Feedback-informed treatment versus usual psychological treatment for depression and anxiety: a multisite, open-label, cluster randomised controlled trial

Jaime Delgadillo, Kim de Jong, Mike Lucock, Wolfgang Lutz, Julian Rubel, Simon Gilbody, Shehzad Ali, Elisa Aguirre, Mark Appleton, Jacqueline Nevin, Harry O'Hayon, Ushma Patel, Andrew Sainty, Peter Spencer, Dean McMillan

Lancet Psychiatry 2018
5: 564–72

Published Online
June 21, 2018
http://dx.doi.org/10.1016/S2215-0366(18)30162-7

This online publication has been corrected. The corrected version first appeared at thelancet.com/psychiatry on June 25, 2018

See Comment page 529

Clinical Psychology Unit, Department of Psychology, University of Sheffield, Sheffield, UK (J Delgadillo PhD); **Institute of Psychology, Leiden University, Leiden, Netherlands** (K de Jong PhD); **Centre for Applied Research in Health, University of Huddersfield, Huddersfield, UK** (Prof M Lucock PhD); **Department of Psychology, University of Trier, Trier, Germany** (Prof W Lutz PhD, J Rubel PhD); **Department of Health Sciences, University of York, York, UK** (Prof S Gilbody PhD, S Ali PhD, D McMillan PhD); **North East London National Health Service (NHS) Foundation Trust, London, UK** (E Aguirre PhD); **Pennine Care NHS Foundation Trust, Hyde, UK** (M Appleton MSc); **Cheshire and Wirral Partnership NHS Foundation Trust, Winsford, UK** (J Nevin MClinPsychol); **Whittington Health NHS Trust, London, UK** (H O'Hayon MSc); **Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK** (U Patel BSc); **Humber NHS Foundation Trust, Hessle, UK** (A Sainty Dip); **and South West Yorkshire Partnership NHS Foundation Trust, Barnsley, UK** (Prof M Lucock, P Spencer DClinPsy)

Correspondence to:
Dr Jaime Delgadillo, Clinical Psychology Unit, University of Sheffield, Sheffield S1 2LT, UK
jaime.delgadillo@nhs.net

Summary

Background Previous research suggests that the use of outcome feedback technology can enable psychological therapists to identify and resolve obstacles to clinical improvement. We aimed to assess the effectiveness of an outcome feedback quality assurance system applied in stepped care psychological services.

Methods This multisite, open-label, cluster randomised controlled trial was done at eight National Health Service (NHS) Trusts in England, involving therapists who were qualified to deliver evidence-based low-intensity or high-intensity psychological interventions. Adult patients (18 years or older) who accessed individual therapy with participating therapists were eligible for inclusion, except patients who accessed group therapies and those who attended less than two individual therapy sessions. Therapists were randomly assigned (1:1) to an outcome feedback intervention group or a treatment-as-usual control group by use of a computer-generated randomisation algorithm. The allocation of patients to therapists was quasi-random, whereby patients on waiting lists were allocated sequentially on the basis of therapist availability. All patients received low-intensity (less than eight sessions) or high-intensity (up to 20 sessions) psychological therapies for the duration of the 1-year study period. An automated computer algorithm alerted therapists in the outcome feedback group to patients who were not on track, and primed them to review these patients in clinical supervision. The primary outcome was symptom severity on validated depression (Patient Health Questionnaire-9 [PHQ-9]) and anxiety (Generalised Anxiety Disorder-7 [GAD-7]) measures after treatment of varying durations, which were compared between groups with multilevel modelling, controlling for cluster (therapist) effects. We used an intention-to-treat approach. This trial was prospectively registered with ISRCTN, number ISRCTN12459454.

Findings In total, 79 therapists were recruited to the study between Jan 8, 2016, and July 15, 2016, but two did not participate. Of these participants, 39 (51%) were randomly assigned to the outcome feedback group and 38 (49%) to the control group. Overall, 2233 patients were included in the trial (1176 [53%] were treated by therapists in the outcome feedback group, and 1057 [47%] by therapists in the control group). Patients classified as not on track 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], p=0·012; GAD-7 d=0·19, B=–0·85 [–1·56 to –0·14], p=0·019).

Interpretation Supplementing psychological therapy with low-cost feedback technology can reduce symptom severity in patients at risk of poor response to treatment. This evidence supports the implementation of outcome feedback in stepped care psychological services.

Funding English NHS and Department of Health Sciences, University of York, York, UK.

Introduction

Several psychological interventions, ranging from brief guided self-help to more intensive psychotherapies, are effective for the treatment of depression and anxiety disorders.[1] Large-scale assessments of such treatments applied in routine care are generally favourable, although it is also known that at least 30% of patients do not show statistically reliable improvement and approximately 10% of patients deteriorate.[2,3] Previous studies have shown that patients at risk of poor response to treatment can be identified early by use of outcome feedback.[4] Outcome feedback is a quality assurance method that involves routinely monitoring a patient's condition with standardised measures, which are compared with data from a normative clinical sample.[5] Individuals who are classified as not on track by data charts or automated electronic monitoring technologies are detected when the symptoms reported are significantly worse than the typical symptoms that are observed in similar cases.

Several reviews of experimental and practice-based studies suggest that using outcome feedback can help to

Research in context

Evidence before this study

Previous research suggests that the use of inexpensive quality improvement strategies such as routine outcome monitoring and feedback can improve psychological treatment outcomes, particularly for individuals who are prone to deterioration. The generalisability of previous trials is limited by their application in specialist university or psychotherapy clinics, and observational studies in primary care were likely to be statistically underpowered.

Added value of this study

This large-scale, pragmatic, randomised controlled trial was adequately powered to detect small effect size differences, and designed to assess the generalisability of feedback effects across multiple primary care psychological therapy services. The findings indicate that feedback-informed treatment can improve outcomes for patients classed as not on track, and with remarkable consistency across different services, therapists, and intensities of treatment.

Implications of all the available evidence

A compelling evidence base now exists to support the implementation of outcome monitoring and feedback technologies in mainstream psychological services. This low-cost, automated feedback and quality assurance system can help to prevent deterioration for individuals who are at risk of poor treatment outcomes.

improve treatment outcomes by comparison with usual psychological care.[4,6–8] Simply collecting patient-reported outcome measures in clinical practice is not associated with improved outcomes,[9] so the so-called risk signal element of feedback technologies could plausibly serve to effectively prompt therapists to identify and to resolve obstacles to improvement. This mechanism of action is supported by evidence from controlled trials in which therapy supported with risk signalling yielded better outcomes than did routine psychological care.[6–8] A 2010 meta-analysis[6] suggested that supplementing the risk signal with clinical decision making and support tools further enhances the effectiveness of feedback technologies; however, a meta-analysis[9] from 2016 contradicts this finding. Outcome feedback has been proposed to specifically help to prevent deterioration in patients classed as not on track.[6–8] A systematic review published in 2016[9] concluded that studies that applied risk-signalling technology showed some evidence of improved outcomes for individuals who were classed as not on track, but the pooled effect size was small (standardised mean difference of –0·22). Furthermore, some studies[10–12] have not found a differential effect of feedback in the not-on-track subgroup, and one study[13] found that the use of feedback possibly deteriorates outcomes for patients with cluster B personality disorders (eg, borderline personality disorder) who are classed as not on track.

Overall, the scientific literature shows mixed and inconclusive evidence for the use of feedback technologies, and the methodological quality of studies has been rated as generally low.[9] This variability raises questions about the generalisability of feedback, justifying the need to carefully assess its acceptability, feasibility, and effectiveness before use in routine care.[14] Some studies have suggested that outcome feedback might be particularly helpful in short-term evidence-based therapies such as cognitive behavioural therapy (CBT), and could enhance the efficiency of treatment.[10,11] One study[11] reported qualitative evidence that feedback-informed brief psychological interventions were acceptable to patients with depression and anxiety disorders, and they were feasible to implement in a routine primary care setting. This study also suggested that outcome feedback could reduce the cost and enhance the efficiency of treatment, although it was limited by the use of historical control group data in a non-randomised design. Despite these promising results, more rigorous experimental evidence is necessary to establish the generalisability and efficacy of feedback in primary care settings. We aimed to address this gap in the scientific literature through a multisite randomised controlled trial applied in primary care psychological services for common mental health problems. The trial was designed to assess the clinical effects of feedback-informed psychological treatments for patients who are classed as not on track.

Methods

Study design

This multisite, open-label, cluster randomised controlled trial was done at eight National Health Service (NHS) Trusts in England. Together, these services covered a large primary care population across London, Cambridge, Cheshire and Wirral, Bury, Heywood, Middleton, Rochdale, Oldham, Stockport, Tameside and Glossop, Trafford, Barnsley, and East Riding.

The study was approved by the London—City & East NHS Research Ethics Committee (Jan 6, 2016; reference 15/LO/2200), and the protocol is available online.

For the **protocol** see http://www.isrctn.com/ISRCTN12459454

Participants

Therapists qualified to deliver evidence-based low-intensity or high-intensity psychological interventions were eligible to take part, with the exception of therapists with short-term employment contracts (<1 year) and trainees who were not yet fully qualified. The trial included adult patients (18 years or older) who accessed

individual (low-intensity or high-intensity) therapy with participating therapists, excluding patients who accessed group therapies and those who attended less than two individual therapy sessions. The two-sessions condition was applied because, first, outcome measures for patients that accessed one session reflect symptom severity for a pretreatment period of 2 weeks and, second, the outcome feedback technology requires at least two sessions to provide a progress feedback signal, taking the first session as a baseline score. No other exclusion criteria were used to select patients.

For more on the **Patient Case Management Information System** see http://www.pc-mis.co.uk

A participant information sheet and consent form were shared via email with all therapists working in participating services. Therapists had an opportunity to clarify questions with the principal investigator before providing signed consent forms directly to the research team. Patients were not required to provide consent since they already provide consent for treatment and for therapists to gather and to examine clinical outcome measures in routine care.

Randomisation

Therapists (and all of their patients meeting the inclusion criteria) were randomly assigned (1:1) to an outcome feedback intervention group or a treatment-as-usual control group. The rationale for this design was that, first, randomly assigning therapists would minimise the risk of contamination of controls through practice effects, which could occur if the same therapist was to treat some patients with and others without outcome feedback technology. Second, this cluster design adequately represents the nesting of patients within therapists, thus enabling us to control for variability in outcomes attributable to therapists (therapist effects[15]). Parallel-group random allocation was independently done by a researcher using a computer-generated randomisation algorithm to prevent selection bias within services. Given the nature of the outcome feedback technology, this trial was open label, whereby therapists and patients were aware of their allocation. The allocation of patients to therapists in the participating services was quasi-random, whereby patients on waiting lists were allocated sequentially based on therapist availability.

Procedures

All participating services were part of the national Improving Access to Psychological Therapies (IAPT) programme, which offers protocol-driven, evidence-based psychological interventions for depression and anxiety disorders organised in a stepped care model.[16] Low-intensity guided self-help based on principles of CBT (LiCBT) was offered as an initial treatment to most patients with mild-to-moderate depression or anxiety problems. LiCBT was delivered by trained coaches (psychological wellbeing practitioners) in various formats (eg, individual or group psychoeducation and computerised CBT with telephone support) and typically involved up to eight sessions. Patients with more severe or complex problems, and those who did not respond to LiCBT, were stepped up to high-intensity psychotherapies (up to 20 sessions), including CBT, interpersonal psychotherapy, and counselling for depression. The specific treatment recommendation for each patient followed national clinical guidelines.[17] Treatment was supported by regular (weekly or fortnightly) clinical supervision delivered in a peer-supervision model organised within each service.

Therapists in all participating services routinely recorded their patients' weekly outcome measures using an electronic clinical record system called Patient Case Management Information System. This system includes outcome-monitoring graphs that chart depression and anxiety symptom severity scores at every session. Therapists randomly assigned to the outcome feedback group had access to enhanced outcome-monitoring graphs, which included expected treatment response curves. The expected treatment response curves represented 80% prediction intervals, which were estimated with growth curve modelling of data from a normative clinical sample.[5,18,19] We calculated expected treatment response curves for subgroups of patients with the same baseline symptom severity, using a large clinical dataset of patients treated in IAPT (further details described elsewhere[20]). These enhanced outcome-monitoring graphs automatically generated a risk signal to alert therapists to patients who were not on track (ie, whose depression or anxiety symptoms surpassed the 80% upper boundary of the expected treatment response curves). Control group therapists only had access to standard outcome-tracking graphs, but without expected treatment response curves or automated risk signals.

Therapists randomly assigned to the outcome feedback group attended a standardised 6·5 h training programme that covered outcome feedback theory and evidence base, instructions on how to use the feedback tool, and clinical trouble-shooting skills. The training required therapists to use the following process: first, review outcome feedback graphs with patients at the start of every session; second, if the graph shows a risk signal, discuss it with the patient to collaboratively identify potential obstacles to improvement; third, prioritise discussing not-on-track cases with a clinical supervisor; fourth, use information from the second and third points to develop a plan to address obstacles; and finally, use outcome feedback graphs to assess how the plan is working. Therapists were also primed to become aware of variables that were empirically associated with treatment outcomes (patient, therapist, process, and wider context factors). This information and evidence base were synthesised in a clinical guideline that therapists assigned to the outcome feedback group received after training.[21]

We collected session-by-session depression (Patient Health Questionnaire-9 [PHQ-9]) and anxiety (Generalised Anxiety Disorder-7 [GAD-7]) outcome measures for all

eligible patients who accessed individual therapy with participating therapists during a 1-year study period.

Outcomes

The primary outcome was depression and anxiety symptom severity assessed at the last treatment session with validated patient-reported outcome measures. Secondary outcomes were work and social adjustment, treatment duration, reliable improvement, reliable deterioration, treatment dropout rates, and the percentage of patients classified as not on track. Patients accessing the participating services routinely self-completed standardised outcome measures before each session. The PHQ-9 is a nine-item screening tool for depression, in which each item is rated on a 0–3 scale, yielding a total depression severity score between 0 and 27.[22] A cutoff score of 10 or above has been recommended to screen for major depression,[22] and a difference of 6 or more points between assessments is indicative of reliable change.[23]

GAD-7 is a seven-item measure developed to screen for anxiety disorders.[24] It is also rated with a 0–3 scale, yielding a total anxiety severity score between 0 and 21. A cutoff score of 8 or above is recommended to identify the possible presence of a diagnosable anxiety disorder,[24] and a difference of at least 5 points is indicative of reliable change.[23]

Secondary data sources included demographics (age, gender, ethnicity, and employment status), stepped care pathway information, number of treatment sessions, primary diagnoses recorded in clinical records, and functional impairment measured using the Work and Social Adjustment Scale (WSAS).[25]

Statistical analysis

Using the Optimal Design Software for Multi-level and Longitudinal Research (version 3.01), we estimated that a minimum of 60 therapists (30 per group), each of whom treated an average of ten patients, was required to detect a small effect size with an α of 0·05 and 80% power. This calculation assumed an intracluster correlation coefficient (ICC) of 0·05, guided by previous studies investigating therapist effects in naturalistic samples.[15,26] We aimed to recruit up to 80 therapists to account for attrition.

We compared patients' characteristics between groups (those included and excluded from the trial sample) using Mann-Whitney *U* tests for continuous variables and χ^2 tests for categorical variables. We imputed missing post-treatment outcome measures by averaging the values from 25 estimated datasets using an expectation maximisation method,[27] so that we could run intention-to-treat analyses, including post-treatment outcomes for all cases regardless of completion or dropout status.

We used multilevel modelling (MLM) with separate models for PHQ-9 and GAD-7 outcomes for the primary analysis. Following conventional model-building guidelines,[28] we initially examined the hierarchical structure of the dataset using unconditional models predicting post-treatment symptom severity. The site variable was not statistically significant in a three-level model (patients within therapists within sites), so subsequent analyses used two-level models (patients within therapists). Next, we considered different covariance structures, assessed non-linear (ie, quadratic or log-linear) trends in the number of treatment sessions, and assessed goodness of fit (using Akaike information criterion [AIC], Bayesian information criterion [BIC], and –2 log likelihood statistics). After initial model checking, we applied a two-level model in the primary analysis, including random intercepts for therapists, with an unstructured covariance matrix, and an identity link-function. No patients included in the trial sample had two interventions delivered by different therapists (eg, low-intensity followed by high-intensity therapy), so we did not model crossed random effects. Continuous variables were grand-mean centred, and we calculated the ICC to assess the proportion of variance in outcomes attributable to therapists. Our initial conditional model included the following predictors: baseline severity of symptoms, log-transformed number of sessions, and group (feedback *vs* control), which analysed between-group

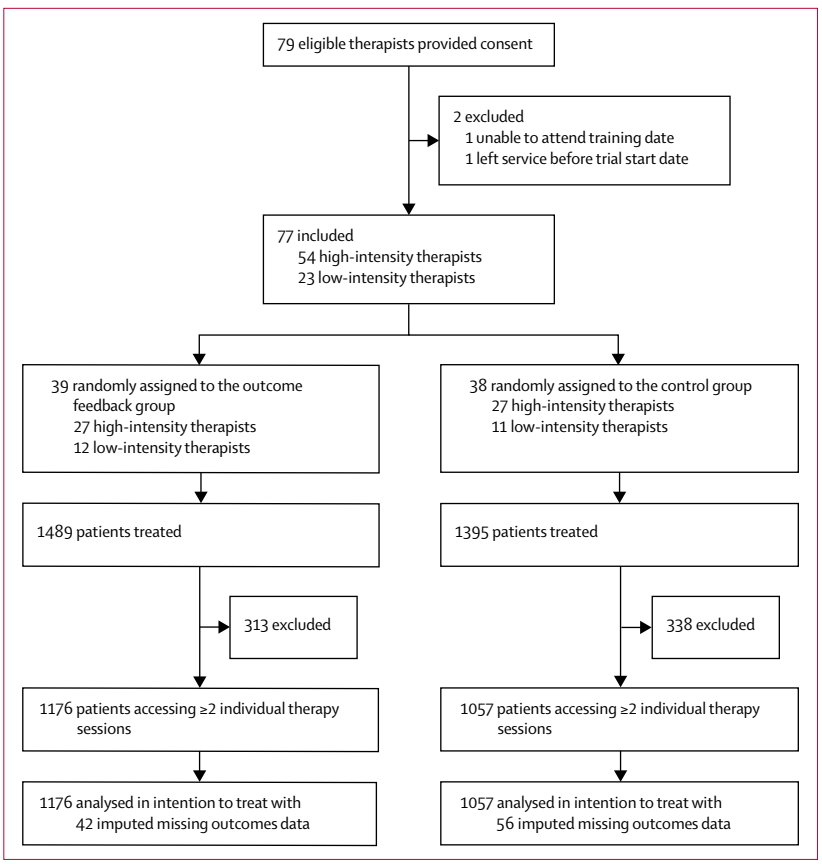

Figure 1: **Trial profile**

	Full sample	Outcome feedback group	Control group
Therapists			
Demographics			
Sex			
Male	12/77 (16%)	9/39 (23%)	3/38 (8%)
Female	65/77 (84%)	30/39 (77%)	35/38 (92%)
Age (years)	40·81 (11·13)	40·26 (11·29)	41·37 (11·10)
Ethnicity			
White British	65/77 (84%)	32/39 (82%)	33/38 (87%)
Other	12/77 (16%)	7/39 (18%)	5/38 (13%)
Professional experience (years)	7·42 (5·79)	7·46 (5·88)	7·38 (5·77)
Treatments			
High intensity	54/77 (70%)	27/39 (69%)	27/38 (71%)
Low intensity	23/77 (30%)	12/39 (31%)	11/38 (29%)
Patients			
Demographics			
Sex			
Male	765/2230 (34%)	424/1175 (36%)	341/1055 (32%)
Female*	1465/2230 (66%)	751/1175 (64%)	714/1055 (68%)
Age (years)	39·22 (15·02)	38·40 (14·66)	40·14 (15·38)
Unemployed*	286/1583 (18%)	164/808 (20%)	122/775 (16%)
Ethnicity*			
White British	1824/2061 (89%)	979/1099 (89%)	845/962 (88%)
Other	237/2061 (12%)	120/1099 (11%)	117/962 (12%)
Clinical characteristics			
Diagnosis			
Affective disorder	771/2233 (35%)	413/1176 (35%)	358/1057 (34%)
Mixed anxiety and depression	316/2233 (14%)	154/1176 (13%)	162/1057 (15%)
Generalised anxiety disorder	326/2233 (15%)	170/1176 (14%)	156/1057 (15%)
Other diagnosis	820/2233 (37%)	439/1176 (37%)	381/1057 (36%)
Baseline PHQ-9 score	15·29 (6·20)	14·96 (5·96)	15·65 (6·43)
Baseline GAD-7 score	13·99 (4·93)	13·82 (4·78)	14·19 (5·09)
Baseline WSAS score	19·29 (9·40)	19·08 (9·22)	19·52 (9·57)
Number of treatment sessions	6·45 (3·67); 6 (4–8)	6·54 (3·73); 6 (4–8)	6·35 (3·60); 6 (4–8)

Data are mean (SD), n (%), and median (IQR). PHQ-9=Patient Health Questionnaire-9. GAD-7=Generalised Anxiety Disorder-7. WSAS=Work and Social Adjustment Scale. *Calculations exclude cases with missing demographic data.

Table 1: **Trial sample characteristics**

differences in post-treatment symptom severity. Next, a fully adjusted model additionally included a case classification (case classified as on track *vs* not on track), and a group by classification interaction term (main hypothesis test). We repeated this MLM strategy in a sensitivity analysis controlling for age and intensity of treatment (low-intensity *vs* high-intensity treatment).

We assessed other relevant clinical outcomes in the secondary analyses. We repeated the fully adjusted MLM with the WSAS as a dependent variable to assess potential effects of feedback on functional impairment. We used Poisson MLM to analyse between-group differences in treatment duration, controlling for baseline PHQ-9 and GAD-7. We used logistic MLM to assess between-group probabilities (odds ratios [ORs]) of meeting post-treatment criteria for reliable improvement, after controlling for baseline severity (PHQ-9 and GAD-7). The reliable-improvement classification required patients to have statistically reliable improvement in at least one of the outcome measures, as long as the other measure did not show reliable deterioration. We also used logistic MLM to estimate between-group ORs for the percentage of patients with reliable deterioration (in at least one outcome measure), the percentage of patients classed as not on track, and the percentage of patients who dropped out of treatment. We computed these models using the full sample and repeated this calculation in the not-on-track subsample.

This trial was prospectively registered with ISRCTN, number ISRCTN12459454.

Role of the funding source

The funding organisations had no role in study design, data collection, data analysis, data interpretation, writing of the report, or the decision to publish the study. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

In total, 79 therapists were recruited between Jan 8, 2016, and July 15, 2016, but two did not participate (figure 1). Of the 77 participating therapists, 39 (51%) were randomly assigned to the outcome feedback group and 38 (49%) to the control group. Of these participants, 48 (62%) delivered high-intensity CBT, 23 (30%) delivered low-intensity CBT, and six (8%) delivered counselling for depression. Most therapists were women (65 [84%] of 77) and from a white British background (65 [84%] of 77), with a mean 7 years (SD 6) of experience in delivering psychological interventions (range from 9 months to 31 years; table 1). The number of patients treated by each therapist ranged between one and 113 (median 25 [IQR 30–50]; mean 30·77 [SD 24·54]).

Altogether, 2233 patients meeting case selection criteria were included in the trial (1176 [53%] were treated by therapists in the outcome feedback group, and 1057 [47%] by therapists in the control group). According to clinical records, 771 (35%) had a primary affective disorder (major depression episode or recurrent depression), 316 (14%) had mixed anxiety and depression disorder, 326 (15%) had generalised anxiety disorder, and 119 (6%) had post-traumatic stress disorder; other anxiety problems were less prevalent (table 1). The median number of weekly therapy sessions was 6 (IQR 4–8; range 2–25) in the full study sample, 6 (4–8; 2–22) in the outcome feedback group, and 6 (4–8; 2–25) in the control group (table 1).

The trial sample excluded 651 patients who did not access individual therapy (eg, group psycho-education cases) or who only attended a single session (figure 1).

	Depression (PHQ-9) fixed effects model			Anxiety (GAD-7) fixed effects model		
	B or variance (SE)	95% CI or Z	p value	B or variance (SE)	95% CI or Z	p value
Intercept	6·94 (0·35)	6·25 to 7·63	<0·0001	6·06 (0·33)	5·42 to 6·70	<0·0001
Sessions (log)*	–9·50 (0·45)	–10·38 to –8·63	<0·0001	–8·86 (0·40)	–9·65 to –8·07	<0·0001
Baseline severity (mean-centred)†	0·54 (0·02)	0·51 to 0·57	<0·0001	0·47 (0·02)	0·43 to 0·51	<0·0001
Group‡	0·19 (0·49)	–0·76 to 1·15	0·69	0·31 (0·45)	–0·57 to 1·20	0·49
Classification§	5·64 (0·30)	5·05 to 6·24	<0·0001	5·18 (0·27)	4·65 to 5·71	<0·0001
Group by classification¶	–1·03 (0·41)	–1·84 to –0·23	0·012	–0·85 (0·36)	–1·56 to –0·14	0·019
Variance components||**						
Residual	22·04 (0·66)	33·30	<0·0001	17·67 (0·53)	33·28	<0·0001
Random intercept	2·63 (0·59)	4·45	<0·0001	2·27 (0·50)	4·52	<0·0001

PHQ-9=Patient Health Questionnaire-9. GAD-7=Generalised Anxiety Disorder-7. *Log-linear transformation for number of treatment sessions. †Mean-centred values for PHQ-9 in the depression model or GAD-7 in the anxiety model. ‡0=controls; 1=outcome feedback patients. § 0=patients classified as on track; 1=patients classified as not on track. Note that there were two symptom-specific classifications, one for PHQ-9 and one for GAD-7. ¶A group by classification interaction term is the main hypothesis test. ||Variance components for PHQ-9 had an intracluster correlation coefficient (ICC) of 0·107 and for GAD-7 had an ICC of 0·114. **Data are variance and Z score only for the variance components.

Table 2: **Multilevel models predicting post-treatment depression and anxiety scores**

Excluded patients had similar baseline characteristics to those of trial patients (appendix), but a higher proportion of excluded patients were unemployed (22% *vs* 18%; p=0·040) and had marginally higher baseline PHQ-9 scores (mean difference 0·35; p=0·0070).

The main effect for the trial group was not significant in the initial conditional models testing between-group differences (appendix) or in the fully adjusted models testing interaction terms (table 2). The negative coefficients for the group by classification interaction terms indicated that patients classed as not on track in the outcome feedback group had lower post-treatment PHQ-9 scores (ie, less severe symptoms) than those in the control group (figure 2). The interaction term was significant for the depression model (B=–1·03 [95% CI –1·84 to –0·23]; p=0·012) and for the anxiety model (B=–0·85 [–1·56 to –0·14]; p=0·019; table 2). Approximately 11% of variability in depression (ICC 0·107) and anxiety (ICC 0·114) outcomes was attributable to therapist effects (table 2). Effect size differences between groups were *d*=0·17 for PHQ-9 and *d*=0·13 for GAD-7 in the whole sample (n=2233); the corresponding values in the not-on-track subsample (n=1288) were *d*=0·23 for PHQ-9 and *d*=0·19 for GAD-7 (table 3). Sensitivity MLM analyses controlling for age and intensity of treatment (low *vs* high) produced the same results (appendix).

The fully adjusted MLM results with WSAS as a dependent variable followed the same pattern as described previously (appendix). The main effect did not differ between groups (B=0·46 [95% CI –1·05 to 1·96]; p=0·55), but the group by classification interaction term was significant (B=–1·75 [–2·95 to –0·54]; p=0·0050), yielding an effect size of *d*=0·22 in the not-on-track subgroup. The Poisson MLM results indicated no differences in treatment duration between groups (B=–0·05 [–0·05 to 0·14]; p=0·37) and no group by classification interaction (B=–0·02 [–0·09 to 0·05]; p=0·62).

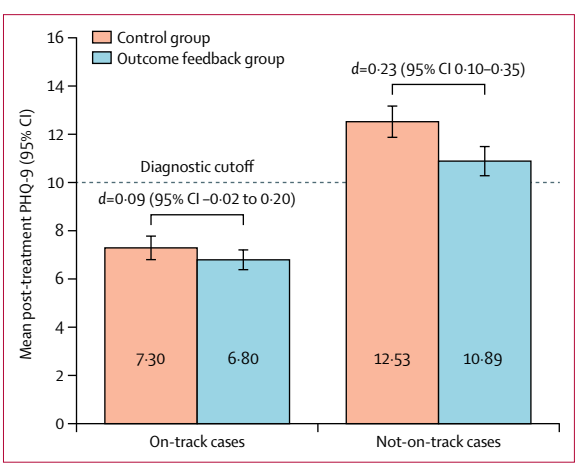

Figure 2: **Differences in post-treatment depression (PHQ-9) between outcome feedback and control groups**
PHQ-9=Patient Health Questionnaire-9. Error bars represent 95% CIs.

See **Online** for appendix

MLM results controlling for therapist effects indicated that no significant between-group differences existed for the odds of reliable improvement in the full sample (OR 1·21; p=0·29) or in the not-on-track subsample (1·32 [95% CI 0·93 to 1·89]; p=0·12). However, patients in the control group had greater odds of reliable deterioration than those in the outcome feedback group (full sample OR 1·48, p=0·023; not-on-track subsample OR 1·73 [1·18 to 2·54], p=0·0050). The odds of treatment dropout or of being classed as not on track did not differ significantly between groups (table 3).

Discussion

This large-scale, multisite trial done in stepped care IAPT services showed that supplementing psychological therapy with low-cost outcome feedback technology can improve outcomes for patients who are at risk of poor response to treatment. No main effect of feedback was

	Full sample (n=2233)			Not-on-track subsample (n=1288)		
	Outcome feedback group (n=1176)	Control group (n=1057)	Cohen's *d* or $OR_{adjusted}$ (95% CI)	Outcome feedback group (n=678)	Control group (n=610)	Cohen's *d* or $OR_{adjusted}$ (95% CI)
PHQ-9 pretreatment	14·41 (5·96)	14·85 (6·46)	..	14·47 (5·80)	15·45 (6·30)	..
PHQ-9 post-treatment	8·61 (6·60)	9·75 (7·12)	0·17	10·89 (7·17)	12·53 (7·37)	0·23
GAD-7 pretreatment	13·42 (4·85)	13·54 (5·24)	..	13·82 (4·77)	14·25 (5·00)	..
GAD-7 post-treatment	7·96 (5·78)	8·76 (6·12)	0·13	10·06 (6·12)	11·26 (6·37)	0·19
WSAS pretreatment	19·58 (8·67)	19·88 (9·12)	..	20·29 (8·70)	21·03 (8·91)	..
WSAS post-treatment	12·65 (9·57)	14·11 (9·98)	0·15	15·54 (10·23)	17·72 (9·95)	0·22
Reliable improvement	796 (68%)*	630 (60%)	1·21 (0·85–1·71)	412 (61%)*	317 (52%)	1·32 (0·93–1·89)
Reliable deterioration	49 (4%)	76 (7%)†	1·48 (1·06–2·07)	44 (6%)	68 (11%)‡	1·73 (1·18–2·54)
Dropout	284 (24%)	253 (24%)*	1·00 (0·70–1·43)	167 (25%)*	151 (25%)	1·03 (0·71–1·50)
Classified as not on track	678 (58%)*	610 (58%)	1·07 (0·86–1·32)	..	..	..

PHQ-9=Patient Health Questionnaire-9. GAD-7=Generalised Anxiety Disorder-7. WSAS=Work and Social Adjustment Scale. OR=odds ratio, adjusting for baseline severity. *Not significant (p>0·05). †p<0·05. ‡p<0·01.

Table 3: **Comparison of clinical outcomes**

found overall; instead an interaction effect indicated that feedback is specifically helpful for patients classified as not on track. These findings are largely consistent with reviews and meta-analyses of previous trials in university and outpatient psychotherapy centres, which conclude that the effect of feedback is mostly observed for patients who are not on track.[4,6,8,9] However, there are also exceptions, such as the trial by Amble and colleagues,[12] which found main effects for feedback in the full sample but not in the not-on-track subgroup. Effect sizes of d=0·23 for depression, d=0·19 for anxiety, and d=0·22 for work and social adjustment favouring the feedback group were observed. These effect sizes are small by conventional standards, but nevertheless remarkable considering the automated nature of the risk-signalling technology and the low cost incurred by services in requiring outcome feedback users to attend a single-day training session. Additionally, given that the feedback intervention prioritises clinical supervision resources for patients who are not on track, the results indicate that this prioritisation of patients did not disadvantage patients who were on track in terms of clinical outcomes or dropout rates. Overall, this low-cost quality assurance system effectively integrates the use of routine outcome measures, outcome prediction technology, and clinical supervision.

Given that usual treatment in IAPT stepped care services uses standard outcome-tracking charts and regular clinical supervision, we might expect modest effect size differences when supplementing such treatment with risk-signalling technology. Patients in the control group were more likely to deteriorate than those in the outcome feedback group, although the ORs in this trial (full sample OR 1·48; not-on-track subsample OR 1·73) were lower than the OR of 2·3 reported in the meta-analysis by Shimokawa and colleagues.[6] This difference might be explained by the low number of patients with reliable deterioration in the participating services (<7·5%), whereas other psychotherapy settings have typically observed deterioration rates of about 10%.[1] This discrepancy is plausibly explained by differences in patient characteristics (eg, case-mix), since IAPT services mostly support people with mild-to-moderate mental health problems.[16]

Contrary to other studies that have used evidence-based CBT interventions,[10,11] we found no significant effects of feedback on treatment duration. One methodological explanation could be that previous quasi-experimental studies did not have contemporaneous controls, and their effects on duration could be explained by other unmeasured factors that changed over time. An alternative explanation could be that the inclusion of counselling and LiCBT interventions in this trial might have obscured effects that could be specific to conventional CBT. The potential effect of feedback on treatment duration and costs requires further investigation.

The inclusion of services across diverse regions in England is a key strength of this study, offering compelling evidence of generalisability in contrast to earlier single-centre pilot studies.[11,29] The risk-signalling technology was developed with historical data from a service and region that did not take part in this trial,[11] thus offering a strong test of the generalisability and predictive power of the outcome feedback model. The present trial was adequately powered to detect a small effect and to control for therapist effects. This central feature is an important advance, confirming that the use of feedback technology improves clinical outcomes after accounting for variations in therapeutic aptitude across multiple practitioners. Of note, the therapist effect estimate (approximately 11%) in our study explains a considerably larger proportion of variance than does the effect of feedback, so attention to the factors that characterise underperforming therapists is clearly warranted. It is, of course, plausible that some therapists might make better use of feedback than others, and future

studies could investigate the personal attitudes, skills, or organisational conditions that optimise adequate use of feedback.[30–32]

Some limitations should also be borne in mind when interpreting the results of this study. Although we included a sizeable group of therapists delivering a range of low-intensity and high-intensity interventions, our study participants volunteered to take part in the trial. We did not have information about the total size or professional characteristics of the workforce across all participating services, so we cannot assume that trial therapists are necessarily representative of the wider workforce. Furthermore, we did not have the resources to closely monitor competence in treatment delivery or in feedback use. A central feature of this feedback model involves discussing risk signals with patients and clinical supervisors; however, we did not have objective data to assess the extent to which these features were adhered to. A further methodological issue relates to potential ceiling effects. Patients with high baseline severity scores (eg, PHQ-9 ≥22) whose symptoms increased during treatment could not be classified as showing reliable deterioration, which is mostly an artifact of the measurement tools and reliable change indices used in the study. Therefore, the true extent of reliable deterioration rates could possibly be underestimated. Furthermore, like most other feedback studies to date,[8] this trial only had a short-term observation period, since outcomes were assessed at the end of treatment. Thus, whether the observed effects of feedback could have a durable impact on longer-term symptoms and functioning is unknown.

Contributors

JD, KdJ, ML, WL, JR, SG, and DM contributed to the study design, data interpretation, and writing and approval of the final manuscript. JD analysed all data. SA provided statistical advice and contributed to the writing and interpretation of results. EA, MA, JN, HO'H, UP, AS, and PS contributed to the management of all trial sites, including recruitment, data collection, data interpretation, and writing and approval of the final manuscript. JD was the principal investigator, led the data collection, and is guarantor.

Declaration of interests

We declare no competing interests.

Acknowledgments

The outcome feedback and risk-signalling technology used in this study was developed by Patient Case Management Information System (PCMIS) at the Department of Health Sciences, University of York, York, UK.

References

1 Lambert MJ. The efficacy and effectiveness of psychotherapy. In Lambert MJ, ed. Bergin and Garfield's handbook of psychotherapy and behavior change, 6th edn. New York, NY: Wiley & Sons, 2013.

2 Hansen NB, Lambert MJ, Forman EM. The psychotherapy dose-response effect and its implications for treatment delivery services. *Clin Psychol-Sci Pr* 2002; **9:** 329–43.

3 NHS Digital. Psychological therapies: annual report on the use of IAPT services, England, 2015–16. Oct 18, 2016. http://digital.nhs.uk/catalogue/PUB22110 (accessed Nov 26, 2017).

4 Carlier IV, Meuldijk D, Van Vliet IM, Van Fenema E, Van der Wee NJ, Zitman FG. Routine outcome monitoring and feedback on physical or mental health status: evidence and theory. *J Eval Clin Pract* 2012; **18:** 104–10.

5 Finch AE, Lambert MJ, Schaalje BG. Psychotherapy quality control: the statistical generation of expected recovery curves for integration into an early warning system. *Clin Psychol Psychother* 2001; **8:** 231–42.

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

7 Castonguay LG, Barkham M, Lutz W, McAleavey AA. Practice-oriented research: approaches and applications. In Lambert MJ, ed. Bergin and Garfield's handbook of psychotherapy and behavior change, 6th edn. New York, NY: Wiley & Sons, 2013.

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

9 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 Syst Rev* 2016; **7:** CD011119.

10 Janse PD, De Jong K, Van Dijk MK, Hutschemaekers GJ, Verbraak MJ. Improving the efficiency of cognitive-behavioural therapy by using formal client feedback. *Psychother Res* 2017; **27:** 525–38.

11 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.

12 Amble I, Gude T, Stubdal S, Andersen BJ, Wampold BE. The effect of implementing the Outcome Questionnaire-45.2 feedback system in Norway: a multisite randomized clinical trial in a naturalistic setting. *Psychother Res* 2015; **25:** 669–77.

13 de Jong K, Segaar J, Ingenhoven T, van Busschbach J, Timman R. Adverse effects of outcome monitoring feedback in patients with personality disorders: a randomized controlled trial in day treatment and inpatient settings. *J Pers Disord* 2017; published online June 8. DOI:10.1521/pedi_2017_31_297.

14 Lutz W, De Jong K, Rubel J. Patient-focused and feedback research in psychotherapy: where are we and where do we want to go? *Psychother Res* 2015; **25:** 625–32.

15 Baldwin SA, Imel ZE. Therapist effects: findings and methods. In Lambert MJ, ed. Bergin and Garfield's handbook of psychotherapy and behavior change, 6th edn. New York, NY: Wiley & Sons, 2013.

16 Clark DM. Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *Int Rev Psychiatry* 2011; **23:** 318–27.

17 NICE. Common mental health disorders: identification and pathways to care. CG123. London: National Institute for Health and Care Excellence, 2011.

18 Lutz W, Martinovich Z, Howard KI. Patient profiling: an application of random coefficient regression models to depicting the response of a patient to outpatient psychotherapy. *J Consult Clin Psychol* 1999, **67:** 571–77.

19 Lutz W, Martinovich Z, Howard KI, Leon SC. Outcome management, expected treatment response and severity adjusted provider profiling in outpatient psychotherapy. *J Clin Psychol* 2002, **58:** 1291–304.

20 Delgadillo J, Moreea O, Lutz W. Different people respond differently to therapy: a demonstration using patient profiling and risk stratification. *Behav Res Ther* 2016; **79:** 15–22.

21 Delgadillo J, Lucock M, de Jong K. Using outcome feedback in psychological therapy: a guideline for IAPT practitioners. York: Department of Health Sciences, University of York, 2015.

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

23 Richards DA, Borglin G. Implementation of psychological therapies for anxiety and depression in routine practice: two year prospective cohort study. *J Affect Disord* 2011; **133:** 51–60.

24 Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med* 2007; **146:** 317–25.

25 Mundt JC, Marks IM, Shear MK, Greist JM. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *Br J Psychiatry* 2002; **180:** 461–64.

26 Schiefele AK, Lutz W, Barkham M, et al. Reliability of therapist effects in practice-based psychotherapy research: a guide for the planning of future studies. *Adm Policy Ment Health* 2017; **44:** 598–613.

27 Schafer JL, Olsen MK. Multiple imputation for multivariate missing-data problems: a data analyst's perspective. *Multivariate Behav Res* 1998; **33:** 545–71.

28 Raudenbush, SW. Hierarchical linear models and experimental design. In Edwards L, ed. Applied analysis of variance in behavioral science. New York, NY: Marcel Dekker, 1993: 459–96.

29 Lucock M, Halstead J, Leach C, et al. A mixed-method investigation of patient monitoring and enhanced feedback in routine practice: barriers and facilitators. *Psychother Res* 2015; **25:** 633–46.

30 de Jong K, van Sluis P, Nugter MA, Heiser WJ, Spinhoven P. Understanding the differential impact of outcome monitoring: therapist variables that moderate feedback effects in a randomized clinical trial. *Psychother Res* 2012; **22:** 464–74.

31 de Jong K, de Goede M. Why do some therapists not deal with outcome monitoring feedback? A feasibility study on the effect of regulatory focus and person–organization fit on attitude and outcome. *Psychother Res* 2015; **25:** 661–68.

32 Lutz W, Zimmermann D, Müller VN, Deisenhofer AK, Rubel JA. Randomized controlled trial to evaluate the effects of personalized prediction and adaptation tools on treatment outcome in outpatient psychotherapy: study protocol. *BMC Psychiatry* 2017; **17:** 306.