| **Study**  Setting, sample size, design | **Intervention and Comparator**  PROM, training, feedback, control | **Outcomes**  Measures, time points, precision | **Strengths and limitations**  Power, allocation to arm, risk of bias | **Study summary and conclusions**  Rating of study quality |
| --- | --- | --- | --- | --- |
| **Systematic reviews, meta-analyses** | |  |  |  |
| *Any setting* |  |  |  |  |
| **Kendrick et al 2016**2  Multidisciplinary mental health care, psychological therapy settings, primary care.  N=17 studies (8787 participants)  Systematic review of RCTs only in adults with common mental health disorders (excluding studies with >10% participants with psychoses, dementia or eating disorders) | **PROM:** self-complete or administered measures of:   * depressive symptoms * anxiety symptoms * health-related QoL * composite measures of symptoms, individual functioning, and social functioning   Meta-analysis was only possible for studies using the OQ-45, and ORS, composite measures.  **Comparator:** TAU | No statistically significant difference between feedback and no-feedback groups  as measured on OQ-45 (MD -1.14, 95% CI -3.15 to 0.86, 9 studies) or ORS (SMD -0.07, 95% CI-0.16 to 0.01, 3 studies). No study in these analyses was in a primary care setting.  For the subgroup of ‘not on track (NOT)’ patients, symptom scores were slightly lower in the feedback group compared to the no feedback group (SMD -0.22, 95% CI -0.35 to -0.09, 10 studies).  For the subgroup of 'on track (OT)’ patients, the mean number of treatment sessions was slightly fewer in the feedback group: (MD -0.69, 95% CI -1.10 to -0.29, 4 studies).  Only one study reported any findings in relation to adverse events, with no immediate suicide risk discerned. Adverse events from prescribed medication were not assessed in any of the included studies. | **Strengths:** Well conducted review, including formal risk of bias assessment.  **Limitations:** Evidence from meta-analyses was considered low quality, due to high risk of bias within included studies. | There is insufficient evidence to support the use of routine outcome monitoring using PROMs in the treatment of common mental health disorders, in terms of improving patient outcomes or in improving management.  The findings are subject to considerable uncertainty however, due to the high risk of bias in the large majority of trials meeting the inclusion criteria, which means further research is very likely to have an important impact on the estimate of effect and is likely to change the estimate.  High quality study |
| *Specialist mental health* |  |  |  |  |
| **Knaup et al 2009**4  Community-based mental health services, university  counselling centres, specialist mental health units, inpatient psychotherapeutic clinics  N= 12 studiesa (5458 participants)  Systematic review of controlled studies (including non-randomised) in adults with mental health problems (including schizophrenia, dementia and eating disorders) | **PROM:** feedback, defined as providing mental healthcare professionals and/or patients with individual information on treatment outcome based on standardised measures. Included studies used: CANE, CANSAS, CNS, CSC, FEP, FLZ, GBB, HADS, HAQ, HAS, ICD-10, MANSA, MSPSS, OQ-45, PAE, SEED, TREAT-EAT  **Comparator:** Control condition (for 10/12 studies this was TAU) | Feeding back outcome showed a small, but significant (d = 0.10; 95% CI 0.01–0.19, 10 studies) positive short-term effect on the mental health of individuals that did not prevail in the long run. Subgroup analysis revealed no significant differences regarding feedback modalities. Outcome management did not contribute to a reduction of treatment duration. | **Strength:** Range of sources searched for relevant trials  **Limitations:** Included non -randomised studies. No formal risk of bias assessment. | Feedback of outcome had a small, albeit statistically significant short-term effect on improving mental health outcomes. This effect was found to be consistent across a variety of outcome measures  Moderate quality study |
| **Shimokawa et al 2010**5  University counselling centre, hospital outpatient setting  N= 6 studiesb (6,151 participants)  Meta-analysis of randomised and non-randomised trials performed by study authors in in patients receiving psychological therapy | **PROM:** OQ-45. 3 types of patient progress feedback interventions using OQ-45: progress feedback to therapists (T-Fb), progress feedback to both patients and therapists (P/T-Fb), and clinical support tools in addition to progress feedback (CST T/P-Fb)  **Comparator:** TAU, or any of the above interventions | In patients at risk of leaving treatment worse off than when they started, a clinically significant improvement was statistically significant more likely with T—Fb versus TAU (OR 1.70, 95% CI 1.17 to 2.46, 4 studies), and with CST T/P-Fb versus T-Fb (OR 1.43, 95% CI 1.08 to 2.18, 3 studies), however there was no difference between P/T-Fb versus T-Fb. Furthermore, worsening or deterioration of condition was statistically significant less likely with T-Fb versus TAU (OR 0.62 (95% CI 0.40 to 0.98, 4 studies), however there was no difference between P/T-Fb versus T-Fb or CST-Fb versus T-Fb (3 studies). | **Limitations:** All studies included in the meta-analysis were carried out by the same research team. Reliability of treatment implementation may have been an issue in individual studies because the use of feedback interventions by therapists was not closely controlled or monitored. Two of the included studies were non-randomised trials. | The evidence appears to support the efficacy and effectiveness of feedback interventions in enhancing treatment outcome.  Low quality study |
| **RCTs** |  |  |  |  |
| *Primary care* |  |  |  |  |
| **Mathias et al 1994**9  Primary care group practices, USA.  N = 618 participants  Cluster randomised by practice. | **PROM:** Mental Health Patient Profile, constructed from SCL-90, DIS, and SF-36. PROMs administered by researchers outside the practice and results summarised for practitioners. Feedback to practitioners only.  **Control group:** completed PROM but without feedback of results. | No significant differences between arms in SCL-90 derived global severity of symptoms (GSI 59.99 in intervention arm vs 60.89 in control arm, p = 0.89), or anxiety (HASS 64.72 vs 68.23, p = 0.74) at 12 weeks. No significant difference in changes in psychotropic medications (OR 1.09, 95% CI 0.64 to 1.85) over 5 months. Referrals to mental health specialists over 5 months significantly more likely in feedback group (OR 1.73, 95% CI 1.11 to 2.70). No significant differences on the SF-36. | **Limitations:** Method of allocation to arm not reported in detail (“physicians were randomized “by call group” to intervention or control arm”).  Not possible to blind practitioners and researchers due to the nature of the intervention and cluster randomisation.  The 45 participants (7.3%) lost to follow-up had higher mean scores for the SF-36 than those followed up. | Symptom monitoring with feedback to physicians did not improve outcomes of depression treatment in primary care.  Feedback to physicians did not result in more changes of psychotropic medication but did result in more referrals to mental health specialists.  Moderate quality study. |
| **Yeung et al 2012**10  Primary care group practices, USA  N = 915 participants  Cluster quasi randomised by practice. | **PROM:** PHQ-9 administered by researchers outside the practice and scores faxed to physicians monthly over 6 months, with % change from baseline, interpretation, and possible treatment adjustment.  Feedback to practitioner and patient.  **Control group:** completed PHQ-9 at 3 and 6 months but results not faxed to physician until 6 months. | Statistically significantly greater odds of remission to PHQ-9 <5 (odds ratio [OR], 1.59, 95% CI, 1.07–2.37) and to response (PHQ score reduced by at least 25%) found in feedback group (OR 2.02, 95% CI 1.36 to 3.02). However, proportions without a change in antidepressant therapy did not differ significantly between study arms (200/352 in feedback group vs 115/252 in no feedback group, OR 1.21, 95% CI 0.78 to 1.88; p = 0.06).  No immediate suicide risk found across both feedback and no-feedback groups combined. | **Strength:** Participating physicians were kept blind to their assignment to intervention or control arm and were not informed of the frequency of feedback of assessment results.  **Limitations:** High risk of bias due to alternate assignment of recruited practices to intervention and control arms.  The authors performed per protocol analyses only, and 273/915 (29.9%) of participants were not included. | Symptom monitoring with feedback to physicians improved outcomes of depression treatment, despite a lack of significant changes in physicians’ management of patients’ depression.  Symptom measurement might have increased patients’ awareness and ability to report relevant symptoms, or made them feel more supported, contributing to a lower medication discontinuation rate in the intervention group.  Moderate quality study. |
| **Kendrick et al 2017**12  Primary care group practices,  UK  N = 49 participants  Randomisation partly clustered by group practice, partly by individual patients. | **PROM:** PHQ-9, Distress Thermometer, and PSYCHLOPS problem profile, administered by researchers outside the practice, following diagnosis and 10–35 days later. Practitioner training minimal. Feedback of score to patient by practitioner.  **Control group:** TAU. | At 3 months, mean BDI-II score lower in intervention arm by 5.8 points (95% CI −11.1 to −0.5). No significant differences found in depression (BDI-II), social function (WSAS), quality of life (EQ-5D) or service use (CSRI) over 6 months. At 6 months, patient satisfaction (MISS) score higher in control arm by 22.0 points (95% CI −40.7 to −3.29). | **Limitations:** Small feasibility trial lacking power to detect differences in outcomes. Contamination when individual randomisation used.  Unable to blind participants to allocation, but self-rated assessments used to avoid observer bias.  Follow-up rate 82% in intervention arm but only 72% in control arm. | Possible benefit from monitoring in terms of outcome of depression. Patient satisfaction may be lower due to GPs ignoring the PROM results. A full trial needs better GP training and improved patient follow-up. Cluster randomisation of practices preferable to avoid contamination between arms.  Very low to low quality study. |
| **Wikberg et al 2017**13  Primary Health Care Centres, Sweden  N = 258 participants  Patients cluster randomised by individual practitioner. | PROM: MADRS 4 times over 3 months. 4 hours of practitioner training. Feedback of results to both practitioners and patients.  **Control group:** TAU. | No significant differences in depression (BDI-II), quality of life (EQ-5D), well-being, prescriptions, or sick leave over 12 months. More intervention arm patients continued antidepressants for 6 months (86/125 vs 78/133 among controls, p < 0.05). | **Limitations:** Only 80% power used for sample size calculation; 3-month follow-up only 72%, so target 80% follow-up not achieved.  Potential bias in allocation of GPs by drawing slips of paper to assign them. Not possible to blind practitioners and researchers due to the nature of the intervention. | Some benefit in terms of increased patient adherence to recommended length of treatment, but outcomes not improved. The study probably lacked power to detect possible clinically meaningful effects on outcomes of depression for patients.  Low quality study. |
| *Improving Access to Psychological Therapies (IAPT) services* | |  |  |  |
| **Delgadillo et al 2018**8  IAPT services  UK  N=77 therapists (2233 patients)  Cluster randomised by therapist | **PROMs:** PHQ-9 for depression and GAD-7 for anxiety. An automated computer algorithm alerted therapists to patients who were not on track and primed them to review these patients in clinical supervision.  **Control group:** TAU | No main effect of feedback was found overall however, patients classified as not on track (NOT) had less severe symptoms after treatment if they were allocated to the outcome feedback group than those in the control group (PHQ-9 d=0·23, B=–1·03, 95% CI –1·84 to –0·23; GAD-7 d=0·19, B=–0·85, 95% CI –1·56 to –0·14). | **Strength:** independent computer-generated randomisation  **Limitations:** lack of monitoring competence in treatment delivery or in feedback use.  Patients with high baseline severity scores (e.g. PHQ-9 ≥22) whose symptoms increased during treatment could not be classified as showing reliable deterioration, which is mostly an artefact of the measurement tools. | Supplementing psychological therapy with low-cost feedback technology can reduce symptom severity in patients at risk of poor response to treatment.  Moderate quality study |
| **Ongoing RCTs** |  |  |  |  |
| *Any setting* |  |  |  |  |
| **NCT02790970**16  (Lead investigator Browning)  Primary and secondary care clinics, UK, France, Spain, Germany and the Netherlands  N = 913 participants  Individual patient randomisation. | **PROM:** on-line PReDicT Test including QIDS-SR-16, facial expression rating task for negative emotional bias, and algorithm predicting response to drug treatment. Feedback to practitioner only.  **Control group:** complete PReDicT test but without feedback of result. | Primary outcome: QIDS-SR-16 at week 8. Secondary outcomes: EQ-5D, OxCAP-MH, MADRS, GAD-7, DSST, SAS-SR and costs over 48-50 weeks follow-up. | **Strength:** Remote computerised allocation to arm.  **Limitations:** Target sample calculated to have 80% power to detect MCID of 10% in MADRS score.  Expected attrition rate over 8 weeks, based on pilot study, is 33%.  Open label. | Trial is ongoing. |
| *Primary care* |  |  |  |  |
| **NCT03162211**17  (Lead investigator Uher)  Primary care group practices, Canada.  N = 304 participants  Unit of randomisation not reported. | **PROM:** On-line collection of self-report and clinician-rated depression severity, role functioning, quality of life and self-defined treatment goals, weekly for 3 months then monthly for 3 months. Feedback to both practitioners and patients.  **Control group:** complete on-line ratings but without feedback of result. | Primary outcome: QIDS-SR over 12 months.  Secondary outcomes: MADRS for remission of depression, WSAS for social function, EQ-5D for quality of life, LEAPS employment measure, healthcare cost and achievement of self-defined treatment goals. | **Strengths:** Remote computerised allocation to arm.  Blinding of outcomes assessor.  **Limitations:** Target sample power calculation not reported. May be underpowered to detect clinically meaningful differences in outcomes. | Trial is ongoing. |
| **ISRCTN** **17299295**18  (Lead investigator Kendrick)  Primary care group practices,  UK  N = 676 participants  Cluster randomised by practice. | **PROM:** PHQ-9 2 times over 1 month, administered remotely and results emailed to practitioners. Two hours of practitioner training. Feedback of results to both practitioners and patients.  **Control group:** TAU. | Primary outcome: BDI-II at 3 months.  Secondary outcomes over 6 months: BDI-II; antidepressant treatment; referrals for therapy; WSAS, EQ-5D, CSRI, MISS. Qualitative process evaluation. | **Strengths:** Target sample calculated to have 90% power to detect MCID of 17% in BDI-II score.  Remote computerised allocation to arm.  **Limitation:** Not possible to blind practitioners and researchers due to the nature of the intervention. | Trial is ongoing. |
| **Quasi experimental studies** | |  |  |  |
| *Improving Access to Psychological Therapies (IAPT) services* | |  |  |  |
| **Delgadillo et al 2017**7  IAPT stepped care service  UK  N= 594 participants  Quasi experimental before and after study | **PROM:** Computerised outcome feedback (OF) tool. The OF tool includes a graphical display of session-to-session PHQ-9 depression and GAD-7 anxiety scores with overlaid clinical benchmarks, which is referred to as expected treatment response (ETR) curves. The OF tool automatically alerted therapists about NOT cases using a ‘red signal’, if their symptoms surpassed the 80% upper boundary of the ETR curves, and were thus progressing substantially worse than other patients.  **Control:** therapists had access to plots of symptom severity scores on a weekly chart, without showing ETR curves or red signals | No statistically significant difference between groups were found in post-treatment PHQ-9 or GAD-7 measures. However, OF cases had significantly lower average duration and cost of treatment compared to controls (SMD £97.54, 95% CI £65.88 to £129.90) | **Strength:** The before-and-after design minimised confounding due to therapist effects, since each therapist was his/her own control.  **Limitations:** lack of random allocation and the use of historical controls raise threat to internal validity. | After adopting OF into their practice, this group of therapists attained similar clinical outcomes but within a shorter space of time and at a reduced average cost per treatment episode.  Low quality study |
| **Observational studies** |  |  |  |  |
| *Primary care* |  |  |  |  |
| **Moore et al 2009**11  Primary care practices  UK  N=604 patients  Retrospective cohort study | **PROM:** Records were examined of patients who had received a new diagnosis of depression, and who had completed PHQ-9 at initial diagnosis and a subsequent PHQ-9 within 6 months. | Controlling for the effects of potentially confounding factors, patients who showed an inadequate response in score change at the time of second assessment were nearly five times as likely to experience a subsequent change to treatment in comparison with those who showed an adequate response (OR 4.72, 95% CI 2.83 to 7.86). | **Strengths:** used routine data from 13 practices recruited from within three PCTs therefore likely to reflect standard use in clinical practice.  **Limitations:** Only 19% of practices that were approached participated in the study. However, it was possible to control for a number of potentially confounding factors.  Due to observational study design cannot determine cause and effect | The study provides evidence regarding the use of depression severity measures to monitor illness severity 5–12 weeks after diagnosis of depression. Those with a poor response to treatment (that is, with either an inadequate change in score or a follow-up score remaining above the case threshold) were five times more likely to experience management changes. The findings show only an association between a lack of change in questionnaire scores and treatment changes and, therefore, cannot determine cause and effect in this study. |
| **Abbreviations**  BDI-II: Beck Depression Inventory. CANE Camberwell Assessment of Need for the Elderly. CANSAS Camberwell Assessment of Need Short Appraisal Schedule. CNS Cardinal Needs Schedule. CSC Client Satisfaction Questionnaire. CSRI: Client Service Receipt Inventory. DIS: Diagnostic Interview Schedule. DSST: Digit Symbol Substitution Test. EQ-5D: EuroQol quality of life measure. FEP Questionnaire to Evaluate the Course of Psychotherapy. FLZ Life Satisfaction Questionnaire. GAD-7: Generalised Anxiety Disorder scale. GBB Physical Complaints Questionnaire. GSI: Global Severity Index. HADS Hospital Anxiety and Depression Scale. HAS Helping Alliance Scale. HASS: Highest Anxiety Subscale Score. ICD-10 10th revision of the International Statistical Classification of Diseases and Related Health Problems. LEAPS: Lam Employment Absence and Productivity Scale. MADRS: Montgomery-Asberg Depression Rating Scale. MANSA Manchester Short Assessment of Quality of Life. MCID: Minimum Clinically Important Difference. MD Mean difference. MISS: Medical Informant Satisfaction Scale. MSPSS Multidimensional Scale of Social Support. NOT: Not On Track (also known as signal cases. These are cases having poorer responses to treatment). OQ-45: Outcome Questionnaire–45-item. OR: Odds Ratio. ORS: Outcome Rating Scale. OxCAP-MH: Oxford CAPabilities questionnaire-Mental Health for wellbeing. PAE Treatment Progress Scale. PCT: Primary Care Trust. PHQ-9: Patient Health Questionnaire. PROM: Patient Reported Outcome Measure. PSYCHLOPS: Psychological Outcome Profiles scale QIDS-SR-16: Quick Inventory of Depression Symptoms, Self-Rated. SAS-SR: Social Adjustment Scale, Self-Rated. SCL-90: Symptom Checklist (90 items). SEED Short Evaluation of Eating Disorders. SF-36: Short Form (36 items) of Medical Outcomes Scale. SMD standardised mean difference. TAU: treatment as usual. TREAT-EAT Outcome Monitoring System. WSAS: Work & Social Adjustment Scale.  **Notes**  a 5 of these studies were included in the Cochrane systematic review by Kendrick et al., 2016  b 3 of these studies were included in the Cochrane systematic review by Kendrick et al., 2016 | | | | |