

Use of standardised outcome measures in adult mental health services

Randomised controlled trial

MIKE SLADE, PAUL McCRONE, ELIZABETH KUIPERS, MORVEN LEESE, SHARON CAHILL, ALBERTO PARABIAGHI, STEFAN PRIEBE and GRAHAM THORNICROFT

Background Routine use of standardised outcome measures is not universal.

Aims To evaluate the effectiveness of standardised outcome assessment.

Method A randomised controlled trial, involving 160 representative adult mental health patients and paired staff (ISRCTN16971059). The intervention group ($n=101$) (a) completed monthly postal questionnaires assessing needs, quality of life, mental health problem severity and therapeutic alliance, and (b) received 3-monthly feedback. The control group ($n=59$) received treatment as usual.

Results The intervention did not improve primary outcomes of patient-rated unmet need and of quality of life. Other subjective secondary outcome measures were also not improved. The intervention reduced psychiatric in-patient days (3.5 v. 16.4 mean days, bootstrapped 95% CI 1.6–25.7), and hence service use costs were £2586 (95% CI 102–5391) less for intervention-group patients. Net benefit analysis indicated that the intervention was cost-effective.

Conclusions Routine use of outcome measures as implemented in this study did not improve subjective outcomes, but was associated with reduced psychiatric in-patient admissions.

Declaration of interest None. Funding by the Medical Research Council.

There is international consensus that outcome should be routinely measured in clinical work (Health Research Council of New Zealand, 2003; Trauer, 2003). However, psychiatrists do not use standardised outcome measures routinely (Gilbody *et al*, 2002a), preferring their care to be judged by other criteria (Valenstein *et al*, 2004). The overall evidence from systematic reviews (Gilbody *et al*, 2001, 2002b) and higher-quality trials (Ashaye *et al*, 2003; Marshall *et al*, 2004) is negative, so clinicians remain unconvinced about the effectiveness of routine outcome measurement (Bilsker & Goldner, 2002). We previously applied the Medical Research Council (MRC) framework for complex health interventions (Campbell *et al*, 2000) to the use of outcome measures in adult mental health services, by reviewing relevant theory (Slade, 2002a) and developing a testable model linking routine use of outcome measures with improved patient outcomes (Slade, 2002b). The aim of the present exploratory randomised controlled trial was to test the model.

METHOD

Design

The trial was intended to extend previous work in three ways. First, sample representativeness was maximised by choosing patients from a site which was demographically representative, and then selecting the sample using stratified random sampling on known prognostic factors. Second, outcome measures were applied longitudinally, i.e. with more than one (as in previous studies) or two administrations, to allow cumulative effects to be investigated. Third, each element of the pre-specified model of the intervention effects was evaluated (Slade, 2002b). In summary, the intervention involved asking staff and patient pairs to separately complete standardised measures, and then providing both with identical feedback. In the model, it

was hypothesised that both completing the assessments and receiving the feedback would create cognitive dissonance (an awareness of discrepancy between actual and ideal states) regarding the content and process of care, which in turn would lead to behavioural change in content and process of care, and consequent improvement in outcome. Therefore the two active ingredients were completion of outcome measures and receipt of feedback, and the intervention might have had an impact on patients as well as staff. Hence, in contrast to previous studies in which staff received feedback on patient-completed assessments (Ashaye *et al*, 2003; Marshall *et al*, 2004; van Os *et al*, 2004), in this model both staff and patients completed assessments and received feedback. The model had the advantage of being explicit about the anticipated effects of the intervention, and therefore testable and falsifiable at each stage.

Participants

The inclusion criteria for patients were that they had been on the case-load of any of the eight community mental health teams in Croydon, South London, on 1 May 2001, for at least 3 months; and that they were aged between 18 and 64 years. Croydon has a nationally representative population of 319 000, with 3500 patients using eight community mental health teams. To ensure epidemiological representativeness, sample selection involved stratified random sampling on known prognostic factors: age (tertiles), gender, ethnicity (White v. Black and minority ethnic), diagnosis (psychosis v. other) and community mental health team. One member of staff was then identified who was working most closely with each selected patient.

Measures

The rationale for the choice of measures is reported elsewhere (Slade, 2002a). Staff completed three measures in the postal questionnaire. The Threshold Assessment Grid (TAG) is a 7-item assessment of the severity of a person's mental health problems (range 0–24, the lower the score, the better) (Slade *et al*, 2000). The Camberwell Assessment of Need Short Appraisal Schedule staff version (CANSAS-S) is a 22-item assessment of unmet needs (current serious problems, regardless of any help received) and met needs (no or moderate problem because of help given) (range for both 0–22, the lower the score, the better) (Slade *et*

al, 1999). The Helping Alliance Scale staff version (HAS-S) is a 5-item assessment of therapeutic alliance (range 0–10, the higher the score, the better) (McCabe *et al*, 1999).

Patients completed three measures in the postal questionnaire. The CANSAS-P is a patient's 22-item assessment of met and unmet needs (scores as for CANSAS-S) (Slade *et al*, 1999). The Manchester Short Assessment (MANSA) is a 12-item assessment of quality of life (range 1–7, the higher the score, the better) (Priebe *et al*, 1999). The HAS-P is a 6-item patient's assessment of therapeutic alliance (score as for HAS-S) (McCabe *et al*, 1999).

Three measures were assessed at baseline and follow-up only. The Brief Psychiatric Rating Scale (BPRS) is an 18-item interviewer-rated assessment of symptoms (range 0–126, the lower the score, the better) (Overall & Gorham, 1988). The Health of the Nation Outcome Scale (HoNOS) is a 12-item staff-rated assessment of clinical problems and social functioning (range 0–48, the lower the score, the better) (Wing *et al*, 1998). The patient-rated Client Service Receipt Inventory (CSRI) was used to assess service use during the previous 6 months (Beecham & Knapp, 2001).

Sample size

The CANSAS-P and MANSA were the primary outcome measures, and a reduction of 1.0 unmet needs on the CANSAS-P or an increase of 0.25 on the MANSA were defined in advance as the criteria for improved effectiveness. Secondary outcome measures were the TAG, BPRS, HoNOS and hospital admission rates. The sample size required for the two arms differed since the study also tested another hypothesis within the intervention group arm only, for which 85 patients needed to receive the intervention (Slade *et al*, 2005). The CANSAS-P unmet needs has a standard deviation of 1.7 (Thorncroft *et al*, 1998) and a pre-post correlation after 24 months of 0.32. Assuming an alpha level of 0.05 and that analysis of covariance is used to compare $t2$ values while adjusting for $t1$ levels, a control group of 50 would detect a change of 1.0 patient-rated unmet need with a power of 0.94. The MANSA has a standard deviation of 0.5 and a pre-post correlation of 0.5 (Thorncroft *et al*, 1998) so, with the same assumptions, this sample size would detect a change of 0.25 in quality-of-life rating with a power of 0.9. To allow for dropping out, 160 patients were recruited.

Procedures

Ethical approval and written informed consent from all staff and patient participants were obtained. A trial steering committee met throughout the study and required interim analysis of adverse events. All researchers were trained in standardised assessments through role-play, vignette rating and observed assessments. Assessment quality was monitored by double-rating 13 patient assessments, showing acceptable concordance: 8 (2.8%) of 286 CAN ratings differed, and there was a mean difference of 0.14 in 216 BPRS ratings.

For each pair, baseline staff and patient assessments by researchers composed the postal questionnaire plus trial measures. Following baseline assessment, patients were allocated by an independent statistician who was masked to the results of the baseline assessment. The statistician used a purpose-written Stata program, to ensure random allocation and balance on prognostic factors of age (tertiles), gender, ethnicity (White *v.* Black and minority ethnic), diagnosis (psychosis *v.* other) and community mental health team. Allocation was concealed until the intervention was assigned. Staff and patients were aware of their allocation status.

The control group received treatment as usual, involving mental healthcare from the multidisciplinary community mental health team focused on mental health and social care needs, together with care from the general practitioner for physical healthcare needs.

The intervention group received treatment as usual and, in addition, staff-patient pairs were separately asked to complete a monthly postal questionnaire and were provided by the research team with identical feedback by post at 3-monthly intervals. Feedback was sent 2 weeks after round 3 and round 6 postal questionnaires, and comprised colour-coded graphics and text, showing change over time and highlighting areas of disagreement. Patients were paid £5 for each round of assessments.

Follow-up assessments were made at 7 months. At follow-up, patients were asked not to disclose their status, and assignment was guessed by the researcher after the postal questionnaire element. Staff and patient self-report data were collected on the cognitive and behavioural impact of the intervention. Written care plans were audited at baseline and follow-up.

Analysis

Differences in administration time were tested using paired sample t -tests, and between patients with and without follow-up data using chi-squared and independent-samples t -tests. Data analysis was undertaken on an intention-to-treat basis, for all participants with follow-up data. Effectiveness was investigated using independent-samples t -tests to compare the outcome at follow-up for intervention- and control-group patients. Sensitivity analyses included:

- (a) analysis of covariance to adjust for the baseline level;
- (b) analysis of covariance including random effects for staff member and community mental health team (to check for any clustering effects);
- (c) t -test on the outcomes, with missing values imputed from baseline data;
- (d) Mann-Whitney tests.

A broad costing perspective was used. Production costs were not included. Service-cost data were obtained by combining CSRI data with unit-cost information to generate service costs. More unit costs were taken from a published source (Netten & Curtis, 2002). Some criminal-justice unit costs were estimated specifically for the study: £100 per court attendance and £50 per solicitor contact. Based on assessment processing time, the average cost of providing the intervention was £400 per patient. This assumed that the two researchers employed on the study for 2 years provided two rounds of the intervention to 100 patients, plus two assessments for 160 patients. It was further assumed that the assessments entailed the same administrative time as the intervention. Per year, therefore, each research worker could provide 130 assessments or interventions, and the salary cost of this was about £200 (i.e. £400 for both rounds of the intervention).

Mean number of service contacts (bed-days for in-patient care) and costs at follow-up were compared using regression analysis, with the allocation status and baseline service use or cost entered as independent variables. Resource use data are typically skewed, so bootstrapping with 1000 repetitions was used to produce confidence intervals for cost differences (Netten & Curtis, 2002). A sensitivity analysis was performed by assessing the significance of the difference in total costs after excluding in-patient care.

Cost-effectiveness was investigated using the net-benefit analysis and cost-effectiveness acceptability curves (not shown). Net-benefit analysis uses the equation $\text{net benefit} = \lambda O - SC$ where O is outcome, SC is service cost and λ is the value placed on one unit of outcome (Briggs, 2001); λ is a hypothetical amount that would be problematic to determine, but net benefits can be compared for different values of λ . This involved regression analysis (controlling for baseline costs), with the net benefits associated with λ s between £0 and £90 as the dependent variables, and allocation status as the main independent variable. For each regression, 1000 bootstrap resamples were produced, and for each of these the proportion of regression coefficients that were above zero indicated the probability that the intervention was more cost-effective than the control condition.

RESULTS

Participants

Between May 2001 and December 2002, 160 patients were recruited, with follow-up completed by July 2003. Socio-demographic and baseline clinical assessments for patients are shown in Table 1.

Among the 74 staff who participated in baseline assessments were 43 psychiatric nurses, 14 social workers and 11 psychiatrists. Postal questionnaire completion rates for staff for rounds 2 to 6 were 78%, 71%, 67%, 59% and 58% respectively; 486 staff postal questionnaires were sent and 325 (67%) returned. For patients, the completion rates for rounds 2–6 were 85%, 84%, 76%, 76% and 76% respectively; 487 postal questionnaires were sent and 386 (79%) returned. Three-monthly summary feedback was sent after round 3 to 96 (95%) staff–patient pairs, and after round 6 to 93 (92%) staff–patient pairs. The trial flow diagram is shown in Fig. 1.

No demographic or baseline clinical variables differed between the 142 patients with and the 18 patients without full follow-up data (Fig. 1).

There was a significant reduction in completion time by the 129 patients for whom completion-time data were available (14.9 to 8.7 min, $P < 0.001$), but not for the 130 staff with these data (7.8 to 7.4 min).

Some researcher masking to allocation status was retained. In 81 (57%) of the 143 staff interviews and in 41 (29%) of

Table 1 Social and baseline clinical characteristics of patients ($n=160$)

| Characteristic | All ($n=160$) | Intervention group ($n=101$) | Control group ($n=59$) |
|--|--------------------|-----------------------------------|-----------------------------|
| Age, years: mean (s.d.) | 41.2 (11.2) | 41.8 (11.4) | 40.2 (10.8) |
| Men, n (%) | 78 (49) | 48 (48) | 30 (51) |
| Ethnicity, n (%) | | | |
| White | 122 (76) | 77 (76) | 45 (76) |
| Black African–Caribbean | 20 (13) | 16 (16) | 9 (15) |
| Indian | 6 (4) | 4 (4) | 2 (3) |
| Other | 12 (8) | 4 (4) | 3 (5) |
| Highest educational level, n (%) | | | |
| No formal qualification | 61 (38) | 38 (38) | 23 (39) |
| GCSE or GCE ¹ | 45 (28) | 28 (28) | 19 (32) |
| A levels ² | 14 (9) | 10 (10) | 3 (5) |
| Higher diploma or degree | 16 (10) | 11 (11) | 4 (7) |
| Not known | 24 (15) | 13 (13) | 10 (17) |
| Primary clinical diagnosis, n (%) | | | |
| Schizophrenia | 60 (38) | 40 (40) | 20 (34) |
| Bipolar affective disorder | 17 (11) | 8 (8) | 9 (15) |
| Other psychoses | 21 (13) | 12 (12) | 7 (12) |
| Affective disorder | 43 (27) | 27 (27) | 16 (27) |
| Personality disorder | 11 (7) | 7 (7) | 4 (7) |
| Other | 8 (5) | 7 (7) | 3 (5) |
| Contact with mental health services, mean (s.d.) | | | |
| Years since first contact | 13.1 (11.8) | 14.2 (12.6) | 11.1 (9.8) |
| Years in this episode of care | 4.1 (4.2) | 4.3 (4.7) | 3.9 (3.3) |
| Measure | | | |
| CANSAS–S unmet score, mean (s.d.) | 2.98 (3.19) | 3.24 (3.31) | 2.54 (2.94) |
| CANSAS–S met score, mean (s.d.) | 5.04 (3.43) | 5.06 (3.29) | 5.02 (3.69) |
| TAG score, mean (s.d.) | 5.21 (3.64) | 5.44 (3.58) | 4.81 (3.73) |
| HAS–S score, mean (s.d.) | 7.34 (1.61) | 7.45 (1.59) | 7.14 (1.64) |
| HoNOS score, mean (s.d.) | 8.87 (6.43) | 9.15 (6.63) | 8.40 (6.10) |
| CANSAS–P unmet score, mean (s.d.) | 4.59 (3.62) | 4.36 (3.36) | 4.98 (4.05) |
| CANSAS–P met score, mean (s.d.) | 4.21 (2.88) | 4.23 (2.81) | 4.17 (3.04) |
| HAS–P score, mean (s.d.) | 7.95 (1.94) | 8.19 (1.79) | 7.54 (2.12) |
| MANSA score, mean (s.d.) | 4.25 (1.01) | 4.25 (0.99) | 4.25 (1.05) |
| BPRS score, mean (s.d.) | 33.51 (9.29) | 33.35 (9.04) | 33.79 (9.78) |

CANSAS, Camberwell Assessment of Need Short Appraisal Schedule, –S, staff version, –P, patient version; TAG, Threshold Assessment Grid; HoNOS, Health of the Nation Outcome Scale; HAS, Helping Alliance Scale, –S, staff version, –P, patient version; MANSA, Manchester Short Assessment; BPRS, Brief Psychiatric Rating Scale.

1. Normally taken at 16 years of age.

2. Normally taken at 18 years of age.

the 140 patient interviews, the researchers were unable to guess allocation status. Where they did rate allocation status, they were correct for 97 (92%) of their 105 intervention-group ratings, and for 53 (95%) of their 56 control-group ratings.

Two adverse events occurred. One intervention-group patient withdrew consent during the study, stating that the questions were ‘too disturbing and intrusive’. One intervention-group patient was sent to prison on remand during the intervention, following a serious assault. There

was no evidence linking the assault with involvement in the study.

Primary outcomes

Follow-up assessments of the two primary outcomes are shown in Table 2.

For the 142 patients with baseline and follow-up patient-rated unmet-need data, 79 (56%) had at least 1 fewer unmet needs at follow-up, comprising 51 (55%) out of 93 in the intervention group and 28 (57%) out of 49 in the control group. There

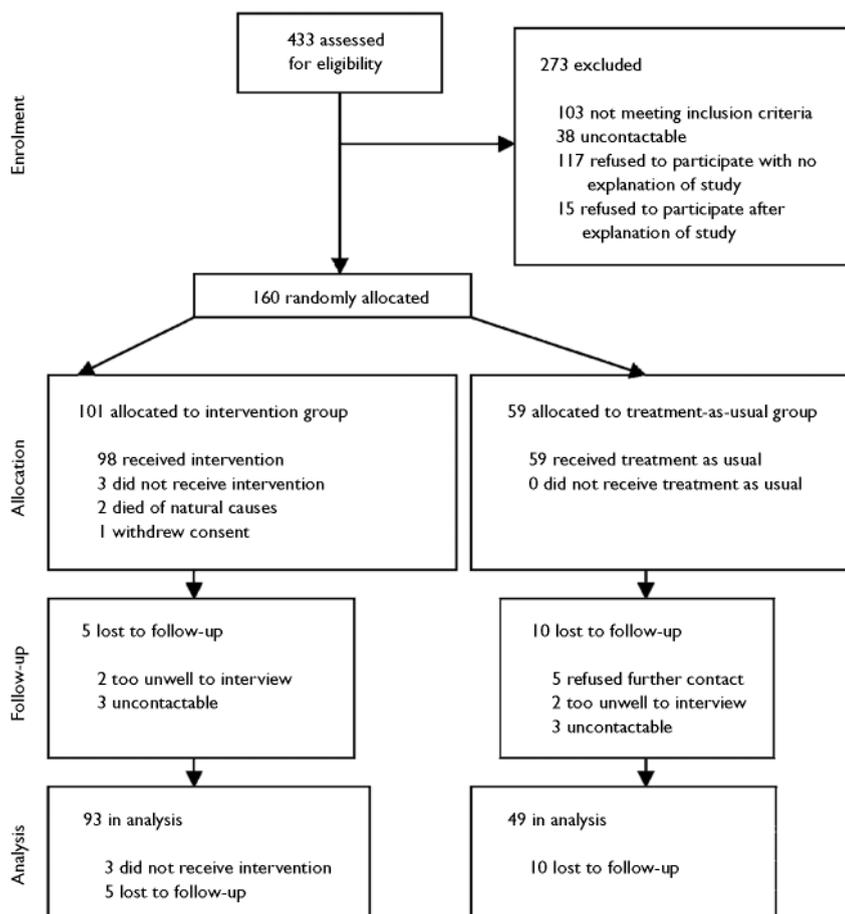


Fig. 1 CONSORT diagram.

Table 2 Follow-up measures

| Measure | Score, mean (s.d.) | | Difference | 95% CI |
|----------------------|---------------------------|----------------------|------------|-------------|
| | Intervention group (n=93) | Control group (n=49) | | |
| CANSAS-S unmet needs | 2.93 (3.56) | 2.02 (2.57) | -0.91 | -2.0 to 0.1 |
| CANSAS-S met needs | 4.06 (2.89) | 5.23 (3.86) | 1.17 | -0.1 to 2.4 |
| TAG | 5.14 (3.58) | 4.58 (3.34) | -0.55 | -1.8 to 0.7 |
| HAS-S | 7.54 (1.62) | 7.33 (1.88) | -0.21 | -0.8 to 0.4 |
| HoNOS | 9.23 (6.55) | 8.88 (6.53) | -0.36 | -2.7 to 2.0 |
| CANSAS-P unmet needs | 3.96 (3.58) | 4.10 (4.31) | 0.15 | -1.2 to 1.5 |
| CANSAS-P met needs | 4.39 (3.32) | 4.63 (4.71) | 0.25 | -1.1 to 1.6 |
| HAS-P | 7.37 (2.15) | 7.12 (2.38) | -0.25 | -1.0 to 0.5 |
| MANSA | 4.27 (1.04) | 4.20 (1.14) | -0.07 | -0.4 to 0.3 |
| BPRS | 31.39 (9.27) | 32.71 (11.39) | 1.3 | -2.2 to 4.8 |

CANSAS, Camberwell Assessment of Need Short Appraisal Schedule, -S, staff version, -P, patient version; TAG, Threshold Assessment Grid; HAS, Helping Alliance Scale, -S, staff version, -P, patient version; HoNOS, Health of the Nation Outcome Scale; MANSA, Manchester Short Assessment; BPRS, Brief Psychiatric Rating Scale.

was no evidence for differences between groups in mean follow-up patient-rated unmet need (mean difference 0.15, 95% CI -1.20 to 1.49, $P=0.83$). The sensitivity analyses all confirmed this conclusion.

There was no evidence for clustering because of staff (intraclass correlation 0.0) and a minimal impact for community mental health team (intraclass correlation 0.01).

For the 141 patients with baseline and follow-up quality-of-life data, 56 (40%) had a MANSA rating at least 0.25 higher at follow-up, comprising 39 (42%) out of 92 in the intervention group and 17 (35%) out of 49 in the control group. There was no evidence for differences between groups in mean follow-up quality of life (mean difference -0.07, 95% CI -0.44 to 0.31, $P=0.72$). The sensitivity analyses all confirmed this conclusion. Intraclass correlations were 0.078 for patients with the same staff member and 0.005 for patients belonging to the same community mental health team.

Secondary outcomes

There was no evidence for differences between groups for the three subjective secondary outcomes: mental health problem severity (mean difference -0.55, 95% CI -1.8 to 0.7, $P=0.38$), symptoms (mean difference 1.3, 95% CI -2.2 to 4.8, $P=0.46$) or social disability (mean difference -0.4, 95% CI -2.7 to 2.0, $P=0.46$). Service use is shown in Table 3.

Intervention-group patients had reduced hospital admissions, with admissions in the 6 months before follow-up being both fewer (means 0.13 *v.* 0.33, bootstrapped 95% CI -0.46 to -0.04) and tending to be shorter (mean 3.5 days *v.* 10.0 days, bootstrapped 95% CI -16.4 to 1.5). Criminal-justice service differences were owing to 121 days spent in prison by one intervention-group patient. Table 4 shows the cost of services used.

Total costs increased by an average of £1109 in the control group and fell by an average of £1928 in the intervention group. Follow-up costs were £2586 less for the intervention group. Most of the difference was owing to reduced in-patient costs and, after excluding these, the mean total cost difference was £338 less for the intervention group, which was not statistically significant (95% CI -£1500 to £731).

Net-benefit analysis indicated that if no value was placed on improved quality of life, the probability that the intervention was cost-effective would be approximately 0.98, and any positive value would raise this probability still higher. A positive value placed on a clinically significant reduction in unmet needs would reduce the probability of the intervention being cost-effective, as unmet needs were marginally less frequent in the control group. However, the value would need to approach £1 million

Table 3 Number of service contacts in 6-month periods before baseline and follow-up interviews

| Contact | Baseline ¹ | | Follow-up ¹ | | 95% CI of follow-up difference ² |
|-------------------------------|-----------------------|----------------------------|------------------------|---------------------------|---|
| | Control group (n=59) | Intervention group (n=101) | Control group (n=49) | Intervention group (n=93) | |
| Psychiatric in-patient | 10.3 (31.4) | 15.6 (37.4) | 16.4 (45.8) | 3.5 (16.1) | -25.7 to -1.6 |
| General in-patient | 1.9 (13.9) | 0.5 (2.9) | 0.8 (4.4) | 0.7 (5.1) | -2.2 to 0.2 |
| Accident and emergency | 0.4 (1.1) | 0.7 (2.1) | 0.4 (1.0) | 0.4 (1.4) | -0.4 to 0.3 |
| General out-patient | 1.0 (3.6) | 0.6 (1.5) | 0.6 (1.6) | 0.6 (2.3) | -0.7 to 0.7 |
| Day care | 14.2 (28.3) | 14.1 (30.2) | 7.1 (17.7) | 9.5 (30.4) | -5.2 to 10.5 |
| Community mental health nurse | 6.2 (7.4) | 9.3 (11.1) | 9.6 (12.9) | 9.6 (13.0) | -6.5 to 1.6 |
| Social worker | 2.5 (5.5) | 3.9 (9.4) | 2.4 (5.3) | 3.8 (10.5) | -1.3 to 3.5 |
| General practitioner | 2.5 (4.2) | 2.1 (3.2) | 2.8 (5.2) | 2.3 (4.5) | -1.7 to 1.2 |
| Psychiatrist | 3.9 (4.5) | 3.7 (4.7) | 3.8 (7.6) | 2.7 (4.0) | -3.4 to 1.0 |
| Psychologist | 1.0 (3.5) | 1.5 (5.2) | 1.5 (7.6) | 1.3 (4.6) | -2.7 to 1.6 |
| Occupational therapy | 4.1 (12.6) | 1.2 (4.3) | 4.7 (26.2) | 1.3 (10.8) | -9.2 to 2.1 |
| Criminal justice services | 0.7 (1.9) | 0.7 (2.9) | 0.0 (0.0) | 2.0 (14.0) | 0.3 to 5.8 |
| Residential care | 3.3 (8.2) | 3.2 (8.1) | 5.2 (10.2) | 3.3 (8.4) | -4.7 to 0.9 |

1. Figures are mean (s.d.) number of contacts, with the exception of in-patient care where the number of days are recorded.

2. Bootstrapped confidence interval using percentile method and controlling for baseline service use.

Table 4 Cost of services used in 6-month periods before baseline and follow-up interviews (2001–2002)

| Service | Baseline ¹ | | Follow-up ¹ | | 95% CI of follow-up difference ² |
|-------------------------------|-----------------------|----------------------------|------------------------|---------------------------|---|
| | Control group (n=59) | Intervention group (n=101) | Control group (n=49) | Intervention group (n=93) | |
| Psychiatric in-patient | 1824 (5558) | 2762 (6624) | 2893 (8100) | 626 (2847) | -4542 to -287 |
| General in-patient | 514 (3803) | 132 (789) | 206 (1196) | 176 (1396) | -596 to 42 |
| Accident and emergency | 32 (79) | 53 (155) | 28 (76) | 33 (101) | -31 to 26 |
| General out-patient | 81 (297) | 47 (127) | 52 (128) | 48 (187) | -55 to 60 |
| Day care | 562 (1324) | 476 (1114) | 177 (443) | 246 (763) | -106 to 275 |
| Community mental health nurse | 251 (395) | 367 (653) | 437 (751) | 325 (553) | -397 to 50 |
| Social worker | 173 (480) | 284 (957) | 98 (224) | 219 (699) | -30 to 285 |
| General practitioner | 77 (143) | 39 (46) | 59 (97) | 45 (82) | -17 to 24 |
| Psychiatrist | 533 (1342) | 412 (902) | 423 (787) | 296 (504) | -390 to 95 |
| Psychologist | 57 (216) | 88 (333) | 49 (194) | 78 (282) | -49 to 95 |
| Occupational therapy | 154 (683) | 33 (130) | 105 (551) | 74 (679) | -256 to 167 |
| Criminal justice services | 14 (45) | 21 (101) | 0 (0) | 152 (1296) | 4 to 467 |
| Residential care | 825 (2077) | 833 (2144) | 1678 (3523) | 900 (2334) | -1841 to 96 |
| Total (all services) | 5097 (7863) | 5548 (7431) | 6206 (9994) | 3620 (4095) | -5391 to -102 |

1. Figures are mean (s.d.) costs in pounds sterling.

2. Bootstrapped confidence interval using percentile method and controlling for baseline service use.

before there would be even a 60% chance that the control condition was more cost-effective. The cognitive and behavioural impacts of the intervention were investigated at follow-up, and are shown in Table 5.

Care plan audit indicated no difference between baseline and follow-up for direct care (possible range 0–10, intervention change 0, control change 0.7, difference in

change 0.7, 95% CI -0.1 to 1.5), planned assessments (range 0–4, intervention change 0.2, control change 0.2, difference -0.1, 95% CI -0.4 to 0.3), referrals (range 0–3, intervention change 0.0, control change 0.1, difference in change 0.1, 95% CI -0.3 to 0.5) and carer support (range 0–6, intervention change 0.5, control change 0.5, difference 0.0, 95% CI -0.6 to 0.6).

DISCUSSION

This randomised controlled trial evaluated the impact over 7 months of monthly assessment of important outcomes by staff and patients, plus feedback to both every 3 months. Routine outcome assessment was not shown to be effective, since means of the subjective outcomes were similar

Table 5 Intervention-group staff (n=81) and patient (n=85) assessment of validity of the model

| Question ¹ | Number (%) replying 'Yes' | |
|--|---------------------------|---------|
| | Staff | Patient |
| Did filling in the postal questionnaires make you think about the care the service user gets? | 72 (94) | 69 (81) |
| Did filling in the postal questionnaires make you think about your relationship with the service user? | 71 (92) | 60 (71) |
| Did you receive the feedback? | 70 (88) | 80 (94) |
| Did you read the feedback? | 69 (96) | 70 (85) |
| Did you understand the feedback? | 61 (88) | 69 (84) |
| Did receiving the feedback make you think about the care the service user is receiving? | 59 (82) | 52 (64) |
| Did receiving the feedback make you think about your relationship with the service user? | 60 (85) | 53 (65) |
| Did receiving the feedback lead you to discuss the content of their care with the service user? | 36 (51) | 26 (31) |
| Did receiving the feedback lead you to change your behaviour with the service user? | 30 (41) | 13 (16) |

I. In the patient version, wording was altered to refer to staff giving care.

across the two groups; it was, however, associated with cost savings, since patients receiving the intervention had fewer psychiatric admissions. Subjective outcomes appeared not to have changed, because the intervention was unsuccessful in promoting behaviour change.

Unchanged subjective outcomes

Subjective outcomes did not significantly improve, so the model did not accurately predict the impact of the intervention. On the basis of their self-report at follow-up, most staff and patients were prompted to consider the process and content of care both by completing the assessments and considering the feedback. However, self-report and care plan audits indicate that behaviour did not change as a result.

The intervention was not entirely implemented as planned, since the turnover of staff was high: 41 (26%) patients had a different member of staff at 7-month follow-up, including 29 (29%) from the intervention group. This may have invalidated some of the intended process-related mechanisms of action. Similarly, there was a progressive reduction in staff return rates, which may indicate a growing lack of enthusiasm if the feedback was not perceived as useful.

More generally, improvement in subjective outcomes may require greater attention

to the context of the intervention (Iles & Sutherland, 2001). Service staff whose shared beliefs are congruent with the use of outcome measures are necessary if the intervention is not to be swimming against the tide. This will involve changing organisational beliefs and working practices, setting up research programmes rather than isolated research studies, and demonstration sites (Nutley *et al*, 2003). A demonstration site in this context would be a service which uses outcome measures as a routine element of care on an ongoing basis. What would such a service look like? The characteristics of such a service would be a focus on the patient's perspective in assessment, the systematic identification of the full range of health and social care needs of the patient, the development of innovative services to address these needs, and the evaluation of the success of the service in terms of impact on quality of life.

The intervention also needs to be more tailored to fostering behaviour change – identifying topics which the patient would like to discuss with staff (van Os *et al*, 2004), or providing (and auditing for level of implementation) more prescriptive advice for staff action (Lambert *et al*, 2001). The feedback was provided every 3 months, which may have been too long a gap – feedback may need to be more prompt (Bickman *et al*, 2000; Lambert *et al*, 2001; Hodges & Wotring, 2004). However,

the objective criterion of admission rates did improve, and so some aspects of behaviour did change. This is considered below.

Reduced admissions

Why were admissions reduced? Reductions in in-patient use and costs may be caused by earlier or different action. Staff received regular clinical information about intervention patients, possibly triggering earlier support and hence avoiding the need for admission. This could be investigated by assessing whether the time between prodromal indications of relapse and keyworker awareness of the need for increased support is reduced when outcome information is routinely collected and available to staff.

Furthermore, staff had more information about intervention-group than control-group patients. Since decisions to admit patients are made using the best clinical information available, there may have been a marginal raising of the admission threshold for intervention patients. Further attention needs to be given to the influences which alter thresholds for in-patient admission.

Finally, the way in which the feedback is used by patients and staff needs to be investigated, for example using qualitative methods such as conversation analysis (McCabe *et al*, 2002).

Limitations

Service use data were obtained via patient self-report, which may be unreliable. However, a number of studies have found adequate correlation between self-report data and information collected by service providers (Caslyn *et al*, 1993; Goldberg *et al*, 2002).

Neither patients nor staff were masked to allocation status. Researchers conducting the follow-up interviews were partially masked – they guessed allocation status correctly for 38% of staff and for 68% of patients.

In the control group, 46 (78%) of the 59 patients had a member of staff who also had an intervention-group patient, indicating that contamination was possible between the two groups. A solution to contamination problems would have been cluster randomisation by the community mental health team. Cluster randomised controlled trials overcome some of the theoretical, ethical and practical problems of investigating mental health services (Gilbody & Whitty, 2002), although they are more complex to design and require larger samples and more complex analysis (Campbell *et al*, 2004). On the basis of intraclass

correlations in this study, a cluster trial randomising by community mental health team would require an increase of 20% in the sample size. Randomisation by staff member would entail an increase of 10%.

Finally, the follow-up period of 7 months may not have been long enough to capture all potential service use changes brought about by the intervention.

Implications for clinicians and policy makers

This study demonstrates that it is feasible to implement a carefully developed approach to routine outcome assessment in mental health services. The staff response rate over the 7 rounds of assessment was 67%, the patient response rate was 79%, and 92% of the intervention group received two rounds of feedback. Furthermore, 84% of staff and patients received, read and understood the feedback.

The intervention cost about £400 per person which, for a primary care trust with a case-load of 3500 people, would equate to about £1.4 million. However, the results of this study suggest that this cost could be more than offset by savings in service use.

This study is the first investigation of the use of standardised outcome measures over time in a representative adult mental health sample. As with previous studies (Ashaye *et al*, 2003; Marshall *et al*, 2004), subjective outcomes did not improve. However, a carefully developed and implemented approach to routinely collecting and using outcome data has been shown to reduce admissions and consequently save money.

ACKNOWLEDGEMENT

We thank Ian White, the trial statistician.

REFERENCES

- Ashaye, O. A., Livingston, G. & Orrell, M. W. (2003) Does standardized needs assessment improve the outcome of psychiatric day hospital care for older people? A randomized controlled trial. *Aging and Mental Health*, **7**, 195–199.
- Beecham, J. & Knapp, M. (2001) Costing psychiatric interventions. In *Measuring Mental Health Needs* (2nd edn) (ed. G. Thornicroft), pp. 200–224. London: Gaskell.
- Bickman, L., Rosof-Williams, J., Salzer, M., *et al* (2000) What information do clinicians value for monitoring adolescent client progress and outcomes? *Professional Psychology: Research and Practice*, **31**, 70–74.
- Bilsker, D. & Goldner, E. M. (2002) Routine outcome measurement by mental health-care providers: is it worth doing? *Lancet*, **360**, 1689–1690.
- Briggs, A. H. (2001) Bayesian approach to stochastic cost-effectiveness analysis. *International Journal of Technology Assessment in Health Care*, **17**, 69–82.
- Campbell, J., Fitzpatrick, R., Haines, A., *et al* (2000) Framework for design and evaluation of complex interventions to improve health. *BMJ*, **321**, 694–696.
- Campbell, M., Elbourne, D., Altman, D. (2004) CONSORT statement: extension to cluster randomised trials. *BMJ*, **328**, 702–708.
- Caslyn, R. J., Allen, G., Morse, G. A., *et al* (1993) Can you trust self-report data provided by homeless mentally ill individuals? *Evaluation Review*, **17**, 353–366.
- Gilbody, S. & Whitty, P. (2002) Improving the delivery and organisation of mental health services: beyond the conventional randomised controlled trial. *British Journal of Psychiatry*, **180**, 13–18.
- Gilbody, S. M., House, A. O. & Sheldon, T. A. (2001) Routinely administered questionnaires for depression and anxiety: systematic review. *BMJ*, **322**, 406–409.
- Gilbody, S. M., House, A. O. & Sheldon, T. A. (2002a) Psychiatrists in the UK do not use outcomes measures. *British Journal of Psychiatry*, **180**, 101–103.
- Gilbody, S. M., House, A. O. & Sheldon, T. A. (2002b) Routine administration of Health Related Quality of Life (HRQoL) and needs assessment instruments to improve psychological outcome – a systematic review. *Psychological Medicine*, **32**, 406–409.
- Goldberg, R. W., Seybolt, D. C. & Lehman, A. (2002) Reliable self-report of health service use by individuals with serious mental illness. *Psychiatric Services*, **53**, 879–881.
- Health Research Council of New Zealand (2003) *Partnerships for Evidence-Based Policy and Practice*. Auckland: HRCNZ.
- Hodges, K. & Wotring, J. (2004) The role of monitoring outcomes in initiating implementation of evidence-based treatments at the state level. *Psychiatric Services*, **55**, 396–400.
- Iles, V. & Sutherland, K. (2001) *Organisational Change*. London: National Co-ordinating Centre for Service Development and Organisation.
- Lambert, M. J., Hansen, N. B. & Finch, A. (2001) Patient focused research: using patient outcome data to enhance treatment effects. *Journal of Consulting and Clinical Psychology*, **69**, 159–172.
- Marshall, M., Lockwood, A., Green, G., *et al* (2004) Systematic assessments of need and care planning in severe mental illness. Cluster randomised controlled trial. *British Journal of Psychiatry*, **185**, 163–168.
- McCabe, R., Roeder-Wanner, U.-U., Hoffman, K., *et al* (1999) Therapeutic relationships and quality of life: association of two subjective constructs in schizophrenia patients. *International Journal of Social Psychiatry*, **45**, 276–283.
- McCabe, R., Heath, C., Burns, T., *et al* (2002) Engagement of patients with psychosis in the consultation. *BMJ*, **325**, 1148–1151.
- Netten, A. & Curtis, L. (2002) *Unit Costs of Health and Social Care*. Canterbury: Personal Social Services Research Unit.
- Nutley, S., Percy-Smith, J. & Solesbury, W. (2003) *Models of Research Impact*. London: Learning and Skills Research Centre.
- Overall, J. E. & Gorham, D. R. (1988) The Brief Psychiatric Rating Scale (BPRS): recent developments in ascertainment and scaling. *Psychopharmacology Bulletin*, **24**, 97–99.
- Priebe, S., Huxley, P., Knight, S., *et al* (1999) Application and results of the Manchester Short Assessment of quality of life. *International Journal of Social Psychiatry*, **45**, 7–12.
- Slade, M. (2002a) What outcomes to measure in routine mental health services, and how to assess them – a systematic review. *Australian and New Zealand Journal of Psychiatry*, **36**, 743–753.
- Slade, M. (2002b) Routine outcome assessment in mental health services. *Psychological Medicine*, **32**, 1339–1344.
- Slade, M., Thornicroft, G., Loftus, L., *et al* (1999) *CAN: The Camberwell Assessment of Need. A Comprehensive Needs Assessment Tool for People with Severe Mental Illness*. London: Gaskell.
- Slade, M., Powell, R., Rosen, A., *et al* (2000) Threshold Assessment Grid (TAF): the development of a valid and brief scale to assess the severity of mental illness. *Social Psychiatry and Psychiatric Epidemiology*, **35**, 78–85.
- Slade, M., Leese, M., Cahill, S., *et al* (2005) Patient-rated mental health needs and quality of life improvement. *British Journal of Psychiatry*, **187**, 256–261.
- Thornicroft, G., Strathdee, G., Phelan, M., *et al* (1998) Rationale and design: PRISM Psychosis Study I. *British Journal of Psychiatry*, **173**, 363–370.
- Trauer, T. (2003) Routine outcome measurement by mental health-care providers. *Lancet*, **361**, 1137.
- Valenstein, M., Mitchinson, A., Ronis, D. L., *et al* (2004) Quality indicators and monitoring of mental health services: what do frontline providers think? *American Journal of Psychiatry*, **161**, 146–153.
- van Os, J., Altamura, A. C., Bobes, J., *et al* (2004) Evaluation of the Two-Way Communication Checklist as a clinical intervention. Results of a multinational, randomised controlled trial. *British Journal of Psychiatry*, **184**, 79–83.
- Wing, J. K., Beevor, A. S., Curtis, R. H., *et al* (1998) Health of the Nation Outcome Scales (HoNOS). Research and development. *British Journal of Psychiatry*, **172**, 11–18.

Correspondence: Dr Mike Slade, Health Services Research Department (Box P029), Institute of Psychiatry, King's College London, London SE5 8AF, UK. Tel.: +44 (0) 20 7848 0795; fax: +44 (0) 20 7277 1462; email: m.slade@iop.kcl.ac.uk

(First received 18 July 2005, final revision 31 October 2005, accepted 6 December 2005)