

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 Hunt¹ and Beth Stuart⁷

¹School of Primary Care, Population Health and Medical Education, University of Southampton, Southampton, UK

²Department of Primary Care and Mental Health, Institute of Population Health, University of Liverpool, Liverpool, UK

³Division of Psychiatry, Faculty of Brain Sciences, University College London, London, UK

⁴Southampton Clinical Trials Unit, University of Southampton and University Hospital Southampton NHS Foundation Trust, Southampton, UK

⁵Leicester Clinical Trials Unit, University of Leicester, Leicester, UK

⁶Faculty of Public Health and Policy, London School of Hygiene and Tropical Medicine, London, UK

⁷Centre 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

Disclosure of interests

Full disclosure of interests: Completed ICMJE forms for all authors, including all related interests, are available in the toolkit on the NIHR Journals Library report publication page at <https://doi.org/10.3310/PLRQ4216>.

Primary conflicts of interest: Tony Kendrick, Christopher Dowrick, Glyn Lewis, Michael Moore, Geraldine Leydon, Adam WA 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 Care Research (NIHR) to carry out this study and other research. In addition, Glyn Lewis has received grant funding from the Medical Research Council and Wellcome. Gareth Griffiths has received funding from Janssen-Cilag (High Wycombe, UK), AstraZeneca (Cambridge, UK), Novartis (Basel, Switzerland), Astex (Cambridge, UK), Roche (Basel, Switzerland), HeartFlow (Mountain View, CA, USA), Celldex (Hampton, NJ, USA), BMS, BioNTech (Mainz, Germany), Cancer Research UK (London, UK), the NIHR, the British Lung Foundation (London, UK), Unitaid (Geneva, Switzerland) and GSK (Brentford, UK) for unrelated academic clinical trials and programme funding and has received personal payments from AstraZeneca for delivering Continuing Professional Development training courses. Mark Gabbay has received consultancy fees from Spectrum Learning and Development (Wakefield, UK) as a board member for substance misuse training courses. Tony Kendrick was a member of the NHS England Quality Outcomes Framework Advisory Committee 2009–14, NICE Quality Indicators Advisory Committee 2015–8 and NICE Depression Guideline Update Committee 2015–22 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 2016–21. Guiqing Lily Yao is a member of the National Institute for Health and Care Excellence Public Health Committee. Glyn Lewis is a member of the NIHR Efficacy and Mechanism Evaluation 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 NIHR Health Technology Assessment 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.

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

Published March 2024
DOI: 10.3310/PLRQ4216

Scientific summary

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

Health Technology Assessment 2024; Vol. 28: No. 17
DOI: 10.3310/PLRQ4216

NIHR Journals Library www.journalslibrary.nihr.ac.uk

Scientific summary

Some text in this chapter has been adapted from the study protocol published as: 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>. This article is published under license 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.

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 $270 \times 2/0.8$ 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 $\geq 50\%$ 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 $-\pounds 163$, 95% confidence interval $-\pounds 349$ to $\pounds 28$). 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 $\pounds 20,000$ and $\pounds 30,000$ 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

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.

Health Technology Assessment

ISSN 2046-4924 (Online)

Impact factor: 3.6

A list of Journals Library editors can be found on the [NIHR Journals Library website](#)

Launched in 1997, *Health Technology Assessment* (HTA) has an impact factor of 3.6 and is ranked 32nd (out of 105 titles) in the 'Health Care Sciences & Services' category of the Clarivate 2022 Journal Citation Reports (Science Edition). It is also indexed by MEDLINE, CINAHL (EBSCO Information Services, Ipswich, MA, USA), Embase (Elsevier, Amsterdam, the Netherlands), NCBI Bookshelf, DOAJ, Europe PMC, the Cochrane Library (John Wiley & Sons, Inc., Hoboken, NJ, USA), INAHTA, the British Nursing Index (ProQuest LLC, Ann Arbor, MI, USA), Ulrichsweb™ (ProQuest LLC, Ann Arbor, MI, USA) and the Science Citation Index Expanded™ (Clarivate™, Philadelphia, PA, USA).

This journal is a member of and subscribes to the principles of the Committee on Publication Ethics (COPE) (www.publicationethics.org/).

Editorial contact: journals.library@nihr.ac.uk

The full HTA archive is freely available to view online at www.journalslibrary.nihr.ac.uk/hta.

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

The research reported in this issue of the journal was funded by the HTA programme as award number 17/42/02. The contractual start date was in November 2018. The draft report began editorial review in November 2022 and was accepted for publication in July 2023. The authors have been wholly responsible for all data collection, analysis and interpretation, and for writing up their work. The HTA editors and publisher have tried to ensure the accuracy of the authors' manuscript and would like to thank the reviewers for their constructive comments on the draft document. However, they do not accept liability for damages or losses arising from material published in this manuscript.

This manuscript presents independent research funded by the National Institute for Health and Care Research (NIHR). The views and opinions expressed by authors in this publication are those of the authors and do not necessarily reflect those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care. If there are verbatim quotations included in this publication the views and opinions expressed by the interviewees are those of the interviewees and do not necessarily reflect those of the authors, those of the NHS, the NIHR, the HTA programme or the Department of Health and Social Care.

Copyright © 2024 Kendrick *et al.* This work was produced by Kendrick *et al.* under the terms of a commissioning contract issued by the Secretary of State for Health and Social Care. This is an Open Access publication distributed under the terms of the Creative Commons Attribution CC BY 4.0 licence, which permits unrestricted use, distribution, reproduction and adaptation in any medium and for any purpose provided that it is properly attributed. See: <https://creativecommons.org/licenses/by/4.0/>. For attribution the title, original author(s), the publication source – NIHR Journals Library, and the DOI of the publication must be cited.

Published by the NIHR Journals Library (www.journalslibrary.nihr.ac.uk), produced by Newgen Digitalworks Pvt Ltd, Chennai, India (www.newgen.co).

