a first line choice for what is possibly more severe depression, and even if he is used to accessing resources on-line, John has a problem with concentrating.

3. Start antidepressant treatment with a SSRI, mirtazapine or lofepramine.

FALSE Score 0. Not at this point – John has cardiac problems including atrial fibrillation and should not start an antidepressant without an ECG to check the QT interval which has to be corrected and interpreted carefully in the presence of AF. His potassium level needs checking too as he is on furosemide and any QT prolongation would be more dangerous in the presence of hypokalaemia. Lofepramine also might put him into urinary retention as he is on finasteride for benign prostatic hypertrophy.

4. Offer John referral to your local psychological therapies service for cognitive behavioural therapy or behavioural activation.

DEBATABLE Score 1. This is a reasonable course of action but probably not the best choice at this point – John needs further assessment to establish whether he has major depressive disorder which is severe and persistent enough to require treatment or referral. A significant proportion of patients have fewer symptoms on follow-up, having had the chance to talk them through. If he turns out to have less severe depression related to bereavement and social isolation then bereavement counselling and help with increasing his social network might be preferable alternatives.

5. Arrange for John to have some blood tests and an ECG and review him for further assessment of his mood in 2 weeks' time.

TRUE Score 2. This is probably the best action at this point. This will enable you to assess whether he has persisting symptoms as well as possible contraindications to antidepressant treatment, and decide on the severity of his depression and appropriate treatment and referral options. In view of his self-professed memory issue, it would also be good practice to assess John for cognitive impairment with the mini-mental state examination (MMSE), 'Mini-Cog', or Addenbrooke's Cognitive Examination–Revised (ACE-R) before starting treatment. You will obviously want to discuss with him your suspicion that he has depression and gauge his response to that, providing some information on why his symptoms suggest possible depression.
</td>
</tr>
</table>

Results

	Frequency	Per cent
Valid 1	4	2.2
2	13	7.3
3	4	2.2
4	63	35.4
5	94	52.8
Total	178	100.0

Case 2

Daniela, who is 23, comes to see you complaining of depression. She says she's come early this time because last time she was depressed it went on for months and became quite bad before she told anyone, and she ended up being referred to a psychiatric clinic. She wants to nip this episode in the bud and is asking for antidepressants. She has been tearful most of the time for the last 3 or 4 weeks, anxious, and not sleeping at all well, waking at 5am every day. She has had some churning in her stomach, palpitations and diarrhoea. She seems agitated but denies any suicidal ideas. She admits on enquiry to poor energy levels, poor concentration, loss of interest in sex and erratic eating with a few pounds of weight loss. She has given up her job as a busy receptionist and is looking for a new one, possibly further away from home as she has just passed her driving test and bought a new car. She says she had amitriptyline 25 mg, two at night for 2 months, 6 months previously, from a doctor in Bulgaria, which she stopped because she was feeling better.

You ask her to complete two symptom questionnaires and find her PHQ-9 score is 18 (moderately severe depression), and her GAD-7 score is 14 (moderate generalised anxiety – see https://patient.info/doctor/generalised-anxiety-disorder-assessment-gad-7)

Which action would be most appropriate at this point?

1. Offer to restart amitriptyline 25 mg at night.

FALSE Score 0. Probably not – tricyclics are no longer recommended as first-line antidepressants in the UK, and amitriptyline can impair driving ability, so although it helped previously it is best avoided if possible.

2. Refer Daniela to the in-practice counsellor.

FALSE Score 0. No – counselling is not recommended as a first-line treatment for more severe depression.

3. Offer to start sertraline 25 mg daily and review Daniela in one week.

TRUE Score 2. Probably the most appropriate action – the usual starting dose of 50 mg dose could increase Daniela's anxiety, so whether you are giving it for anxiety or depression, 25 mg may be a preferable starting dose, and the new NICE guideline recommends seeing patients aged under 30 one week after starting an antidepressant.

4. Offer to start mirtazapine 15 mg at night and review in four weeks.

DEBATABLE Score 1. Not a bad choice – mirtazapine would help the sleep and agitation and is an appropriate choice in someone who has been depressed before, is developing what seems to be more severe depression, and is coming asking for antidepressants. However, starting her on mirtazapine 15 mg might make her anxiety worse initially, and for that reason the new NICE guideline recommends seeing patients aged under 30 one week after starting an antidepressant.

5. Decline Daniela's request for antidepressants, refer her for psychological therapy, and ask her to come back if that does not help her.

FALSE Score 0. While referring Daniela for psychological therapy is a good idea (having explored whether she wants this), this should probably be in combination with an antidepressant rather than instead of drug treatment. Daniela feels she is getting worse and is specifically asking for drug treatment. Referral for psychological treatment usually involves a wait of at least some weeks, and she seems to want help sooner than that. If you are not starting treatment today she should at least be actively reviewed with a specific date for follow-up within 2 weeks.

Results

		Frequency	Per cent
Valid	1	14	7.9
	2	5	2.8
	3	68	38.2
	4	88	49.4
	5	3	1.7
	Total	178	100.0

Case 3

Hanna is 30 and tells you she is fed up. She would like to go back on Prozac which she had some years previously for postnatal depression. She has three kids under six and never stops running around after them. One has just been in hospital with an asthma attack. Her husband has lost his job again and is drinking more. She admits she is having wine every night with him once the kids have gone to bed. She is sleeping OK but gets tired in the evenings so she goes to bed at 9.30 pm. She is eating more than usual but not putting on weight, which she thinks is because she has been smoking more cigarettes too. She is managing to g*et al*l her jobs done although it is an effort. She still enjoys watching her favourite TV programmes and likes going out twice a week when her mum babysits for them. She is looking forward to next month when her middle child will start school. She finds it hard to talk to her husband and when she gets tearful he goes off down to the pub. Her PHQ-9 score is 8 (mild depression). In response to your direct enquiry she denies any history of domestic abuse. There is no evidence in her records that her children have ever been considered to be at risk.	**Which action would be most appropriate at this point?** 1. Agree to restart Prozac. **FALSE Score 0**. Not at this point. Hanna probably does not have major depressive disorder. She has fewer than five symptoms of depression, her mood appears situational rather than being persistently down, and her day-to-day functioning is not significantly impaired. Drug treatment may expose her to side effects, especially if she continues to drink every day, and especially if she becomes pregnant again. These reasons should be explained, together with an offer to review her mood within a few weeks if it is not improving. 2. Offer referral to the local psychological therapies service. **DEBATABLE Score 1**. Probably not the best choice at this point. Hanna has fewer than five symptoms of depression, her mood appears situational rather than being persistently down and her day-to-day functioning is not significantly impaired. Referral would probably not be cost-effective and would add to the waiting list for the service. She is likely to improve with time anyway, so the best approach would be to reassess her mood over time. However she might refer herself and be assessed in the meantime. 3. Advise Hanna on safe drinking levels, talk through what options she has for getting help with childcare, and ask her to come back for a review in a month if things have not improved. **TRUE Score 2**. Probably the best first step at this point. Hanna may not realise that daily wine puts her in a relatively high bracket for alcohol consumption, or that it tends to depress the mood over time, even though it makes her feel better at the time she drinks it. You could signpost her to specific alcohol counselling services or groups if they are available locally. The situation is likely to improve when the middle child starts school, and it is safe to temporise as there is no evidence of domestic or child abuse. 4. Refer Hanna to the health visitor. **FALSE Score 0**. Probably not at this point. This is a reasonable course of action if the problems continue and a health visitor is available, but Hanna is an experienced mother, and this may not be the best use of a scarce resource. You could however discuss her family with the health visitor if you have one available, to check whether there has been any suggestion previously that the children may be at risk. 5. Refer Hanna to social services. **FALSE Score 0**. Not at this point. This is a very common situation and referral may not be the best use of a scarce resource unless any evidence emerges that the children are at risk.

Results

		Frequency	Per cent
Valid	1	2	1.1
	2	27	15.2
	3	143	80.3
	4	6	3.4
	Total	178	100.0

*No respondents gave an answer of "5. Refer Hanna to social services."

Case 4

Christina is 45 and runs her own specialised pottery business. She tells you she has been low for some months, with no energy, poor appetite, interrupted sleep, poor concentration and loss of interest in her usual trips to the opera and theatre. Everything seems to be going wrong. She lost a fairly sizeable contract recently, although generally her business is in good shape. A romantic relationship has also ended after only a few weeks, the third one in recent years. She says she knows it's silly, but everything seems so pointless. She would not do anything to end her life, as this would devastate her widowed mother, although she admits the thought has crossed her mind. She should be happy as she is now quite wealthy, but nothing seems to be worth the effort any more. She can't remember what it was that she used to enjoy so much about her business. The future looks grim. She thinks she needs a head transplant. She can see she is not thinking straight and lacks self-esteem, and wonders if it is because she was never really valued, either by her parents or by her ex-husband. She disapproves of pills because all they do is dampen down your feelings and stop you thinking through the problem, and can be addictive. She tried them before and they gave her unpleasant side effects. She also had a course of CBT in the past without much help. Her PHQ-9 score is 18 (moderate to severe depression).	**Which action would be most appropriate at this point?** 1. Advise Christina that antidepressants are not addictive and prescribe an SSRI. **DEBATABLE Score 1**. Probably not the best choice of action at this point – while Christina might be persuaded that antidepressants are not addictive, she has tried medication already and is unlikely to want to take it again at this point. She is seeking an alternative. 2. Advise her to join a gym and start aerobic exercise three times a week for 45 minutes a time. **FALSE Score 0**. Probably not – exercise is not recommended as a first-line treatment for more severe depression, and in any case supervised group exercise classes are preferable to unsupervised self-directed exercise. 3. Refer her to the local psychological therapy service for CBT or behavioural activation. **DEBATABLE Score 1**. This is a reasonable course of action, either alone or in combination with an offer to restart antidepressant treatment if things don't improve, but probably not the best choice given she has tried CBT before, which includes a behavioural component, and it didn't help. 4. Refer her for short-term psychodynamic therapy and ask her to return for a review of her mood in 2 weeks. **TRUE Score 2**. This is probably the most appropriate choice, given that she has had problems with relationships which she relates to childhood and which have recurred repeatedly through her life, so a developmental approach might suit her best. If it is not available quickly through the NHS she may have to go privately, but fortunately for her she can probably afford this. Seeing her again in 2 weeks is a safety net in case she gets worse in the meantime. You would need to explore whether at that point she would like to restart antidepressant treatment (with an alternative to the drug which gave her side effects previously). 5. Reassure her things will improve with time and arrange to review her in a few weeks if she is not feeling better. **FALSE Score 0**. This is probably not the best choice as it is unlikely she will improve within a few weeks given the problem has been going on for some months. She has quite marked depression, she is at risk of developing suicidal plans, and waiting will only delay referring her for help.

Results

		Frequency	Per cent
Valid	1	63	35.4
	2	1	.6
	3	32	18.0
	4	81	45.5
	5	1	.6
	Total	178	100.0

Total scores

Score	
Valid	178
Missing	0
Mode	6
Std. Deviation	1.357
Minimum	2
Maximum	8

		Frequency	Per cent
Valid	2	1	.6
	3	4	2.2
	4	25	14.0
	5	39	21.9
	6	50	28.1
	7	33	18.5
	8	26	14.6
	Total	178	100.0

PHQ9 Quiz

GP	Score 1	Score 2	Score 3	Difference between first and last score
1	5	10		5
2	10	8		-2
3	7	10		3
4	6	7		1
5	8	10		2
6	7	9		2
7	7	9	10	3
8	10	9		-1
9	7	10		3
10	6	10		4
11	7	10		3
12	10	9		-1
13	6	10		4
14	5	9		4
15	9	10		1
16	10	10		0
17	10	10		0
18	10	9		-1
19	8	7		-1
20	7	7		0
21	6	10		4
22	7	7		0

Case Vignette Quiz

GP	Score 1	Score 2	Score 3	Difference between first and last score
1	4	4		0
2	5	4		-1
3	4	4		0
4	6	5		-1
5	4	5	8	4
6	6	6		0
7	5	6		1
8	5	6		1
9	6	6		0
10	6	6		0
11	6	7	8	2
12	7	7		0
13	7	7		0
14	5	8		3
15	2	8		6
16	5	8		3
17	6	8		2
18	7	8		1
19	4	8		4
20	7	8		1
21	6	8		2
22	5	8		3

Appendix 2 PROMDEP RCT standard operating procedure for suicidal ideation

1. PURPOSE

This standard operating procedure (SOP) has been written to describe the procedure for suicidal ideation that may be expressed by PROMDEP participants.

2. BACKGROUND

People with distress can experience suicidal thoughts and ideas (suicidal ideation), and this SOP is in place to ensure that when this is identified in PROMDEP participants it is recorded and reported. This SOP sets out instructions to achieve consistent practice and includes specific instructions for recording suicidal ideation, and reporting lines.

3. SCOPE OF THIS SOP

This SOP applies to all PROMDEP participant contacts, both pre and post consent.

4. RESPONSIBLE PERSONNEL

Researchers – reporting instances of suicidal ideation to study chief investigator and participant GPs, documenting instances on paper and on study database.

Principal investigator with clinical responsibility (Tony Kendrick, Michael Moore, Chris Dowrick, Glyn Lewis or nominated deputy) – for training researchers in the procedure, advising them, reviewing reports of suicidal ideation, monitoring adherence to procedure and management oversight.

5. PROCEDURE

5.1 Definition of suicidal ideation:

Any participants who disclose information during an interview (telephone or face to face) to the researcher indicating that they have been thinking of taking their own life will be considered to have suicidal ideation.

At screening:

In the PROMDEP study, each participant is asked the following two questions over the telephone at screening, before being seen face to face for the baseline assessment:

1.   Have you attempted suicide in the past few months?
2.   Are you currently thinking of ways to end your life?

If the answer to either screening question is yes, the researcher should then implement the suicidal ideation procedure described in 5.2 below.

Patients with active suicidal plans should be excluded, but patients with suicidal ideation but no active plans may be admitted into the study (after discussion with the principal investigator). The GP will be informed as soon as the suicidal ideation SOP is implemented.

At face-to-face assessment:

Each participant in the PROMDEP study is administered the Beck Depression Inventory (BDI-II), at baseline, 12-week follow-up, and 26-week follow-up assessments. The question used to identify suicidal ideation within the BDI-II is question 9:

Pick out the one statement in each group which best describes the way you have been feeling in the PAST WEEK, INCLUDING TODAY:

I don't have any thoughts of killing myself   0

I have thoughts of killing myself, but I would not carry them out   1

I would like to kill myself   2

I would kill myself if I had the chance   3

A score of 1, 2 or 3 – that is, any score other than zero on question 9 of the BDI-II – will alert the researcher to possible suicidal ideation, and the researcher should then implement the suicidal ideation procedure described in 5.2 below.

In addition, participants in the intervention arm are given the Patient Health Questionnaire (PHQ-9) to complete as a patient-reported outcome measure (PROM) at baseline. The question used to identify suicidal ideation within the PHQ-9 is question 9:

Over the last 2 weeks, how often have you been bothered by thoughts that you would be better off dead or of hurting yourself in some way?

Not at all   0

Several days   1

More than half the days   2

Nearly every day   3

Again, a score of 1, 2 or 3 – that is, any score other than zero on question 9 of the PHQ-9 – will alert the researcher to possible suicidal ideation, who will then follow the procedure in 5.2 below.

Receiving online assessments

Please follow the instructions above for the face-to-face assessment, ensuring that the BDI-II and PHQ-9 (if applicable) are checked as soon as possible after receiving the online assessment, and ideally within 24 hours.

5.2 Implementation of procedure:

General practitioners are responsible for the ongoing clinical care of participants. Therefore, researchers have a duty of care to ensure that the person's GP is aware of suicidal ideation.

Researchers must initiate the suicidal ideation SOP each time a participant expresses thoughts of suicide or self-harm. This may be as a result of responses to the screening questions or the baseline or follow-up assessment questionnaire items, or the participant may disclose other information during an

interview that leads the researcher to believe that there is a suicide risk. In any of these instances, the researcher, with the participant's permission, should inform the participant's GP (by NHS e-mail and/or telephone, to maintain confidentiality). **Before notifying the GP, however, the researcher should explore the participant's ideas further if possible by asking them the four questions of the PS4 suicide screener.**

PS4 suicide screener questions:

1. Have you ever attempted to harm yourself in the past?
   NO YES
2. Have you thought about how you might actually hurt yourself?
   NO YES [How?]
3. There's a big difference between having a thought and acting on a thought. How likely do you think it is that you will act on these thoughts about hurting yourself or ending your life some time over the next month?
   a. Not at all likely; b. Somewhat likely; c. Very likely
4. Is there anything that would prevent or keep you from harming yourself?
   NO YES [What?]

Risk category	Shaded ('risk') response	
	Items 1 and 2	**Items 3 and 4**
Minimal	Neither is shaded	Neither is shaded
Lower	At least one item is shaded	Neither is shaded
Higher		At least one item is shaded

The researcher should then ascertain whether or not the participant has talked to his/her doctor. The researcher should reinforce the importance of maintaining a dialogue with the GP **and ask for permission to pass the information to the GP (see appendix – suggested scripts).** Researchers are not clinically trained and should not attempt to assess the seriousness of the disclosure but should adhere to the policy. A letter (**Form 1 below**) should be completed, scanned, and sent by NHS e-mail to the GP practice, informing them of the risk disclosed.

Minimal risk:

If the participant agrees for this information to be disclosed to their GP, **the researcher should e-mail the GP letter Form 1 by NHS e-mail to the participant's GP practice as soon as possible, ideally within 2 working days,** to pass on the information obtained**.**

Lower or higher risk:

If the participant agrees for this information to be disclosed to their GP, the researcher should e-mail the GP letter Form 1 by NHS e-mail to the participant's GP practice as soon as possible, AND ALSO TELEPHONE THE PRACTICE as soon as possible, on the same day or next working day to pass on the information obtained.** If the participant's GP is not available to speak, then the researcher should ask to speak to the duty doctor. The researcher should make it clear to the GP that no clinical risk assessment has been performed, and that the clinical responsibility for the study participant remains with the GP.

> **If the researcher believes the participant is in immediate danger, the researcher must immediately contact the GP, or duty doctor, who will take appropriate action, or if necessary call an ambulance.

If the participant refuses permission for the researcher to inform the GP then the researcher should immediately consult the Principal Investigator (Tony Kendrick or Michael Moore in Southampton, Chris Dowrick in Liverpool, and Glyn Lewis at UCL, or a nominated deputy if they are not available) to discuss the participant's data and answers. The principal investigator will explore further with the patient if necessary. If it is concluded that there is a significant risk, the participant's GP will be notified **with or without** the participant's consent. In addition, the decision should be explained to the participant as soon as possible.

PROMDEP **Southampton** **LIVERPOOL** **UCL**

Form 1- GP Letter

STRICTLY CONFIDENTIAL

Dear Doctor,

PROMDEP Study notification

Participant Name:……………………… DOB: ……/……/……….

I am writing to notify you that the above PROMDEP study participant reported the following information and agreed to this information being passed to you.

The patient indicated on the BDI-II questionnaire that over the past week they had had:

Thoughts of killing themselves, but would not carry them out ☐
Thoughts they would like to kill themself ☐
Thoughts they would kill themself if they had the chance ☐

The patient indicated on the PHQ-9 questionnaire that over the past 2 weeks they had thoughts that they would be better off dead or of harming themselves, for:

Several days ☐
More than half the days ☐
Nearly every day ☐

The patient indicated they had attempted to harm themself in the past
Yes ☐
No ☐

The patient had thought about how they might actually hurt themself
Yes ☐
No ☐

The patient thought that acting on these thoughts about hurting themself or ending their life some time over the next month was:

Not at all likely ☐
Somewhat likely ☐
Very likely ☐

The patient indicated there was something that would prevent or keep them from harming themself:

NO ☐

YES ☐

The patient further disclosed the following information:

………

………

………

……

………

………

……

……

Furthermore, the patient informed us that they:

Have notified you of these thoughts ☐

Have not notified you of these thoughts ☐

We would remind you that no clinical risk assessment has been carried out as part of the PROMDEP study, and that ongoing clinical care remains with you.

Yours sincerely,

Name: Date :

Phone: E-mail:

Appendix

Suggested Scripts:

Disclosure via questionnaire

Thank you very much for answering these questions to us. We really appreciate your taking part in the study; the information you give us is invaluable and may help others who suffer from distress. However, I notice from the questionnaire you completed that you have had thoughts of harming yourself. Have you spoken to your doctor about these thoughts? These thoughts sometimes happen in distress, and it is really important to talk (or keep talking) to your doctor about them. We'd like to tell your GP about these thoughts and hope this is ok with you?

Disclosure during interview (face-to-face or telephone)

I am concerned about some of the things you have told me. Have you spoken to your doctor about them? It is important that your GP knows about the way you feel, as she/he will be able to make sure that you have the necessary support in place.

If participant is hesitant or refuses

Many people find it hard to bring these things up during a consultation, but your GP can offer you help with these feelings. If he/she knows how you are feeling, he/she will be able to talk to you about it and together you can decide on the best way to treat you. The doctors in charge of this study strongly recommend we tell your GP.

If participant continues to refuse

That fine, but as I am not a medical doctor, I do have to let my colleague (Prof Tony Kendrick, Michael Moore, Chris Dowrick or Glyn Lewis) know about the way you are feeling. We will have to decide whether or not to inform your GP without your consent. If we do that we will inform you that we have informed your GP of our concerns.

Appendix 3　Normalisation process theory framework

NPT components	Questions to consider within the NPT framework
Coherence (i.e. meaning and sense making by participants)	Is the intervention easy to describe?
	Is it clearly distinct from other interventions?
	Does it have a clear purpose for all relevant participants?
	Do participants have a shared sense of its purpose?
	What benefits will the intervention bring and to whom?
	Are these benefits likely to be valued by potential participants?
	Will it fit with the overall goals and activity of the organisation?
Cognitive participation (i.e. commitment and engagement by participants)	Are target user groups likely to think the intervention is a good idea?
	Will they see the point easily?
	Will they be prepared to invest time, energy and work in it?
Collective action (i.e. the work participants do to make the trial function)	How will the intervention affect the work of user groups?
	Will it promote or impede their work?
	What effect will it have on consultations?
	Will staff require extensive training before they can use it?
	How compatible is it with existing work practices?
	What impact will it have on division of labour, resources, power and responsibility between different professional groups?
	Will it fit with the overall goals and activity of the organisation?
Reflexive monitoring (i.e. participants reflect on or appraise the trial)	How are users likely to perceive the intervention once it has been in use for a while?
	Is it likely to be perceived as advantageous for patients or staff?
	Will it be clear what effects the intervention has had?
	Can users/staff contribute feedback about the intervention once it is in use?
	Can the intervention be adapted/improved on the basis of experience?