Kendrick, T, Dowrick, C, Lewis, G, Moore, M, Leydon, G, Geraghty, AWA, Griffiths, G, Zhu, S, Yao, G, May, C, Gabbay, M, Dewar-Haggart, R, Williams, S, Bui, L, Thompson, N, Bridewell, L, Trapasso, E, Patel, T, McCarthy, M, Khan, N, Page, H, Corcoran, E, Hahn, JS, Bird, M, Logan, MX, Ching, BCF, Tiwari, R, Hunt, A and Stuart, B

Depression follow-up monitoring with the PHQ-9: open cluster-randomised controlled trial

Accepted Manuscript

British Journal of General Practice

Depression follow-up monitoring with the PHQ-9: open cluster-randomised controlled trial

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

DOI: https://doi.org/10.3399/BJGP.2023.0539

To access the most recent version of this article, please click the DOI URL in the line above.

Received 17 October 2023
Revised 16 January 2024
Accepted 19 February 2024

Depression follow-up monitoring with the PHQ-9: open cluster-randomised controlled trial

Tony Kendrick[1] BSc MD FRCGP FRCPsych FHEA, professor of primary care, https://orcid.org/0000-0003-1618-9381 Christopher Dowrick[2] BA MSc MD FRCGP, emeritus professor of primary medical care, https://orcid.org/0000-0002-4245-2203 Glyn Lewis[3] PhD FRCPsych professor of epidemiological psychiatry, https://orcid.org/0000-0001-5205-8245 Michael Moore[1] MSc emeritus professor of primary care research, https://orcid.org/0000-0002-5127-4509 Geraldine M Leydon[1] BA MA PhD, professor of medical sociology and communication, https://orcid.org/0000-0001-5986-3300 Adam W A Geraghty[1] PhD CPsychol FHEA, associate professor of psychology and behavioural medicine, https://orcid.org/0000-0001-7984-8351 Gareth Griffiths[4] BSc MSc PhD, professor of clinical trials, https://orcid.org/0000-0002-9579-8021 Shihua Zhu[1] PhD, senior research fellow in health economics, https://orcid.org/0000-0002-1430-713X Guiqing Lily Yao[5] PhD, professor of health economics, https://orcid.org/0000-0002-9095-6676 Carl May[6] PhD FAcSS, professor of medical sociology, https://orcid.org/0000-0002-0451-2690 Mark Gabbay[2] MD FRCGP DFSRH Dip Psychotherapy, professor of general practice, https://orcid.org/0000-0002-0126-8485 Rachel Dewar-Haggart[1] BSc MSc PhD CPsychol, research fellow, https://orcid.org/0000-0002-3757-1152 Samantha Williams[1] BSc MRes, senior research assistant, https://orcid.org/0000-0001-9505-6485 Lien Bui[1] BA MSc PhD, research fellow, https://orcid.org/0000-0003-3434-4066 Natalie Thompson[1] BSc MSc, senior research assistant, https://orcid.org/0000-0002-1880-6438 Lauren Bridewell[1] BMedSci, medical student, https://orcid.org/0000-0001-9569-7813 Emilia Trapasso[2] BSc MSc PhD, postdoctoral research fellow in public engagement, https://orcid.org/0000-0002-7539-7204 Tasneem Patel[2] PhD, research assistant, https://orcid.org/0000-0003-4496-7603 Molly McCarthy[2] BSc MSc, research assistant, https://orcid.org/0000-0002-2504-6799 Naila Khan[2] PhD, research associate, https://orcid.org/0000-0003-3400-7190 Helen Page[2] PhD, research associate, https://orcid.org/0000-0002-5781-9282 Emma Corcoran[3] MSci DClinPsy, clinical psychologist, https://orcid.org/0000-0001-5811-4615 Jane Sungmin Hahn[3] BA MSc, research assistant, https://orcid.org/0000-0002-4584-9441 Molly Bird[3] MSc, research assistant, https://orcid.org/0000-0003-1676-1907 Mekeda X Logan[3] BSc MSc, research assistant, https://orcid.org/0000-0001-7899-7531 Brian Chi Fung Ching[3] BSc MSc, research assistant, https://orcid.org/0000-0002-2179-9793 Riya Tiwari[1] MSc, senior research assistant, https://orcid.org/0000-0002-3002-3276 Anna Hunt[1] BSc MSc, senior research assistant, https://orcid.org/0000-0002-0864-4113 Beth Stuart[7] PhD, professor of medical statistics and clinical trials, https://orcid.org/0000-0001-5432-7437

[1]School of Primary Care, Population Science, and Medical Education, Faculty of Medicine, University of Southampton, Aldermoor Health Centre, Aldermoor Close, Southampton, SO16 5ST, UK.
[2]Department of Primary Care and Mental Health, Institute of Population Health, University of Liverpool, Waterhouse Block B 1-5 Brownlow St Liverpool L69 3GL, UK.
[3]Division of Psychiatry, University College London, 6th Floor Maple House, 149 Tottenham Court Road, London W1T 7NF, UK.
[4]Southampton Clinical Trials Unit, University of Southampton and University Hospital Southampton NHS Foundation Trust, Tremona Road, Southampton, SO16 6YD, UK.
[5]Leicester Clinical Trials Unit, University of Leicester, Maurice Shock Building, University Road, Leicester LE1 7RH, UK.
[6]Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, 15-17 Tavistock Place, London, WC1H 9SH, UK.
[7]Centre for Evaluation and Methods, Wolfson Institute of Population Health, Faculty of Medicine and Dentistry, Queen Mary University of London, Yvonne Carter Building, 58 Turner Street, London E1 2AB, UK.

Correspondence to: Tony Kendrick, Professor of Primary Care, School of Primary Care, Population Health & Medical Education, University of Southampton, Aldermoor Health Centre, Southampton, SO16 5ST. Email: ark1@soton.ac.uk ORCid ID: https://orcid.org/0000-0003-1618-9381

Disclosure of interest

All authors completed a disclosure form at: http://www.icmje.org/coi_disclosure.pdf.

Tony Kendrick, Christopher Dowrick, Glyn Lewis, Michael Moore, Geraldine Leydon, Adam Geraghty, Gareth Griffiths, Shihua Zhu, Guiqing Lily Yao, Carl May, Mark Gabbay and Beth Stuart have received grant funding to their employer universities from the National Institute for Health and Social Care Research (NIHR) to carry out this study and other research. In addition Glyn Lewis has received grant funding from the MRC and Wellcome. Gareth Griffiths has received funding from Janssen-Cilag, AstraZeneca, Novartis, Astex, Roche, Heartflow, Celldex, BMS, BioNTech, Cancer Research UK, the NIHR, the British Lung Foundation, Unitaid, and GSK for unrelated academic clinical trials and programme funding, and has received personal payments from AZ for delivering CPD training courses. Mark Gabbay has received consultancy fees from Spectrum Learning and Development as a board member for substance misuse training courses. Tony Kendrick was a member of the NHS England Quality Outcomes Framework (QOF) Advisory Committee 2009-2014, NICE Quality Indicators Advisory Committee 2015-2018, and NICE Depression Guideline Update Committee 2015-2022, and has been a member of the NHS England Improving Access to Psychological Therapies Expert Advisory Committee since 2020. Christopher Dowrick chaired the WONCA Working Party on Mental Health 2016-2021. Guiqing Lily Yao is a member of the NICE Public Health Committee. Glyn Lewis is a member of the NIHR EME Funding Committee. Gareth Griffiths is Director of the Southampton Clinical trials Unit which is part-funded by the NIHR. Beth Stuart is a member of the HTA Commissioning Committee. Tony Kendrick, Glyn Lewis, Michael Moore, Gareth Griffiths, Guiqing Lily Yao, and Beth Stuart have been members of trial steering committees for other NIHR funded studies. All the other authors declared no relevant interests.

Funding details:

This project was funded by the National Institute for Health and Social Care Research (NIHR) Health Technology Assessment programme (grant reference no. HTA 17/42/02). The views expressed in this publication are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Study registration:

International Standard Randomised Controlled Trial Number (ISRCTN) no: 17299295.

ABSTRACT

Background: Outcome monitoring of depression is recommended but lacks evidence of patient benefit in primary care.
Aim: To test monitoring depression using the PHQ-9 questionnaire with patient feedback.
Design and setting: Open cluster-randomised controlled trial in 141 group practices.
Method: Adults with new depressive episodes were recruited through records searches and opportunistically. Exclusion criteria: dementia, psychosis, substance misuse, suicide risk. The PHQ-9 questionnaire was to be administered soon after diagnosis, and 10-35 days later. Primary outcome: Beck Depression Inventory (BDI-II) score at 12 weeks. Secondary outcomes: BDI-II at 26 weeks; Work and Social Adjustment Scale and EuroQol EQ-5D-5L quality of life at 12 and 26 weeks; antidepressant treatment, mental health service use, adverse events, and Medical Informant Satisfaction Scale over 26 weeks.
Results: 302 intervention arm patients were recruited and 227 controls. At 12 weeks 252 (83.4%) and 195 (85.9%) were followed-up respectively. Only 41% of intervention arm patients had a GP follow-up PHQ-9 recorded. There was no significant difference in BDI-II score at 12 weeks (mean difference -0.46; 95% CI -2.16,1.26), adjusted for baseline depression, baseline anxiety, sociodemographic factors, and clustering by practice). EQ-5D-5L quality of life scores were higher in the intervention arm at 26 weeks (adjusted mean difference 0.053; 95% CI 0.093,0.013). A clinically significant difference in depression at 26 weeks could not be ruled out. No significant differences were found in social functioning, adverse events, or satisfaction. In a per-protocol analysis, antidepressant use and mental health contacts were significantly greater in intervention arm patients with a recorded follow-up PHQ-9.
Conclusions: No evidence was found of improved depression outcome at 12 weeks from monitoring. The findings of possible benefits over 26 weeks warrant replication, investigating possible mechanisms, preferably with automated delivery of monitoring and more instructive feedback.

Key words: Primary Health Care; Mental Health; Mood Disorders; Depression; Patient Reported Outcome Measures.

HOW THIS FITS IN

- Follow-up monitoring of people with depression, using patient-reported outcome measures, is recommended but lacks evidence of benefit in primary care
- Monitoring patients' progress with the PHQ-9 produced no benefit in terms of depressive symptoms at 12 weeks follow-up, but at 26 weeks a significant difference in depression could not be ruled out, and patients' quality of life was significantly improved.
- Only 41% of intervention arm patients had a follow-up PHQ-9 recorded in the GP records
- Further research should test PROMS which measure anxiety as well as depression, are automatically delivered and integrated into patients' records, and produce specific treatment recommendations

INTRODUCTION

Guidelines on the management of depression in adults recommend practitioners consider using validated patient-reported outcome measures (PROMs) to inform treatment at diagnosis and follow-up of people with depression [1-4], but there is insufficient evidence that they improve depression management and outcomes for patients in primary care [5,6].

Relatively few studies of PROMs for depression have been conducted in primary care, and there is almost a complete lack of evidence on important outcomes including social functioning, patient satisfaction, quality of life, cost-effectiveness, and possible adverse effects [5,6].

The aim of the study was to answer the research question: What is the effectiveness of assessing primary care patients with depression or low mood after diagnosis and again at follow-up 10-35 days later, using the PHQ-9 questionnaire as a PROM, giving practitioners guidance on assessment and feedback to patients on their progress?

METHODS

Design and setting:

Parallel group open cluster-randomised superiority trial set in 141 group general practices in England and Wales. A cluster-randomised design was chosen on the basis of a significant risk of contamination between arms identified through qualitative interviews with GPs in a prior feasibility trial, i.e. that it would be difficult to forget and avoid using the PHQ-9 questions when treating a control patient in an individually randomised trial.

Randomisation:

Carried out remotely by a Clinical Trials Unit statistician using computerised sequencing, with minimisation by recruiting centres; small/large practices (dichotomised around 8,000 patients); and Local Authority urban/rural locations.

Inclusion and exclusion criteria

Adults with new episodes of depression were recruited mainly through frequent practice records searches, but also opportunistically in consultations. Exclusion criteria were existing treatment for depression; dementia; psychosis; substance misuse; or suicide risk.

Intervention

The PHQ-9 questionnaire[7] was administered by a researcher as soon as possible after recruitment (within 2 weeks), and the GP was asked to repeat the PHQ-9 at a follow-up consultation 10-35 days later. Patients were given written feedback on their PHQ-9 scores and potential treatments to discuss with their GPs. The GPs were given two hours online training in interpreting PHQ-9 scores and taking them into account in management (Appendix 1). They were tested on their understanding of the trial processes together with the strengths and limitations of the PHQ-9, and how it might be used in practice (Appendix 2). Use of the PHQ-9 in practice was 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 (Appendix 3).

Outcomes

The primary outcome was depression on the Beck Depression Inventory (BDI-II)[8] at 12 weeks. Secondary outcomes at 26 weeks were BDI-II scores; social functioning (Work and Social Adjustment Scale (WSAS)[9]); quality of life (EuroQol EQ-5D-5L[10]); patient satisfaction (Medical Informant Satisfaction Scale (MISS)[11]); antidepressant treatment; mental health and social service contacts; and adverse events.

Blinding

Blinding of participants to allocation was impossible given the pragmatic cluster randomised design, but self-report outcome measures were used to prevent observer bias, and analysis was blind to allocation.

Sample size calculation

We assumed a baseline mean BDI-II score of 24.0 with a standard deviation (SD) of 10.0 (derived from a feasibility trial[12]), and mean scores of 14.0 and 17.0 at 12 weeks in the intervention and control groups respectively. An effect size of 0.3SDs represented the minimum clinically important difference (MCID) on the BDI-II[13]. At 5% significance, for 90% power, we needed 235 patients analysed per group. We assumed an intracluster correlation coefficient (ICC) of 0.03 (from the feasibility trial[12]) and mean cluster size of six which gave a design effect of 1.15, giving 270 per group. With 20% loss to follow-up, the target was 676 patients recruited from 113 practices.

We subsequently (10th June 2021) revised the target on finding a correlation of greater than p=0.5 between baseline and follow-up for the primary outcome, meaning we needed only 222 patients analysed per group and total target of 554 recruited.

Analysis

A detailed Statistical Analysis Plan was drawn up prior to analysis of the results (Appendix 4). Differences between arms in depressive symptoms, social functioning, and quality of life at 12 and 26 weeks follow-up were analysed using linear mixed models, adjusting for baseline values; baseline anxiety (measured using the GAD-7[14]; sociodemographic factors, past history of depression, and clustering including a random effect for practice. Patient satisfaction was compared between arms over the 26 week period.

Differences between arms in the process of care for depression were also analysed from practice medical record data over 26 week including PHQ-9s recorded; antidepressant prescribing; and mental health and social service contacts.

Suicide risk

If patients scored other than 0 on suicide/self-harm questions on the BDI-II or PHQ-9 at screening, baseline or follow-up, or indicated suicidal ideas in other ways, a standard operating procedure (SOP) was implemented, requiring further assessment using the p4 suicide risk assessment[15]. Based on the patient's responses, the risk of suicide was categorised as minimal, lower, or higher, and the GP was informed immediately. Care of all patients remained the responsibility of participating GPs as in usual practice.

Ethics approval

The study was approved by the West of Scotland NHS Research Ethics Committee 5, on 21st September 2018 (ref: 18/WS/0144).

More information on the methods can be found in the published protocol[16].

RESULTS

Recruitment of practices

We aimed to recruit 113 practices between November 2018 and August 2019. Due to slow recruitment of patients, we had to continue much longer than planned (made worse by the Covid-19 pandemic), eventually reached a total of 189 by December 2021. Then 48 withdrew before recruiting patients (24 in each arm), leaving 141: 72 intervention and 69 control arm practices. Minimisation ensured practice characteristics were balanced by arm (Table 1).

Recruitment of patients

Of 11,468 patients approached in consultations or through mailed invitations, 1,058 (9.2%) returned reply slips; 574 (10.6%) in the intervention arm and 484 (8.0%) in the control (Figure 1). After exclusion of patients declining participation, ineligible, or uncontactable, 529 were assessed at baseline: 302 (5.5%) in the intervention arm and 227 (3.8%) in the control, between January 2019 and March 2022. The ratio of intervention to control arm patients was therefore 1.3 to 1.

Follow-up of patients

Of 529 patients recruited, 453 (85.6%) were followed up at 12 weeks: 254 intervention arm (84.1%) and 199 controls (87.7%). At 26 weeks 414 (78.3%) were followed-up: 230 intervention arm (76.2%) and 184 controls (81.1%) (Figure 1). Medical records data were collected for 259 intervention arm patients (85.8%) and 201 controls (88.5%).

Baseline characteristics

Mean baseline BDI-II score was higher in the intervention arm (24.1 (SD 8.89) compared to 22.4 (9.52) in the control (Table 2). Baseline anxiety and quality of life were also worse in the intervention arm. Control arm patients were more likely to have had 2+ previous depressive episodes. Sociodemographic characteristics were relatively well balanced, apart from control arm patients being more likely to have no dependents (Table 2).

Primary outcome

At 12 weeks follow-up the mean BDI-II score was 18.5 (SD 10.2) in the intervention arm and 16.9 (10.3) in the control (Table 3 and Figure 2). The adjusted mean score was slightly lower in the intervention arm, but not statistically significant (adjusted mean difference -0.46; 95% confidence interval (CI) -2.16 to 1.26; p=0.60).

At 26 weeks both groups improved further on the BDI-II (Table 3 and Figure 2). The score was slightly lower in the intervention arm, but not significantly (adjusted mean difference -1.63; 95% CI -3.48 to 0.21; p=0.08). The 95% CI included a difference favouring the intervention by more than 3.0 points on the BDI-II so we could not exclude a clinically important difference in depression at 26 weeks.

As a sensitivity analysis, we re-analysed the primary outcome using a multiple imputation model including the baseline value, clustering by practice and all covariates included in the model. The inferences at 12 and 26 weeks were unchanged (adjusted mean difference at 12 weeks -0.18; 95% CI -1.82 to 1.45; p=0.83, and at 26 weeks -0.93; -2.69 to 0.83; p=0.30).

Secondary outcomes

A similar pattern was seen for social functioning at 12 and 26 weeks, with scores improving between baseline and 12 weeks, and further by 26 weeks, but no significant difference between arms (Table 3).

Quality of life improved in both arms between baseline and 12 weeks, then improved further in the intervention arm, but went down slightly in the control (Table 3 and Figure 3). The difference between arms was not statistically significant at 12 weeks but was significant at 26 weeks, favouring the intervention (adjusted mean difference 0.053; 95% CI 0.013 to 0.093; p=0.01).

Patient ratings in the two arms were similar at baseline on the EQ-5D-5L subscales for mobility, self-care, and pain/discomfort, and remained so (Supplementary table). Slightly more intervention arm patients declared severe or extreme problems for anxiety/depression at baseline (23.5% versus 19.5%). At 26 weeks follow-up the proportions declaring no problem with anxiety/depression were 22.6% in the intervention arm versus 13.5% in the control. Improvement in the anxiety/depression dimension therefore explained the overall greater improvement in scores in the intervention arm.

Total scores for satisfaction with care looking back over 26 weeks were very similar between arms (Table 3). The same was found for all four satisfaction subscales (Distress-relief, Communication-comfort, Rapport, and Compliance-intent).

Post-hoc analysis of 50% improvement, and remission, at 26 weeks

We conducted a post-hoc analysis of categorical improvements in BDI-II scores at 26 weeks, to further investigate differences in depression, given the wide confidence intervals around the mean difference, and because we found the difference in proportions of patients reporting no anxiety/depression on the EQ-5D-5L at 26 weeks (Supplementary Table). We compared the proportions in each arm who improved by 50% or more on the BDI-II, and the proportions who scored above 13 at baseline (the threshold for `caseness') and subsequently remitted to 13 or less by 26 weeks.

The proportions of patients improving by 50% or more were not significantly different (102/226 intervention (45.1%) versus 69/185 (37.3%) controls, OR 1.53; 95% CI 0.92 to 2.56; p=0.10), but the proportion of patients remitting in the intervention arm was significantly greater (100/201 (49.8%) versus 59/148 (39.9%); OR 2.18; 95%CI 1.12 to 4.24; p=0.02).

Post-hoc per-protocol analysis of depression outcome

In the intervention arm 190 patients (73.4%) had PHQ-9s recorded in the medical record, and in the control arm 35 patients (17.4%). However, around half of those recorded were the baseline PHQ-9s carried out by the researchers. Only 124 patients had recorded PHQ-9s carried out by their GPs during follow-up: 106 in the intervention arm (40.9%) and 18 in the control arm (8.9%).

Post-hoc, we defined GP compliance with the protocol in the intervention arm as carrying out and recording a follow-up PHQ-9, and in the control arm with not carrying out and recording one. On that basis we carried out a post-hoc per-protocol analysis of depression outcome for the 106 intervention participants with a recorded follow up PHQ-9, compared to that for the 209 control participants without a recorded follow-up PHQ-9. At 12 weeks the fully adjusted difference in BDI-II score was -1.57 points (95% CI -3.47, 0.35; p=0.108) and at 26 weeks -1.08 (95% CI -3.40, 1.24; p=0.361), so there were no significant differences in depression symptom counts at either point.

Use of antidepressants

Medical records data were obtained for 258 intervention arm patients (85.4%) and 201 controls (88.5%). Of these 174 (67.4%) and 112 (55.7%) respectively had antidepressant prescriptions recorded over 26 weeks, but the difference between arms was not significant, (odds ratio (OR) 1.83; 95% CI 0.96 to 3.48; p=0.07, adjusted for baseline depression, baseline anxiety, baseline antidepressant use, sociodemographics, and practice).

In a post-hoc per-protocol analysis, we found that, of the 106 intervention arm patients with a recorded follow up PHQ-9, 71 (67.0%) received a prescription for antidepressants, compared to 102/183 (55.7%) of the 209 controls with no recorded follow-up PHQ-9. The adjusted OR was 2.80 (95% CI 1.14, 6.88; p=0.025), showing significantly more antidepressant prescribing in those with a recorded follow up PHQ-9 administered by the GP.

Contact with mental health and social services

In their records 90 intervention arm patients (34.6%) and 68 controls (33.8%) had contacts over 26 weeks with mental health and social services (mental health nurse, counsellor, psychologist, psychiatrist, and social workers): not significantly different between arms (adjusted OR 1.37; 95% CI 0.71 to 2.63; p=0.342).

In a post-hoc per-protocol analysis, 48 (45.3%) of the 106 intervention arm patients with a recorded follow-up PHQ-9 had a mental health service contact, compared to 57/183 (31.2%) of the controls with no follow-up PHQ-9. The adjusted OR was 3.96 (95% CI 1.38, 11.34; p=0.010) showing significantly more mental health contacts for those with a recorded follow up PHQ-9 administered by the GP.

Adverse events

There were two serious adverse events. One control arm patient reported suicidal ideas, was assessed by the trial principal investigator and found to be at higher risk, and the GP was informed immediately. The patient was referred to a community mental health team (CMHT) for immediate assessment and withdrawn from the study. One intervention arm patient was hospitalised with Covid-19 and ketoacidosis: a severe event, but not related to the trial.

The suicidal ideation SOP was triggered 318 times, 180 times for intervention arm patients, and 138 times for controls, in proportion to patient numbers in each arm. Altogether, 267 (146 intervention, 121 control) were rated 'minimal risk', 38 (25 intervention, 13 control) 'lower risk', and 13 (nine intervention, four control) 'higher risk'. In four cases (two intervention and two control) participants were withdrawn from the study.

DISCUSSION

Summary

We found no significant difference between intervention and control arms in the primary outcome, depression on the BDI-II at 12 weeks. However it was not possible to rule out a clinically significant benefit at 26 weeks, given the upper limit of the 95% confidence interval included the MCID of 3.0 points [13]. We found evidence of benefit in a categorical analysis of remission of depression to a BDI-II less than 13 at 26 weeks, but this was a post-hoc analysis, and there was no significant difference in a similar analysis of the proportions of patients with a 50% improvement in depression.

There were no significant differences found in social functioning and satisfaction with care, although the differences found tended to favour the intervention. Quality of life scores were however significantly higher in the intervention arm at 26 weeks. The better quality of life score was due to a greater proportion of intervention arm patients reporting no anxiety/depression. We did not measure anxiety symptoms specifically at follow-up, but it may be that some patients were reassured to see their depression was improving, and therefore felt less anxious.

Overall, more intervention arm patients had recorded antidepressant prescriptions than controls over 26 weeks (67.4% versus 55.7%), but this difference was not statistically significant. There was no overall difference in mental health and social service contacts either, a third of patients in both arms having at least one. However, in post-hoc per-protocol analyses including only those intervention arm patients who had follow-up PHQ-9s administered and recorded by their GPs, there was significantly greater antidepressant prescribing and contact with mental health services than among controls with no follow-up monitoring.

Strengths and limitations

A strength of the study is that its design was informed by a feasibility trial [12], which led to choosing the cluster design, avoiding contamination between arms in applying the intervention, and optimising adherence to study procedures in practices. However, a cluster randomised design increases the risk of selection bias among practitioners deciding whether or not to approach patients opportunistically in consultations. More than twice as many intervention arm than control arm patients were recruited opportunistically, and overall the ratio of patients randomised was 1.3 to 1, which may have reflected lower motivation to take part on the part of control arm patients, who were offered only usual care. Selection bias may explain higher baseline depression and anxiety scores, and lower quality of life, in the intervention arm, although the two arms were relatively well balanced in terms of patient demographics, and analyses were adjusted for baseline differences.

Participating practitioners were trained in both the use of the PHQ-9 and treatment choices related to severity scores, while taking into account contextual factors. The amount of training was limited to two hours, but was considered an amount feasible to offer at scale.

Recruitment to the trial was very challenging, particularly during the Covid-19 pandemic, when practices had significant extra pressures. We did not quite achieve the revised target of 554 patients, by 25, but the follow-up rate of 85.6% was better than predicted and we gathered primary outcome data on sufficient participants to answer the main research question with precision, so the result for the primary outcome may be regarded as robustly negative. It is possible however that there was a difference in depression at 26 weeks, which was missed due to lacking power at that point.

It was not possible to blind participants and researchers given the pragmatic cluster design, but self-report outcome measures avoided possible observer bias, and the statistical analyses were all carried out blind to allocation.

Delivering the intervention was challenging, and not as it would be in routine practice. Practitioners could not administer the PHQ-9 when patients first presented with depression, because patients had to be given information about the study and at least 24 hours to consider taking part before consenting. This was a requirement of the NHS Research Ethics Committee. To avoid asking the GP to bring the patient back to administer the first PHQ-9, the researcher administered it at baseline assessment instead. Treatment could therefore have started at the initial consultation before the baseline score could be taken into account.

As only 73% of intervention arm patients had PHQ-9s recorded in their records, the GPs obviously did not record their scores routinely, since we know 100% had PHQ-9s administered by the researchers at baseline, and these were all communicated to the practices. Only 41% of intervention arm patients had follow-up, GP-administered PHQ-9s recorded, although the actual numbers of follow-up PHQ-9s carried out may well have been higher. We asked the intervention arm GPs to administer follow-up PHQ-9s with all their participating patients, but did not insist that they recorded the follow-up PHQ-9s, which is a limitation of the study. Effectively we tested instituting a policy of monitoring using the PHQ-9 which we knew would not necessarily be carried out per protocol, which would likely be the case to a greater extent in routine practice.

We did not have the resources to collect detailed information in real-time of individual GPs' patient treatment plans and whether they were changed following PHQ-9 assessment at follow-up. However, the post-hoc per-protocol analyses, indicating that significantly increased antidepressant prescribing and mental health service contacts were associated with carrying out and recording follow-up PHQ-9s, suggested that the GPs may have increased antidepressant treatment and referrals to specialist services on finding less than desired improvements in scores at follow-up.

A smaller proportion (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 them. These may have been administered outside practices in psychology services, or by temporary practitioners within practices. However, this was a relatively low level of use, so there was good differentiation between the arms, and the pr-specified analyses were conducted on an intention to treat basis.

There were relatively few exclusion criteria, tending to increase the heterogeneity of the sample and generalisability of the findings. The computer codes used to identify patients through the records searches included symptom codes (e.g. `low mood') in addition to specific diagnoses (e.g. 'depressive disorder'), to avoid missing patients not given a specific diagnosis.

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, only 5.5% of those approached in the intervention arm, and 3.8% in the control.

Comparison with previous literature

The findings are consistent with previous trials which have mostly shown no benefit for depression outcome. Only one trial found a reduction in depression[17], but no changes in management to

explain the benefit[18]. Two others found changes in management but not outcomes[19,20]. The most recent found no difference in depression, but reduced anxiety at 8 weeks, and improved functioning at 24 weeks follow-up[21].

Evidence of benefit from PROMs has been found in psychological therapy settings including improved outcome[22], and making therapy more efficient[23]. However in psychological services PROMs are given multiple times during therapy, and facilitate adjustment of treatment. With only 1-2 PHQ-9s given in our study, the information available to GPs was more restricted. PROMs used in monitoring progress in psychological services are also more extensive than the PHQ-9 alone, and the implications of results for therapy are discussed with supervisors between sessions[22, 23]. Finally, psychological services offer a range of evidence-based treatments, whereas GP treatment is largely antidepressants alone, and may be less effective in changing depression outcome[1].

Implications for practice and research

The absence of evidence for improvement in the outcome of depression from studies of follow-up monitoring with PROMs in primary care suggests that guidelines that recommend their use[1-4] should continue to make them discretionary rather than mandatory, at least outside psychological therapy settings, where there is good evidence of benefit. Monitoring patients who like to see improvement in their scores is justifiable, as it may improve their quality of life. Our post-hoc analyses also suggest that conducting and recording follow-up PHQ-9s may lead to greater antidepressant prescribing and referrals to mental health services. However their use is not without cost, in terms of the time taken, even though they are relatively cheap. The cost-effectiveness of using the PHQ-9 in this study will be reported separately.

In addition, continuing to recommend outcome monitoring with PROMs may be justified on the basis of providing greater transparency to health service funders and the public about the management of depression and patients' responses to particular treatments.

Future research on depression monitoring in primary care should improve the delivery of monitoring and test PROMs which cover anxiety and social functioning as well as depression. PROMS should be completed remotely between consultations; facilitated by automated analysis and feedback of the results to practitioners and patients; and deliver specific recommendations for treatment. Practitioners interpreting PROM results will still need to consider the circumstances surrounding individuals' histories of depression, and response to treatments.

ACKNOWLEDGEMENTS

Tony Kendrick, Christopher Dowrick, Glyn Lewis, Michael Moore, Geraldine Leydon, Adam Geraghty, Gareth Griffiths, Shihua Zhu, Guiqing Lily Yao, Carl May, Rachel Dewar-Haggart, Samantha Williams, and Beth Stuart wrote the funding application. Mark Gabbay replaced Christopher Dowrick on his retirement. Rachel Dewar-Haggart, Samantha Williams, Lien Bui, Natalie Thompson, 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 Lauren Bridewell collected and entered data. Beth Stuart, Guiqing Lily Yao and Shihua Zhu analysed the quantitative data. Rachel Dewar-Haggart and Brian Chi Fung Ching led on the analysis of the qualitative data. All authors were involved in drafting the work or revising it critically, gave final approval for the version to be published, and agree to be accountable for all aspects of the work related to its accuracy or integrity.

Tony Kendrick, Michael Moore, Adam Geraghty, and Beth Stuart have been supported by the NIHR School for Primary Care research (SPCR). Mark Gabbay is supported by the NIHR Applied Research Collaboration North West Coast (ARC NWC). Glyn Lewis is supported by the University College London Hospital Biomedical Research Centre (UCLH BRC).

Professor Bernd Löwe of the University of Hamburg was an international advisor.

Sophie Johnson, Ben Hammond, and Charlotte Hookham in Southampton; Daniel Lawrence and Taylor Hooper in Liverpool; and Paula Beharry and Thomas Gant at UCL; provided administrative support.

Tammy Thomas and Louisa Little provided oversight from the Southampton Clinical Trials Unit.

Claire Winch, Debbie Kelly, Nicola Blakey, Heather Kenyon and Nnenna Omeje, of Thames Valley and South Midlands Clinical Research Network, provided assistance with data collection.

Thanks to the Trial Steering committee (TSC) members who were: Michael Barkham (academic psychologist and Chair); Susan Collinson (PPI representative); Laura Gray (statistician); Stavros Petrou (health economist); and Linda Gask (academic psychiatrist).

Thanks also to the trial Independent Data Monitoring Committee (IDMC) members who were Richard Byng (academic GP and Chair), Jill Mollison (statistician), and Jaime Delgadillo (academic psychologist).

Most of all, many thanks to the participating patients and practitioners of the following 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 & Liss Surgery, Longrove Surgery, Mann Cottage Surgery, Marine Lake Medical Practice, Meddygfa Bronyffynnon Surgery, Milman Road Health Centre, Mitcham Family Practice , Mulberry & 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 & 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.

Patient and public involvement

We recruited two mental health service users Margaret Bell and Bryan Palmer to join the study team, very early on, at the point of designing the study and applying for funding. They 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.

Bryan 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, Bryan helped the CI 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.

They were both very active members of the study group, attending study team meetings and commenting on relevant documents through email throughout the four 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 intervention arm patients on the meaning of their PHQ-9 scores and possible treatments related to the level of severity of their depression. They also reviewed the semi-structured 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. We gave our PPI colleagues regular feedback on our interactions with them and asked for theirs, both of which were very positive.

REFERENCES

[1] National Institute for Health and Care Excellence. Depression in adults: treatment and management (update). Published: 29 June 2022 https://www.nice.org.uk/guidance/ng222/chapter/Recommendations#recognition-and-assessment (Accessed 31 May 2023)

[2] US Federal Health Resources and Services Administration. Uniform Data System Clinical Quality Measures 2020. https://bphc.hrsa.gov/sites/default/files/bphc/data-reporting/2020-clinical-measures-handout.pdf (Accessed 31 May 2023)

[3] Kaiser Permanente Health Maintenance Organisation. Mental Health Monitoring tool. https://wa-provider.kaiserpermanente.org/static/pdf/provider/patient-ed/screenings/bhi-monitoring-tool.pdf (Accessed 31 May 2023)

[4] Nederlands Huisartsen Genootschap (Dutch Society of General Practitioners). NHG Guideline on Depression. https://richtlijnen.nhg.org/standaarden/depressie#volledige-tekst-3-beoordelen-van-de-ernst-van-de-depressieve-stoornis (Accessed 31 May 2023)

[5] Kendrick T, El-Gohary M, Stuart B, *et al.* Routine use of patient-reported outcome measures (PROMs) for improving treatment of common mental health disorders in adults. Cochrane Database of Systematic Reviews 2016, Issue 7. Art. No.: CD011119. DOI: 10.1002/14651858.CD011119.pub2. https://www.cochrane.org/CD011119/DEPRESSN_using-patient-reported-outcome-measures-monitor-progress-among-adults-common-mental-health-disorders

[6] 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

[7] 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.

[8] Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. Psychological Corporation, San Antonio, Texas, 1996.

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

[10] EuroQol Research Foundation. EQ-5D-5L. EuroQol Office, Rotterdam. https://euroqol.org/EQ-5D-5L-instruments/EQ-5D-5L-about/

[11] Meakin R, Weinman J. The 'Medical Interview Satisfaction Scale' (MISS-21) adapted for British general practice. Fam Pract 2002;19(3):257-63.

[12] Kendrick T, Stuart B, Leydon GM, et al. Patient-reported outcome measures for monitoring primary care patients with depression: PROMDEP feasibility randomised trial. BMJ Open 2017;7(3):1–11. [e015266]. https://doi.org/10.1136/bmjopen-2016-015266.

[13] Button KS, Kounali D, Thomas L, et al. Minimal clinically important difference on the Beck Depression Inventory – II according to the patient's perspective. Psychol Med 2015;45:3269–3279 doi:10.1017/S0033291715001270.

[14] Spitzer RL, Kroenke K, Williams JB, *et al*; A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166:1092-7.

[15] 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

[16] Kendrick, T., Moore, M., Leydon, 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

[17] Yeung AS, Jing Y, Brenneman SK, *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. doi: 10.1002/da.21983

[18] Chang TE, Jing Y, Yeung AS, *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(2):105–112.

[19] 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. doi: 10.1007/BF02600303

[20] Wikberg C, Westman J, Petersson EL, *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 Practice 2017;18:6 DOI 10.1186/s12875-016-0578-9.

[21] Browning M, Bilderbeck AC, Dias R, 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(7):1307-1314. doi: 10.1038/s41386-021-00981-z.

[22] Delgadillo J, de Jong K, Lucock M, *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. doi: 10.1016/S2215-0366(18)30162-7.

[23] Delgadillo J, Overend K, Lucock M, *et al.* Improving the efficiency of psychological treatment using outcome feedback technology. Behav Res Ther 2017;99:89-97. doi: 10.1016/j.brat.2017.09.011

Table 1 Cluster level (participating general practice) characteristics at baseline

Characteristic - N	Intervention	Control	Total
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	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			
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 (%) (Self-declared gender)	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%)
- 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%)

Characteristic	Intervention (n=302)	Control (n=227)	Total (n=529)
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 & 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 dependents 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%)

CSE is the Certificate of Secondary Education, a qualification in a specific subject formerly taken by school students aged 14–16, at a level below O (Ordinary) level. Both the CSE and O level were replaced in 1988 by the GCSE, or General Certificate of Secondary Education. NVQ Level 1 is the first level National Vocational Qualification, a work-based job-specific qualification. A Level is the Advanced secondary education qualification in a specific subject taken by school students aged 17-19. BTEC is the Business and Technology Education Council certificate work-based vocational qualification taken after secondary school above the age of 16. HNC (Higher National Certificate), HND (Higher National Diploma), and City & Guilds are more advanced vocational qualifications.

Table 3 Primary and secondary outcomes at baseline, 12 weeks, and 26 weeks follow-up

	Baseline		12 weeks			26 weeks		
	N	**Mean Score (SD)**	**N**	**Mean score (SD)**	**Mean adjusted difference* (95% CI); p-value**	**N**	**Mean score (SD)**	**Mean adjusted difference* (95% CI); p-value**
Depression (BDI-II score)								
Intervention	302	24·1 (8·96)	252	18·5 (10·17)	-0·46 (-2·16 to 1·26); p=0·602	226	15·1 (10·84)	-1·63 (-3·48 to 0·21); p=0·082
Control	227	22·4 (9·52)	195	16·9 (10·30)	REF	184	14·7 (10·65)	REF
Social functioning (WSAS score)								
Intervention	302	17·3 (9·94)	237	14·7 (9·54)	0·48 (-1·03 to 2·00); p=0·531	212	11·6 (9·59)	1·34 (-3·20 to 0·53); p=0·160
Control	227	16·6 (10·06)	195	13·2 (9·90)	REF	183	12·0 (9·99)	REF
Quality of Life (EQ-5D-5L score)								
Intervention	302	0·659 (0·232)	256	0·694 (0·236)	-0·002 (-0·0412 to 0·0372) p=0·94	221	0·718 (0·249)	0·053 (0·013 to 0·093); p=0·01
Control	226[#]	0·667 (0·226)	197	0·708 (0·213)	REF	183	0·696 (0·225)	REF
Satisfaction with care (MISS total score)								
Intervention	302	N/A		N/A		217	121·8 (27·37)	5·39 (-1·39 to 12·16); p=0·119
Control	227	N/A		N/A		176	116·0 (26·75)	REF

*Adjusted for baseline value, baseline anxiety (GAD-7 score), sociodemographics, past history of depression, and practice as a random effect· REF = reference value

Figure 1 CONSORT Diagram

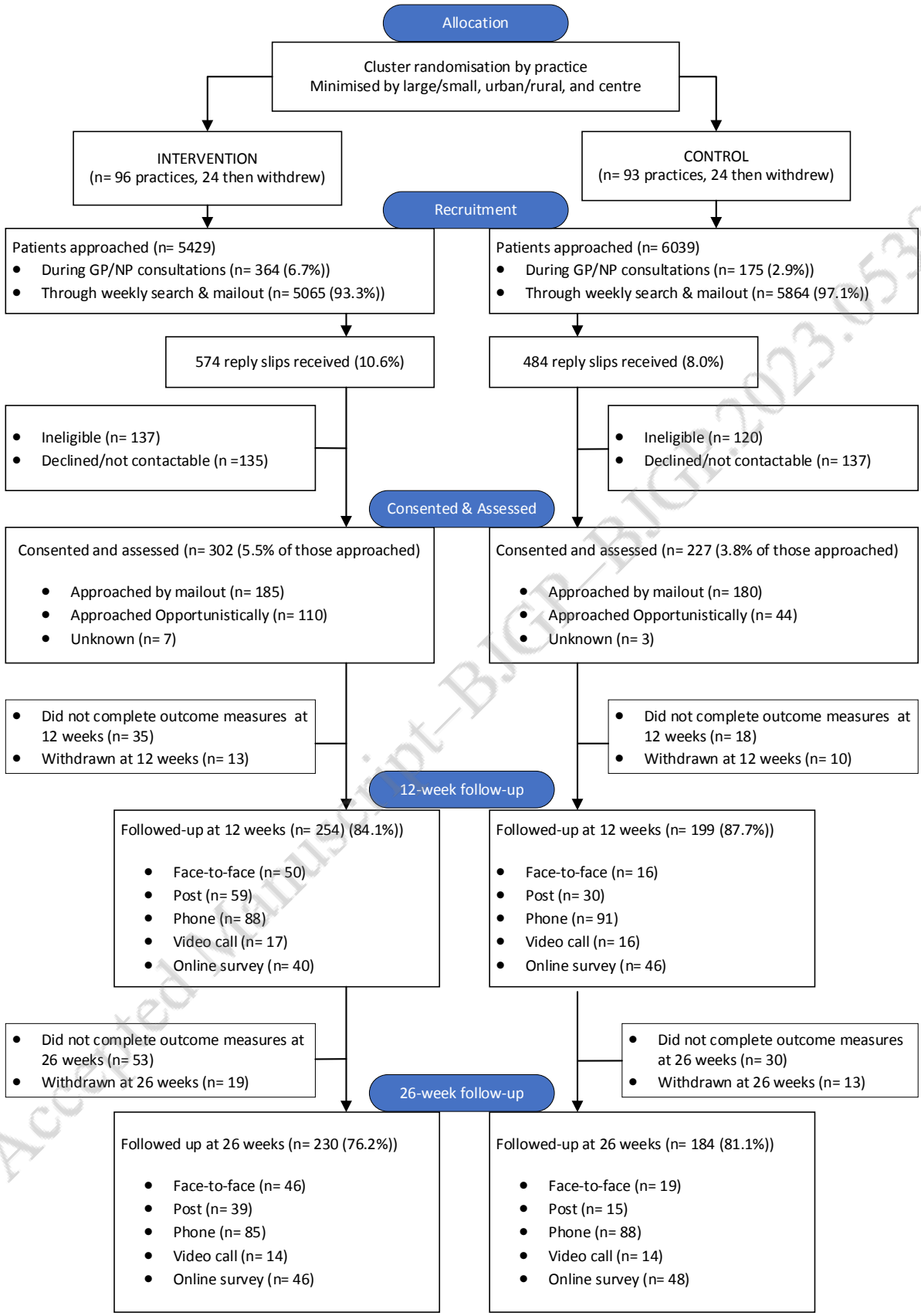

Figure 2 Mean BDI-II depression scores at baseline, 12 weeks, and 26 weeks follow-up