Collaborative care for depression in UK primary care: a randomized controlled trial

D. A. Richards[1]*, K. Lovell[2], S. Gilbody[1], L. Gask[3], D. Torgerson[1], M. Barkham[4], M. Bland[1], P. Bower[3], A. J. Lankshear[5], A. Simpson[1], J. Fletcher[2], D. Escott[2], S. Hennessy[1] and R. Richardson[1]

[1] *Department of Health Sciences, University of York, UK*
[2] *Department of Nursing, Midwifery and Social Work, University of Manchester, UK*
[3] *National Primary Care Research and Development Centre, University of Manchester, UK*
[4] *Centre for Psychotherapy Services Research, University of Sheffield, UK*
[5] *School of Nursing and Midwifery Studies, University of Cardiff, UK*

Background. Collaborative care is an effective intervention for depression which includes both organizational and patient-level intervention components. The effect in the UK is unknown, as is whether cluster- or patient-randomization would be the most appropriate design for a Phase III clinical trial.

Method. We undertook a Phase II patient-level randomized controlled trial in primary care, nested within a cluster-randomized trial. Depressed participants were randomized to 'collaborative care' – case manager-coordinated medication support and brief psychological treatment, enhanced specialist and GP communication – or a usual care control. The primary outcome was symptoms of depression (PHQ-9).

Results. We recruited 114 participants, 41 to the intervention group, 38 to the patient-randomized control group and 35 to the cluster-randomized control group. For the intervention compared to the cluster control the PHQ-9 effect size was 0.63 (95% CI 0.18–1.07). There was evidence of substantial contamination between intervention and patient-randomized control participants with less difference between the intervention group and patient-randomized control group (-2.99, 95% CI -7.56 to 1.58, $p=0.186$) than between the intervention and cluster-randomized control group (-4.64, 95% CI -7.93 to -1.35, $p=0.008$). The intra-class correlation coefficient for our primary outcome was 0.06 (95% CI 0.00–0.32).

Conclusions. Collaborative care is a potentially powerful organizational intervention for improving depression treatment in UK primary care, the effect of which is probably partly mediated through the organizational aspects of the intervention. A large Phase III cluster-randomized trial is required to provide the most methodologically accurate test of these initial encouraging findings.

Introduction

Depression is a major health problem causing substantial disability and set to become the second largest cause of global disability by 2020 (WHO, 2001). In the UK depression and anxiety are estimated to cost the economy £17bn in lost output with a £9bn impact on the Exchequer through benefit payments and lost tax receipts (Layard, 2006). Only one in four depressed people receive effective pharmacological treatment and less than 10% a talking therapy (Singleton *et al.* 2001). Bower & Gilbody (2005) have identified four types of organizational strategies to improve this

situation: training and the use of guidelines for general practitioners and primary care staff; 'consultation liaison', whereby specialist mental health practitioners advise on the care of individual patients in primary care; 'collaborative care', an enhanced form of consultation liaison which also includes a case manager to deliver care and liaise between GP, specialist and patient; and 'replacement referral' which refers to the deployment of specialists in secondary or primary care to whom GPs can refer.

A systematic review (Gilbody *et al.* 2003) of 36 such organizational intervention studies concluded that effective strategies require complex interventions at the 'systems level', consisting of: (1) a multi-professional approach to patient care; (2) a structured patient management plan; (3) scheduled patient follow-ups; and (4) enhanced inter-professional communication (Wagner *et al.* 1996; Gunn *et al.* 2006). The most

* Address for correspondence: Professor D. A. Richards, Department of Health Sciences, Seebohm Rowntree Building, University of York, University Road, Heslington, York, North Yorkshire, United Kingdom.

(Email: dr17@york.ac.uk)

effective systems-level intervention in this review was 'collaborative care' (Von Korff & Goldberg, 2001; Simon, 2006).

Although collaborative care improves outcomes over usual care (Katon *et al.* 1999; Wells *et al.* 2000; Unutzer *et al.* 2002), two recent systematic reviews found small to medium mean effect sizes of either 0.24 (95% CI 0.17–0.32) (Gilbody *et al.* 2006) or 0.40 (95% CI 0.20–0.60) (Gensichen *et al.* 2005). The effects associated with individual studies varied significantly, reflecting variation in the content of these 'complex' interventions (MRC, 2000). Further, most of the studies originated from the USA (Gilbody *et al.* 2006).

Although there have been calls for the implementation of collaborative care in the UK (Simon, 2006), these have not been supported by UK clinical guidelines (National Institute for Health and Clinical Excellence, 2004) and may be premature given that it is not known exactly which models of collaborative care work best, and whether the model will generalize to the UK. In other areas of mental health such as assertive community treatment (Killaspy *et al.* 2005), the adoption of complex interventions based on international/US data before UK evaluation has resulted in ineffective UK service developments.

We adopted the phased approach (Campbell *et al.* 2000) recommended by the Medical Research Council (MRC) for investigating complex interventions (MRC, 2000). We developed a UK-specific collaborative care intervention for depression, based on analysis of 'active ingredients' in published interventions (Gilbody *et al.* 2006) and in-depth qualitative research with stakeholders (Richards *et al.* 2006*a*). We then tested it in an exploratory Phase II randomized controlled trial.

Cluster-randomized controlled trials are recommended for testing systems-level interventions such as collaborative care (Ukoumunne *et al.* 1999), since patient-randomized trials may be vulnerable to contamination. Patients in the control group may be influenced by system-level changes such as advice from specialists and changes to the process of care. Contamination in a patient-randomized trial may result in underestimating the real effect size of collaborative care. However, cluster-randomized trials require larger patient samples and often greater resources. Our Phase II trial, therefore, used an unusual design, nesting patient-level randomization within a cluster-randomized controlled trial to investigate the presence and magnitude of contamination (Fig. 1). We report the results of that randomized controlled trial here, the first UK trial of collaborative care and an early test of the utility of the MRC's complex-interventions framework.

Ethical approval was given by the South West Multi-site Research Ethics Committee.

Objectives

The objectives were:

(1) To estimate an effect size for a UK-specific collaborative care protocol.
(2) To determine whether cluster or patient randomization would be the most appropriate design for a Phase III trial.

Method

General practice sites were randomly allocated to treatment or cluster control conditions from four primary-care trusts (PCT) in the northern UK, stratified by PCT. Almost all practices had a deprivation index higher than the UK national average and a number were from areas where black and minority ethnic groups were strongly represented. Patients in the treatment cluster group were then individually randomized to either collaborative care or usual care control. Allocation was by a remote computer-generated number sequence concealed from researchers and conducted independently after patients were enrolled in the study by research interviewers. The randomization team at the trials unit informed patients, GPs and, where appropriate, case managers, of participant allocation. This created three study groups (cluster-randomized controls, individually randomized intervention patients, and individually randomized control patients). Fig. 2 details the consort diagram. To try to reduce the possibility of recruitment bias GPs were given no information about the allocation of their practice.

Sample

We recruited patients from primary care aged >18 years diagnosed as depressed by a GP, confirmed by a score of ⩾5 on the depression section of the Standard Clinical Interview for DSM-IV (SCID; Spitzer *et al.* 1992) undertaken by trained research assistants. We excluded patients with post-natal, bereavement or physical causes for their depression. We only included patients with a newly identified episode of major depression, defined as a current episode of GP-initiated treatment of not more than 1 month's duration. We excluded patients reporting active suicidal plans and those with a primary drug or alcohol dependence. Although a standard power calculation based on detecting treatment effects would be the conventional approach to determining the sample size for a clinical trial, our Phase II study was designed to help inform

The collaborative care intervention for depression consists of two aspects:

- A component **A** that impacts on the individual patient (e.g. medication management from the case manager).
- A component **B** that impacts on the general practice (e.g. feedback of patient information to the GP).

In a standard, individually randomized trial, component **A** only influences patients who are randomly allocated to receive it, whereas component **B** may potentially influence all patients who are under the care of a particular practice. Contamination occurs if the outcome of patients randomized to the control group is influenced by **B**, even though those patients receive no benefit from **A**.

In the design adopted in this study, a patient level randomization study is nested within a cluster level randomized study.

```
                    ┌──────────────────────────┐
                    │  Cluster randomization of │
                    │         practices         │
                    └──────────────────────────┘
              ┌────────────────┴────────────────────┐
    ┌──────────────────┐                    ┌──────────────────┐
    │   Intervention   │                    │     Control      │
    │    practices     │                    │    practices     │
    └──────────────────┘                    └──────────────────┘
              │                                       │
    ┌──────────────────┐                              │
    │    Individual    │                              │
    │ randomization of │                              │
    │     patients     │                              │
    └──────────────────┘                              │
        ┌─────────┴──────────┐                        │
┌───────────────┐  ┌────────────────┐  ┌──────────────────┐
│   Group 1     │  │    Group 2     │  │     Group 3      │
│ Individually  │  │  Individually  │  │     Cluster      │
│  randomized   │  │   randomized   │  │    randomized    │
│ intervention  │  │    control     │  │  control patients│
│   patients    │  │    patients    │  │                  │
│               │  │                │  │   Receive no     │
│  Receive A+B  │  │   Receive B    │  │   intervention   │
└───────────────┘  └────────────────┘  └──────────────────┘
```
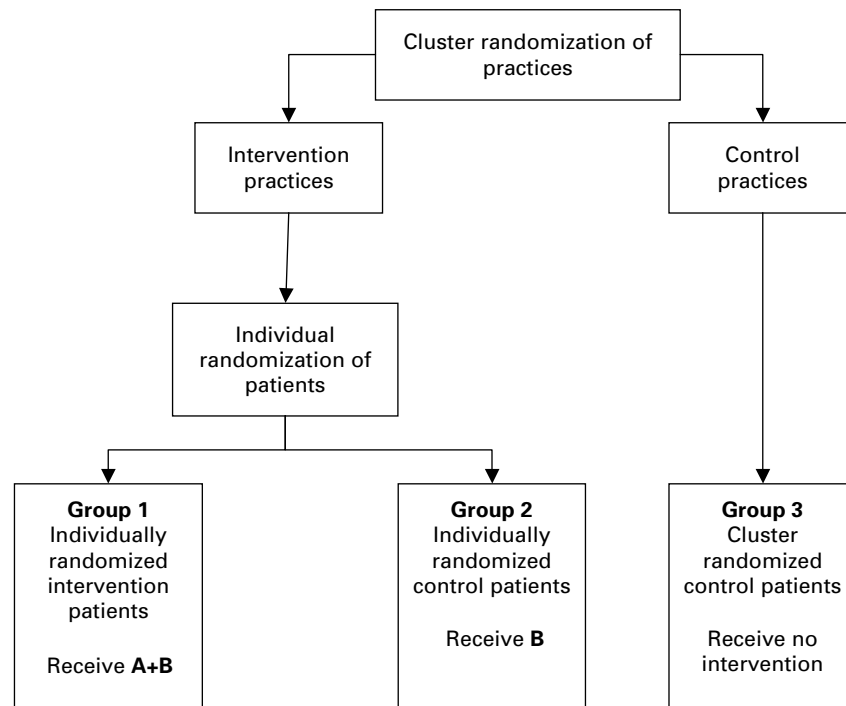
This has the advantage of allowing two analyses:

- The analysis of Group 1 *versus* Group 3 compares individually randomized intervention patients (who have received the individual patient component **A** and the practice component **B**) with cluster-randomized control patients (who have received neither **A** nor **B**). This analysis therefore provides the best overall indication of the effect of the entire collaborative care intervention (i.e. **A+B**).
- The analysis of Group 2 *versus* Group 3 compares individually randomized control patients (who have received the practice component **B**) with cluster-randomized control patients (who have received neither **A** nor **B**). This analysis therefore provides the best overall indication of the effect of the practice component (i.e. **B**). As noted above, this component **B** represents potential contamination in an individually randomised trial comparing Group 1 and Group 2.

Fig. 1. Investigating contamination in a trial of a complex organizational intervention.

the overall power calculation for a definitive Phase III trial to be used together with estimates of the '*expected treatment effect*' (MRC, 2000) from published studies to allow us to estimate a plausible effect size for use in the sample size calculation for the definitive Phase III

trial. Recruiting 32 patients in each of the patient-randomized groups within the treatment cluster would have given us a 95% confidence interval width of half an effect size, allowing us to test for an effect of 0.25 in either direction. Therefore, we aimed to recruit
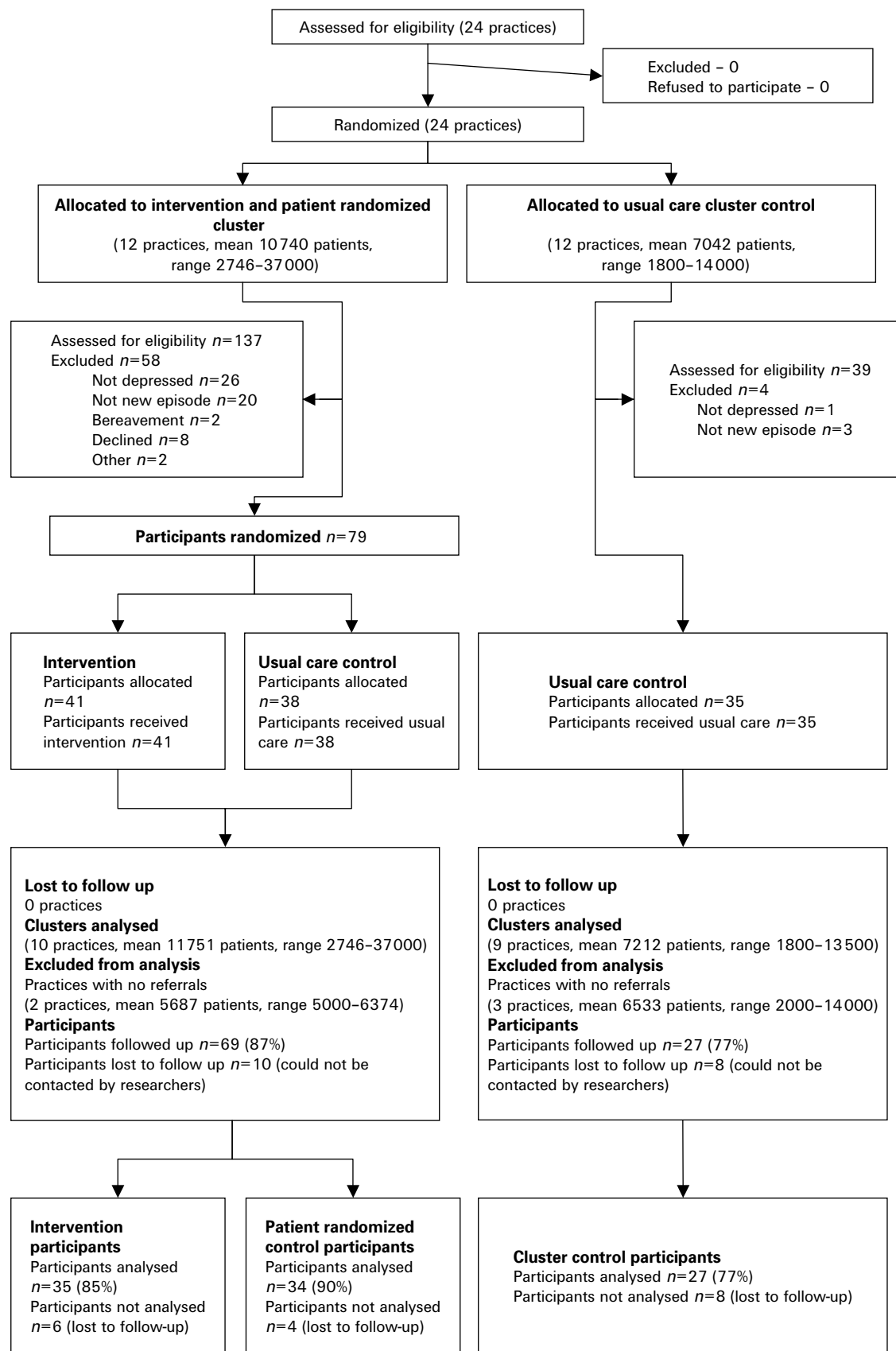
Assessed for eligibility (24 practices)

Excluded – 0
Refused to participate – 0

Randomized (24 practices)

Allocated to intervention and patient randomized cluster
(12 practices, mean 10 740 patients, range 2746–37 000)

Allocated to usual care cluster control
(12 practices, mean 7042 patients, range 1800–14 000)

Assessed for eligibility *n*=137
Excluded *n*=58
    Not depressed *n*=26
    Not new episode *n*=20
    Bereavement *n*=2
    Declined *n*=8
    Other *n*=2

Assessed for eligibility *n*=39
Excluded *n*=4
    Not depressed *n*=1
    Not new episode *n*=3

Participants randomized *n*=79

Intervention
Participants allocated *n*=41
Participants received intervention *n*=41

Usual care control
Participants allocated *n*=38
Participants received usual care *n*=38

Usual care control
Participants allocated *n*=35
Participants received usual care *n*=35

Lost to follow up
0 practices
Clusters analysed
(10 practices, mean 11 751 patients, range 2746–37 000)
Excluded from analysis
Practices with no referrals
(2 practices, mean 5687 patients, range 5000–6374)
Participants
Participants followed up *n*=69 (87%)
Participants lost to follow up *n*=10 (could not be contacted by researchers)

Lost to follow up
0 practices
Clusters analysed
(9 practices, mean 7212 patients, range 1800–13 500)
Excluded from analysis
Practices with no referrals
(3 practices, mean 6533 patients, range 2000–14 000)
Participants
Participants followed up *n*=27 (77%)
Participants lost to follow up *n*=8 (could not be contacted by researchers)

Intervention participants
Participants analysed *n*=35 (85%)
Participants not analysed *n*=6 (lost to follow-up)

Patient randomized control participants
Participants analysed *n*=34 (90%)
Participants not analysed *n*=4 (lost to follow-up)

Cluster control participants
Participants analysed *n*=27 (77%)
Participants not analysed *n*=8 (lost to follow-up)

Fig. 2. Consort diagram.

144 patients in total between the intervention and control clusters, 32 of the extra patients compensating for the design effect, the remainder to account for any attrition.

Intervention

Experimental

We developed a UK-specific collaborative care protocol in the modelling phase of our complex-interventions trial (Gilbody *et al.* 2006; Richards *et al.* 2006*a*), which met the four criteria for an organizational, systems-level intervention (Wagner *et al.* 1996; Gunn *et al.* 2006): (1) a *multi-professional approach to patient care* provided by a case manager working with the GP under weekly telephone supervision from specialist mental health medical and psychological therapies clinicians; (2) a *structured management plan* of medication support and behavioural activation – a structured cognitive-behaviourally based, depression-specific psychological intervention which has equivalent efficacy to other more complex CBT interventions (Dimidjian *et al.* 2006; Cuijpers *et al.* 2007) but is simpler to use and thus more suitable for collaborative care (Jacobson *et al.* 1996; Martell *et al.* 2001). No other interventions were permitted for the duration of the trial; (3) *scheduled patient follow-ups* via a maximum of ten scheduled contacts over a period of 3 months, predominantly using the telephone; (4) *enhanced inter-professional communication* patient-specific written feedback to GPs via electronic records and personal contact. Case managers were a mix of professionals (nurse, counsellor and occupational therapist) and para-professionals (graduate primary-care mental health workers) all of whom received 2 days of protocol-specific training in addition to their existing clinical training and 30–45 min of supervision per week for the duration of the trial.

Control

Usual care management of depression by patients' GPs, including access to secondary services, and to best practice guidance published in local NHS depression protocols in the trial localities.

Outcome measures

The primary outcome was symptoms of depression as measured by the Patient Health Questionnaire-9 (PHQ-9; Kroenke *et al.* 2001). Secondary outcomes were the Clinical Outcomes in Routine Evaluation – Outcome Measure (CORE-OM; Barkham *et al.* 2001), measuring general wellbeing; the Short Form Health Survey (SF-36 v.2; Ware *et al.* 2000), measuring health-related quality of life. All assessments were completed at baseline and 3 months post-randomization by trained assessors blind to participant allocation.

Analysis

We aimed to determine a point estimate of the effect size of collaborative care, specific to the UK primary-care setting. We conducted analysis of covariance accounting for baseline imbalances in depression scores and clustering within the units of randomization using the Huber–White sandwich estimator (White, 1980) within Stata 8 (Stata Corp., College Station, TX, USA). We used an intention-to-treat approach to examine the mean differences between the three groups and the associated confidence intervals and calculated coefficients to represent the difference between the cluster-randomized controls and the intervention group in follow-up outcome measures. For the main analysis of the effectiveness of the intervention, we calculated the standardized effect size (mean difference divided by the pooled standard deviation) between the intervention and cluster-randomized control groups. We examined the degree of contamination by comparing the coefficients of individually randomized and cluster-randomized control groups. We examined clustering of outcomes within practices by calculating the intra-class correlation coefficient (ICC).

Results

We recruited 114 patients, 41 to the intervention group, 38 to the patient randomized control group and 35 to the cluster-randomized control group (Fig. 2) from February 2005 until March 2006. Table 1 details the sample characteristics. The average number of case manager/patient contacts was 6.46 (S.D. = 1.69), taking a mean time per patient of 191.13 min (S.D. = 70.68).

We found an effect size on PHQ-9 depression symptoms of 0.63 (95% CI 0.18–1.07) for the intervention compared to the cluster control (Table 2). We found the intervention to be more effective than the cluster control on the CORE-OM (0.45, 95% CI 0.11–1.01) and the mental component score of the SF-36 (0.67, 95% CI 0.19–1.16) but not more effective on the physical component score of the SF-36 (0.11, 95% CI −0.49 to 0.72). No adverse events were reported in any group.

Evidence for substantial contamination was observed, as there was less difference in PHQ-9 depression outcomes between the intervention group and patient-randomized control group (−2.99, 95% CI −7.56 to 1.58, $p = 0.186$) than between the intervention group and cluster-randomized controls (coefficient −4.64, 95% CI −7.93 to −1.35, $p = 0.008$). The ICC for our primary outcome was 0.06 (95% CI 0.00–0.32).

Table 1. *Baseline characteristics of sample*

	Intervention (*n* = 41)	Patient-randomized control (*n* = 38)	Cluster-randomized control (*n* = 35)	Total (*n* = 114)
Gender				
Female	32 (78%)	29 (76%)	27 (77%)	88 (77%)
Male	9 (22%)	9 (24%)	8 (23%)	26 (23%)
Age (yr)				
Mean (S.D.)	42.63 (12.33)	43.00 (14.88)	41.71 (13.99)	42.47 (13.62)
Range	20–72	18–77	19–71	18–77
Ethnic origin				
White British	38 (93%)	33 (87%)	27 (77%)	98 (86%)
Other	3 (7%)	5 (13%)	8 (23%)	16 (14%)
Education				
None	9 (22%)	9 (24%)	10 (29%)	28 (25%)
GCSE/0 Level	11 (27%)	9 (24%)	13 (37%)	33 (29%)
Post-GCSE/0 level	16 (39%)	15 (39%)	9 (26%)	40 (35%)
Degree or higher	3 (7%)	4 (11%)	3 (9%)	10 (9%)
Other or 'don't know'	2 (5%)	1 (3%)	0	3 (3%)
Employment status				
Employed/self-employed	24 (59%)	14 (40%)	17 (49%)	56 (49%)
Not working	17 (41%)	24 (60%)	18 (51%)	58 (51%)
Marital status				
Married/cohabiting	26 (63%)	25 (66%)	10 (29%)	61 (54%)
Baseline score means (S.D.)				
PHQ-9	17.51 (4.90)	16.29 (4.47)	18.17 (5.58)	17.31 (5.00)
SCL-20	47.34 (12.15)	43.84 (12.38)	47.85 (14.60)	46.34 (13.02)
CORE-OM	2.02 (0.58)	1.88 (0.61)	2.12 (0.55)	2.00 (0.58)
SF-36 MCS	19.06 (11.42	20.33 (10.19)	18.64 (10.98)	19.36 (10.80)
SF-36 PCS	50.80 (10.88)	50.99 (11.05)	49.2 (14.18)	50.38 (11.94)
EQ5D	0.55 (0.26)	0.56 (0.30)	0.43 (0.32)	0.52 (0.29)

PHQ-9, Patient Health Questionnaire-9; SCL-20, Hopkins Symptoms Checklist Depression Scale; CORE-OM, Clinical Outcomes in Routine Evaluation – Outcome Measure; SF-36 MCS, Short Form Health Survey mental component score; SF-36 PCS, Short Form Health Survey physical component score; EQ5D, EuroQuol 5 Dimension Scale.

Discussion

We found a moderate to large effect (Cohen, 1988; Lipsey & Wilson, 2001) of collaborative care, an effect which would be considered clinically significant under the guidelines for depression produced by the National Institute for Health and Clinical Excellence (2004), the first time this has been demonstrated in the UK. This effect is greater than that determined by systematic reviews (Gensichen *et al.* 2005; Gilbody *et al.* 2006) and equates to a mean difference between treated and usual care patients of 5 points on the PHQ-9. Five points is the difference between symptoms of mild or moderate/severe intensity and between symptoms of moderate/severe and severe intensity. Furthermore, change in PHQ-9 scores achieved by the intervention patients from baseline to follow-up equates to a clinical shift of almost two categories of depression severity.

The optimal design for a full Phase III trial has also been clarified by the phased approach. We detected clear evidence of contamination, which has not always been accounted for in previous collaborative care studies (Gilbody *et al.* 2003). The individually randomized controls were closer to the intervention patients than to the cluster-randomized controls. Although striking, the precise mechanism of the contamination is unclear, but may relate to the sharing of information between case manager and GP. Examination of this mechanism may prove fruitful for the further development of interventions. As in any cluster-randomized trial, however, unmeasured differences between doctors could also explain some differences in outcomes, although we were careful to stratify cluster randomization by our four clinical sites to protect against this source of bias.

The main limitation of this study is the relatively small numbers in what was a Phase II trial.

Table 2. *Follow-up scores, coefficients of difference between intervention and patient-randomized control group with cluster controls, 95% CI and p values of the difference*

	Intervention	Patient-randomized control	Cluster-randomized control
PHQ-9	*n*=35	*n*=34	*n*=27
Mean (S.D.)	8.80 (7.02)	10.27 (7.51)	13.82 (8.32)
Coefficient	−4.64	−2.99	N.A.
95% CI	−7.93 to −1.35	−7.56 to 1.58	
p value	0.008	0.186	
CORE-OM	*n*=39	*n*=38	*n*=32
Mean (S.D.)	2.02 (0.58)	1.88 (0.61)	2.12 (0.55)
Coefficient	0.34	−0.24	N.A.
95% CI	−0.76 to 0.08	−0.69 to 0.20	
p value	0.109	0.268	
SF-36 MCS	*n*=39	*n*=37	*n*=33
Mean (S.D.)	19.06 (11.42)	20.33 (10.19)	18.64 (10.98)
Coefficient	9.89	2.95	N.A.
95% CI	2.79 to 17.00	−6.00 to 11.92	
p value	0.009	0.497	
SF-36 PCS	*n*=39	*n*=37	*n*=33
Mean (S.D.)	50.80 (10.88)	50.99 (11.05)	49.2 (14.18)
Coefficient	0.91	1.49	N.A.
95% CI	−3.89 to 5.72	−3.02 to 6.00	
p value	0.694	0.497	

PHQ-9, Patient Health Questionnaire-9; CORE-OM, Clinical Outcomes in Routine Evaluation – Outcome Measure; SF-36 MCS, Short Form Health Survey mental component score; SF-36 PCS, Short Form Health Survey physical component score.

Consequently the results have wide confidence intervals around the mean and the effect size we obtained requires confirmation in a full trial. Although we did control for baseline depression severity as a covariate in our analysis, small numbers also prevented us balancing potentially important variables such as ethnicity, marital status and gender in our randomization which may or may not have influenced our results. Further, although there were no differences in consultations with GPs between groups, we do not have full detailed information on what constituted usual care in the control groups, which may have varied substantially and affected our results. However, Phase II trials are an important stage in carefully developing and testing new interventions and these results lend support to the utility of the MRC's complex-interventions research framework (MRC, 2000), which provided a logical and systematic structure to help us in the process of designing and testing collaborative care in the UK. For example, during our development work, we found four previous UK studies of collaborative care which in contrast to our results produced no or inconclusive effects (Wilkinson *et al.* 1993; Blanchard *et al.* 1995; Mann *et al.* 1998; Peveler *et al.* 1999). These studies were early trials in the development of collaborative care, and had not used the systematic framework to develop their interventions; our review indicates they had used suboptimal intervention ingredients (Gilbody *et al.* 2006). Although the principle of carefully phased intervention development is an effective way to think about designing interventions and is supported by our results, the framework is not prescriptive and lacks close detail. Our specific approach is only one of many methodological possibilities.

The research implications are that a fully powered Phase III cluster-randomized trial should be the next step of the MRC's complex-interventions phased approach (Campbell *et al.* 2000; MRC, 2000) to investigating this complex intervention. Such a design will provide the best protection against both over- and underestimating the real effect size of collaborative care in the UK and will allow us to achieve a better balance of baseline demographic characteristics. A parallel qualitative investigation to this trial (Richards *et al.* 2006*b*) has shown the clinical procedures to be acceptable to patients, mental health workers and GPs. If such a trial were to confirm the effect size of our Phase II trial results, we will have evidence to enable the NHS to substantially improve the organization of its care for depressed patients in primary care and to assist primary-care providers to deliver an effective

model of enhanced depression service within the GP contract.

Acknowledgements

Declaration of Interest

None.

References

Barkham M, Margison F, Leach C, Lucock M, Mellor-Clark J, Evans C, Benson, L, Connell J, Audin K, McGrath G (2001). Service profiling and outcomes benchmarking using the CORE-OM: towards practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology* **69**, 184–196.

Blanchard M, Waterreus A, Mann A (1995). The effect of primary care nurse intervention upon older people screened as depressed. *International Journal of Geriatric Psychiatry* **10**, 289–298.

Bower P, Gilbody S (2005). Managing common mental health disorders in primary care: conceptual models and evidence base. *British Medical Journal* **330**, 839–842.

Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, Tyrer P (2000). Framework for design and evaluation of complex interventions to improve health. *British Medical Journal* **321**, 694–696.

Cohen J (1988). *Statistical Power Analysis for the Behavioral Sciences*. Erlbaum: Hillsdale, NJ.

Cuijpers P, van Straten A, Warmerdam L (2007). Behavioural activation treatments of depression: a meta analysis. *Clinical Psychology Review* **27**, 318–326.

Dimidjian S, Hollon S, Dobson K, Schmaling KB, Kohlenberg RJ, Addis ME, Gallo R, McGlinchey JB, Markley DK, Gollan JK, Atkins DC, Dunne, DL, Jacobson NS (2006). Randomized trial of behavioural activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology* **74**, 658–670.

Gensichen J, Beyer M, Muth C, Gerlach FM, Von Korff M, Ormel J (2005). Case management to improve major depression in primary health care: a systematic review. *Psychological Medicine* **36**, 7–14.

Gilbody S, Bower P, Fletcher J, Richards DA, Sutton A (2006). Collaborative care for depression: a systematic review and cumulative meta-analysis. *Archives of Internal Medicine* **166**, 2314–2321.

Gilbody S, Whitty P, Grimshaw J, Thomas R (2003). Educational and organizational interventions to improve the management of depression in primary care: a systematic review. *Journal of the American Medical Association* **289**, 3145–3151.

Gunn J, Diggens J, Hegarty K, Blashki G (2006). A systematic review of complex system interventions designed to increase recovery from depression in primary care. *BMC Health Services Research* **6**, 88.

Jacobson N, Dobson K, Traux P, Addis ME, Koerner K, Gollan JK, Gortner E, Prince SE (1996). A component analysis of cognitive-behavioural treatment of depression. *Journal of Consulting and Clinical Psychology* **64**, 295–304.

Katon W, Von Korff M, Lin E, Simon G, Walker E, Unützer J, Bush T, Russo J, Ludman E (1999). Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. *Archives of General Psychiatry* **56**, 1109–1115.

Killaspy H, Bebbington P, Blizard R, Johnson S, McCrone P, Nolan F, Pilling S, King M (2005). The REACT study: a Randomised Evaluation of Assertive Community Treatment in North London. *British Medical Journal* **331**, 599–604.

Kroenke K, Spitzer RL, Williams JB (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine* **16**, 606–613.

Layard R (2006). The case for psychological treatment centres. *British Medical Journal* **332**, 1030–1032.

Lipsey M, Wilson D (2001). *Practical Meta-Analysis*. Sage: New York.

Mann A, Blizard R, Murray J, Smith JA, Botega N, MacDonald E, Wilkinson G (1998). An evaluation of practice nurses working with general practitioners to treat people with depression. *British Journal of General Practice* **48**, 875–879.

Martell CR, Addis ME, Jacobson NS (2001). *Depression in Context: Strategies for Guided Action*. Norton: New York.

MRC (2000). *A Framework for Development and Evaluation of RCTs for Complex Interventions to Improve Health*. Medical Research Council: London.

National Institute for Health and Clinical Excellence (2004). *Depression: Management of Depression in Primary and Secondary Care* (Clinical guideline 23). NICE: London.

Peveler R, George C, Kinmonth A, Campbell M, Thompson C (1999). Effect of antidepressant drug counselling and information leaflets on adherence to drug treatment in primary care: randomised controlled trial. *British Medical Journal* **319**, 612–615.

Richards DA, Barkham M, Bower P, Gask L, Gilbody S, Lovell K, Rogers A, Torgerson D, Escott D, Fletcher J, Hennessy S, Kendall S, Lankshear AJ, Richardson R, Simpson A (2006*b*). A Trial Platform of Enhanced Care for Depression in Primary Care: Final Report. University of York: York.

Richards DA, Lankshear AJ, Fletcher J, Rogers A, Barkham M, Bower P, Gask L, Gilbody S, Lovell K (2006*a*). Developing a UK protocol for collaborative care: a qualitative study. *General Hospital Psychiatry* **28**, 296–305.

Simon G (2006). Collaborative care for depression. *British Medical Journal* **332**, 249–250.

Singleton N, Bumpstead R, O'Brien M, Lee A, Meltzer HY (2001). *Office of National Statistics: Psychiatric Morbidity Among Adults Living in Private Households*. HMSO: London.

Spitzer RL, Williams JB, Gibbon M, First MB (1992). The Structured Clinical Interview for DSM-III-R (SCID). I: History, rationale, and description. *Archives of General Psychiatry* **49**, 624–629.

Ukoumunne OC, Gulliford MC, Chinn S, Sterne JA, Burney PG, Donner A (1999). Methods in health service research. Evaluation of health interventions at area and organisation level. *British Medical Journal* **319**, 376–379.

Unutzer J, Katon W, Callahan CM, Williams Jr. JW, Hunkeler E, Harpole L, Hoffing M, Della Penna RD, Noel PH, Lin EH, Arean PA, Hegel MT, Tang L, Belin TR, Oishi S, Langston C, IMPACT Investigators (2002). Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *Journal of the American Medical Association* **288**, 2836–2845.

Von Korff M, Goldberg D (2001). Improving outcomes in depression. *British Medical Journal* **323**, 948–949.

Wagner EH, Austin BT & Von Korff M (1996). Organizing care for patients with chronic illness. *Milbank Quarterly* **74**, 511–544.

Ware JE, Kosinski M, Dewey JE (2000). *How to Score Version Two of the SF-36 Health Survey*. Quality Metric Incorporated: Lincoln, RI.

Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unutzer J, Miranda J, Carney MF, Rubenstein LV (2000). Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *Journal of the American Medical Association* **283**, 212–220.

White H (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**, 817–838.

Wilkinson G, Allen P, Marshall E, Walker J, Browne W, Mann AH (1993). The role of the practice nurse in the management of depression in general practice: treatment adherence to antidepressant medication. *Psychological Medicine* **23**, 229–237.

WHO (2001). *The World Health Report 2001: Mental Health: New Understanding, New Hope*. World Health Organisation: Geneva.