A pragmatic cluster randomized controlled trial of an educational intervention for GPs in the assessment and management of depression

L. GASK,[1] C. DOWRICK, C. DIXON, C. SUTTON, R. PERRY, D. TORGERSON
AND T. USHERWOOD

From the School of Psychiatry and Behavioural Science, University of Manchester, Department of Primary Care, University of Liverpool; Statistics Group, Department of Physics Astronomy and Mathematics, University of Central Lancashire, Preston and Centre for Health Economics, University of York; and Department of General Practice, University of Sydney at Westmead Hospital, NSW, Australia

ABSTRACT

Background. General practitioners (GPs) can be provided with effective training in the skills to manage depression. However, it remains uncertain whether such training achieves health gain for their patients.

Method. The study aimed to measure the health gain from training GPs in skills for the assessment and management of depression. The study design was a cluster randomized controlled trial. GP participants were assessed for recognition of psychological disorders, attitudes to depression, pre-scribing patterns and experience of psychiatry and communication skills training. They were then randomized to receive training at baseline or the end of the study. Patients selected by GPs were assessed at baseline, 3 and 12 months. The primary outcome was depression status, measured by HAM-D. Secondary outcomes were psychiatric symptoms (GHQ-12) quality of life (SF-36), satis-faction with consultations, and health service use and costs.

Results. Thirty-eight GPs were recruited and 36 (95%) completed the study. They selected 318 patients, of whom 189 (59%) were successfully recruited. At 3 months there were no significant differences between intervention and control patients on HAM-D, GHQ-12 or SF-36. At 12 months there was a positive training effect in two domains of the SF-36, but no differences in HAM-D, GHQ-12 or health care costs. Patients reported trained GPs as somewhat better at listening and understanding but not in the other aspects of satisfaction.

Conclusions. Although training programmes may improve GPs' skills in managing depression, this does not appear to translate into health gain for depressed patients or the health service.

INTRODUCTION

Depression is a major source of distress and disability (Ormel *et al.* 1994) and an important public health concern (Secretary of State for Health, 1998). There appears to be greater potential health benefit in improving its man-agement rather than its detection (Dowrick & Buchan, 1995; Goldberg *et al.* 1998). General practitioners (GPs) can be provided with effec-tive training in the skills to manage depression (Gask *et al.* 1998; Van Os *et al.* 1999). What remains uncertain is how effective such train-ing may be in producing health gain with some authors reporting positive effects (Rutz *et al.* 1989*a,b*, 1992; Tiemens *et al.* 1999) and others (Lin *et al.* 1997, 2001; Thompson *et al.* 2000; King *et al.* 2002) minimal or no effect.

This study aimed to measure the health gain from providing GPs with a multifaceted, inter-active (Oxman *et al.* 1995; Davis *et al.* 1999)

[1] Address for correspondence: Dr Linda Gask, Guild Academic Centre, Royal Preston Hospital, Sharoe Green Lane, Preston, PR2 9HT.

training package in the assessment and management of depression, previously demonstrated to change clinical behaviour and attitudes (Gask *et al.* 1998). The primary hypothesis was that training GPs would lead to a reduction in severity of depression for patients. Secondary hypotheses were that training would reduce psychiatric symptoms, improve quality of life and satisfaction with GP consultations, and reduce health care use and costs.

METHOD

The educational intervention

A 10 h approved training course (Gask *et al.* 1998) was provided over five evenings in both Liverpool and Manchester by L.G. and C.D. (Appendix).

Subjects

The doctors

We aimed to recruit 40 GP principals, drawn equally from Manchester and Liverpool. Within each city, half would be randomized to receive training at the beginning of the study (Intervention group) and half to receive training at the end of the study (Control group). Following research ethics committee approval, all GP principals in Manchester, Liverpool, Sefton and Wirral, were approached by letter. Recruited GPs were assessed on recognition of psychological disorders (Dowrick *et al.* 2000), attitudes to depression (Botega *et al.* 1992) and prescribing patterns (assessed by Prescribing Analysis and CosT data for the preceding 3 months). They also provided information on previous experience of psychiatry and communication skills.

The patients

All study GPs were asked to select patients attending surgery who fulfilled the following criteria: aged 16 to 65 years; intention to treat or current treatment for depression; symptoms for <6 months; not referred to specialist mental health service (except counsellors); not in recovery; not suffering from psychotic illness.

Procedure

Outcome measures

The primary measures were Hamilton Depression (HAM-D) (Hamilton, 1960) score at baseline, 3 and 12 months. The secondary measures were: General Health Questionnaire (GHQ-12) (Goldberg & Williams, 1988) at baseline, 3, 6 (by post) and 12 months (scored by the original '0,0,1,1,' method); SF-36 (Garratt *et al.* 1993) at baseline, 3 and 12 months; a modified version of the Medical Interview Satisfaction Scale (MISS) (Wolf *et al.* 1978; Gask *et al.* 1998) at baseline; economic questionnaire at baseline (assessing preceding 3 months) and 1 year; and total GP consultations, prescriptions and referrals during previous and study year.

Exclusion criteria

Following baseline assessment, only patients with HAM-D scores of $\geqslant 13$ were recruited, to ensure inclusion of patients most likely to benefit from medical intervention for depression (Paykel *et al.* 1998). Patients who did not speak English or were physically too ill to participate were excluded.

Economic analysis

We undertook a cost analysis; cost data were collected in two ways. Patients' GP notes were searched and data on prescriptions, GP and hospital visits were recorded. Additionally, questionnaires at baseline and 12 months asked patients about their use of services during the previous 12 months. Resource items were costed using published data (Netten & Dennet, 1997). The price year was for 1997. Hospital referrals were costed using figures derived from a sample of North-West hospitals and costs were priced in 1997 figures.

Target sample size determination

For the primary outcome measure, the HAM-D, we expected to observe an average baseline score of 19 points. We deemed an excess improvement of 15% (2·85 points) in the trained group compared with the control group as significant. To account for clustering within GPs, we assumed a between-patients (within-GP) variance of 25 and within-patients variance of 5 (conservative). This required us to allocate 40 GPs to the two treatments in equal numbers, with each GP subsequently treating 10–11 patients to achieve 90% power to detect an excess improvement of 2·85 points (using a two-sided 5% significance level).

Statistical analysis

Rationale and method

Analysis was carried out on an 'intention to train' basis although only GPs who recruited patients could be included. The descriptive analyses of patient and GP data used SPSS Version 8.0.0 (SPSS Inc. 1997). To investigate the training effects, mixed model analysis of covariance (ANCOVA) was applied using the 'xtreg' command in Stata Version 6.0 (Stata-Corp 1999), to allow for clustering effects of the GPs on primary and secondary outcome measures, the HAM-D, the GHQ-12 and the SF-36 component scores and the total patient costs to the NHS. The covariate in each case was the baseline value of the measure, as pre-specified in the protocol. The modified MISS scores were analysed similarly but mixed model analysis of variance (ANOVA) was used as there was no corresponding baseline assessment. For the clinical outcome measures, the evaluation of efficacy was made using an intention-to-treat approach. This was implemented via a last observation carried forward (LOCF) approach for patients who prematurely dropped out of the study. Similar methodology was used to investigate the patterns in both patient exclusions and patient drop-outs. Bootstrapping (Efron & Tibshirani, 1993) was used when the parametric assumptions were found not to hold. Bootstrapping is a means of obtaining standard errors and hence confidence intervals for statistical parameters in cases where no standard method is available. An estimation procedure (in this case mixed model ANCOVA), for a dataset based on N observations, leads to a single estimate (of the training effect). By drawing repeated samples of size N, with replacement, from the original dataset, one can repeat this estimation procedure. Then, using the standard deviation of the set of estimates thus obtained, one has obtained a bootstrap estimate of the standard error.

Subsequently, due to unanticipated differences in relevant skills and experience between the GPs in the two centres, it was deemed appropriate to explore possible differential training effects on the clinical outcome measures between the cities by adding a training-by-city interaction to the mixed models described above.

The aim of this analysis was to suggest possible hypotheses for future investigation.

Assignment

Cluster randomization was used, the unit of randomization being the GP. GPs were ranked by size of practice and practice Townsend deprivation scores (Townsend *et al.* 1988) to ensure uniformity of doctors and patients between groups. GPs were matched in pairs by an independent statistician (B.F.). Using computer generated random number tables, one GP in each pair was allocated to the intervention group: by default the other GP in each pair was allocated to the control group. The intervention group was then notified of training course details. Recruitment of patients began immediately for the control group, and after training for the intervention group.

Blinding

It was not feasible, in this trial, to blind the research interviewers to the training status of the GPs, since regular liaison took place between GPs and researchers. However, patients were blind to the training status of their GP.

RESULTS

Baseline assessment of the GPs

Control and intervention GPs were similar in age, year of qualification, practice deprivation status, recognition of depression and attitudes to depression. No significant differences were found between GPs recruited from Liverpool and Manchester, although the former tended to work in more deprived areas. Liverpool GPs were more experienced in videotape training (Liverpool 19/20, 95%; Manchester 13/18, 72%); more likely to have had at least 6 months training in psychiatry (Liverpool 9/20, 45%; Manchester 3/18, 17%) and to have undergone extensive training in counselling and communication skills (Liverpool 6/20, 30%; Manchester 1/18, 6%).

Participant flow and follow-up

Some 280 GPs in Manchester and 250 GPs in Liverpool were invited to participate. In Manchester only 18 GPs expressed an interest, and all were recruited. In Liverpool 37 GPs returned acceptance forms: a random sample of
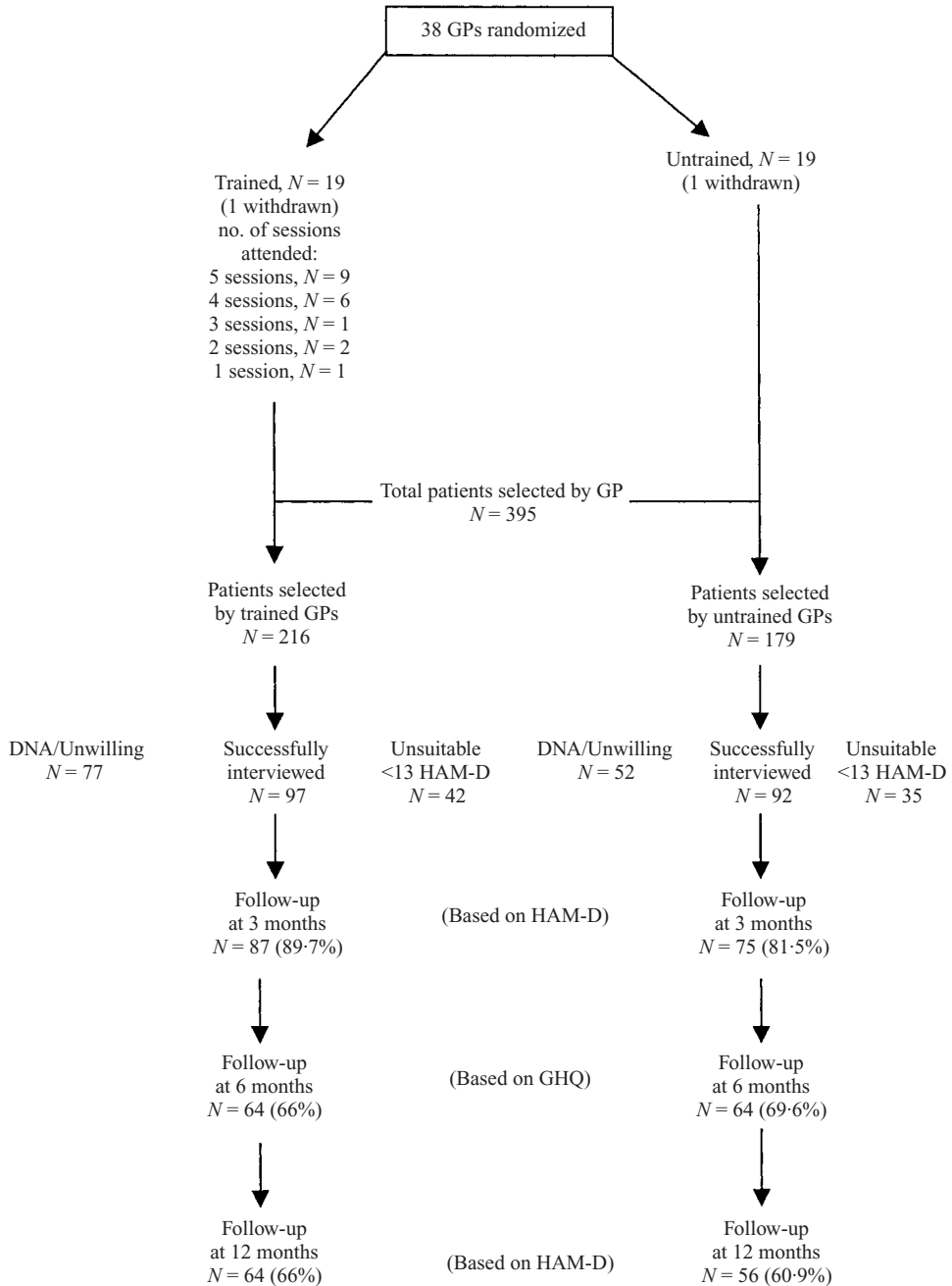
Fig. 1. Consort diagram (DNA, did not attend; HAM-D, Hamilton Depression Scale; GHQ, General Health Questionnaire).

20 was selected. Follow-up is summarized in Fig. 1. Of these 38 GPs, 18 (47%) were women, and 36 (95%) completed the study. They identified 318 patients meeting the eligibility criteria: 189 (59%) patients were successfully interviewed and recruited. Ages ranged from 18 to 67 years, and 128 (68%) were women. Townsend deprivation scores ranged from −4·88 (affluent) to +13·00 (high deprivation), heavily weighted towards deprivation. Median

Table 1. *Characteristics of patients at baseline by group*: *values are medians* (*interquartile ranges*) *unless stated otherwise*

	Group				
	Control ($N=92$)			Intervention ($N=97$)	
Measure	Value	Number missing		Value	Number missing
Female/Male, N	56/36	0		72/25	0
Townsend deprivation score	+4·0 (+0·2 to +7·0)	4		3·7 (+0·0 to +7·0)	5
HAM-D	20 (16 to 24)	0		19 (15 to 22)	0
GHQ-12 score	11 (5 to 12)	2		10·5 (5 to 12)	1
Age	37·5 (31 to 45)	2		38 (30 to 46)	0
Physical function	75 (55 to 90)	5		75 (55 to 90)	1
Role limitation					
Physical	25 (0 to 100)	4		25 (0 to 100)	2
Mental	0 (0 to 0)	4		0 (0 to 33·3)	1
Social function	33·3 (11·1 to 55·6)	5		33·3 (11·1 to 55·6)	1
Mental function	26·5 (16 to 40)	4		32 (16 to 44)	1
Energy/vitality	20 (10 to 30)	4		25 (10 to 40)	1
Pain	55·6 (33·3 to 77·8)	4		50 (33·3 to 66·7)	1
Health perception	42 (25 to 64)	4		41 (25 to 59)	1
Change in health	25 (12·5 to 50)	4		25 (25 to 50)	1

Townsend scores were similar for control (+4·0) and intervention (+3·7) groups (see Table 1), but were markedly different in Liverpool (+5·0) and Manchester (+3·1).

Mixed model analysis of variance showed that the average between-city difference was 1·80 (95% CI −0·42 to 4·01, $P=0·11$). Other baseline variables were similar in the control and intervention groups, although there were more women in the intervention (72/97, 74·2%) than the control group (56/92, 60·9%). The patients in Liverpool tended to be older (median 39 years) than in Manchester (median 34 years) and to have higher initial SF-36 scores.

GPs successfully recruited a median 5·5 patients each (range 0–14, interquartile range 2 to 7). Thirty-four GPs recruited one or more patients. The only significant predictor of non-recruitment, following backward elimination, was city: 70/150 (47%) of eligible patients selected by Manchester GPs were recruited, compared with 119/168 (71%) in Liverpool. Reasons for this were higher instance of failure to attend for appointment with the researcher in Manchester and higher referral to the study of patients found to score <13 on the HAM-D. Twenty-seven (14%) patients had dropped out by 3 months and 69 (37%) by 12 months: rates were similar between intervention (33/97 (34%)) and control (36/92 (39%)) groups. Patients dropping out tended to be from Liverpool

($P=0·088$), have higher baseline HAM-D scores ($P=0·085$) or GHQ-12 scores ($P=0·086$) (univariate analyses). Those recruiting less than the median number of patients had similar attrition rate (55/150, 36·7%) to those recruiting more than the median (14/39, 35·9%). Case notes were available for 152 (80%) patients.

Analysis

Clinical outcome

The modelling of group differences, using baseline score as covariate and allowing for clustering by GP, showed no significant difference between the intervention and control groups for the HAM-D and GHQ-12 (Table 2). The differences observed were, however, in the direction of improvement for all five measures. Modelling of SF-36 scores showed significantly better patient scores among the trained group for 'role limitation mental' ($P=0·015$) and 'health perception' ($P=0·007$) at 12 months, and a similar trend for social function at 12 months ($P=0·062$) (Table 2). Furthermore, the observed difference favoured the trained group for all three and twelve month SF-36 scores, except the 3 month physical function score.

We noted that training effects were consistently different in Liverpool and Manchester, and made a *post hoc* decision to add a training-by-city interaction to the model for each of the

Table 2. *Effect of training: using the last observation carried forward method for patients dropping out*

Measure	Difference (95% CI): intervention–control
HAM-D score	
At 3 months	−0·9 (−3·1 to 1·3)
At 12 months	−1·9 (−4·1 to 0·3)
GHQ-12 score	
At 3 months	−0·9 (−2·6 to 0·8)
At 6 months	−0·4 (−1·7 to 0·9)
At 12 months	−0·8 (−2·1 to 0·6)
SF-36	
Physical function	
At 3 months	−1·8 (−10·4 to 6·9)
At 12 months	1·4 (−7·5 to 10·2)
Role limitation	
At 3 months, physical	2·2 (−14·8 to 19·1)
At 12 months, physical	13·2 (−3·5 to 29·9)
At 3 months, mental	9·1 (−3·9 to 22·0)
At 12 months, mental	17·9 (3·4 to 32·5)*
Social function	
At 3 months	6·9 (−6·8 to 20·7)
At 12 months	11·9 (−0·6 to 24·4)
Mental function	
At 3 months	2·3 (−4·2 to 8·7)
At 12 months	4·7 (−1·9 to 11·2)
Energy/vitality	
At 3 months	1·4 (−6·0 to 8·9)
At 12 months	5·8 (−1·6 to 13·1)
Pain	
At 3 months	3·7 (−6·8 to 14·1)
At 12 months	4·0 (−6·8 to 14·9)
Health perception	
At 3 months	3·5 (−2·2 to 9·3)
At 12 months	9·2 (2·5 to 15·8)**
Change in health	
At 3 months	0·0 (−7·9 to 7·8)
At 12 months	1·5 (−8·0 to 11·0)

HAM-D, Hamilton Depression Scale; GHQ-12, General Health Questionnaire; SF-36, Health Survey Questionnaire, short-form.
   * $P < 0.05$; ** $P < 0.01$.

outcome variables. There was little observed change in HAM-D scores in Liverpool over the 12 months of the trial (3 months, 0·8, 95% CI −2·1 to 3·6; 12 months, −0·4, 95% CI −3·4 to 2·5). A training effect was observed in Manchester between baseline and 3 months (−3·4, 95% CI −6·9 to 0·0), maintained at 1 year (−4·1, 95% CI −7·8 to −0·4). Training effects were also observed for Manchester patients on GHQ-12 and four SF-36 domains. These observations suggest a differential training effect in the two cities.

Costs

Modelling of the total costs for the 12 months of the study, adjusted for costs over 3 months

Table 3. *Differences in patient satisfaction for GPs' listening and understanding skills by treatment group: 100 mm Visual Analogue Scale scores (positive values indicate a higher degree of agreement with the statement in the intervention group)*

Statement	Difference (95% CI)
Listening	
I was able to discuss my concerns with the doctor.	4·6 (−1·5, 12·7)
The doctor didn't take my problem seriously.	−9·5 (−16·9, −1·7)
The doctor dismissed some of the problems I mentioned.	−6·5 (−14·4, −0·7)
Understanding	
The doctor understood the nature of my symptoms.	3·3 (−2·0, 9·7)
The doctor understood how much my symptoms were worrying me.	3·8 (−1·4, 9·4)
The doctor understood how I had been feeling emotionally.	5·2 (0·6, 11·1)
The doctor understood the problems I have been experiencing in my life.	2·2 (−3·7, 10·1)

before randomization, showed no significant difference in mean costs between intervention and control groups ($P = 0.16$). Despite the positive skewness of the distribution of total costs, bootstrapping gave similar estimates to those from the parametric analysis: the effect of the intervention was to increase mean total costs by £121 (95% CI: £48 decrease to £291 increase).

Patient satisfaction

Fifteen statements were assessed by the patient on a Visual Analogue Scale (VAS) modified from the MISS, with 0 representing 'Strongly Disagree' and 100 representing 'Strongly Agree' with the statement. GPs in the intervention group were generally rated higher on listening skills (2/3 95% CIs exclude no difference, see Table 3) and understanding (1/4 CIs exclude no difference, all favour training, see Table 3), but the differences in explaining, gaining the patient's trust and length of appointment were small and non-significant (not shown).

Process of care

Health service use increased during the study, unrelated to either study group or city. Mean (S.D.) consultation rates were 3·6 (2·2) the year before the study, and 11·7 (6·4) during it. Prescribing increased from 360 items the year

before, to 814 during the study, of which 187 (52%) and 289 (36%) respectively were for psychotropics. Psychiatric out-patient appointments increased sixfold, from 9 to 56. There was only one psychiatric admission (in Liverpool) during the study period.

DISCUSSION

We were unable to demonstrate significant effects of giving GPs a 10-hour skills-based training to treat depression. Desirable though such training may appear, it did not have a significant effect on patient outcome in this study. Because of lack of patient recruitment and patient attrition the study was underpowered to draw firm conclusions, it can be inferred that the size of any training effect overall effect is small or moderate. However, having powered our study to detect a 15% (2·85) effect of training on patient HAM-D, the 95% CI limit favouring training for the 3 and 12-month effects (3·1 and 4·1 respectively) do not enable us to exclude a moderate effect of training. No significant economic differences were demonstrated as a result of training. Differences in clinical outcome were noted in *post hoc* analysis of the two training centres, and there were indications of limited impact on patient quality of life and satisfaction.

Methodological issues

Although 41% of selected patients were subsequently not recruited, none of the baseline variables was found to be associated with non-recruitment. We do not know how many eligible patients were not selected by GPs although only five GPs (all from Liverpool) successfully selected the desired 10 patients or more. We are not alone in experiencing problems in trials where GPs recruit patients (Ward *et al.* 1999). A problem associated with selection of cases by GPs in such a study is that the 'trained' arm may recruit cases more readily and also recruit 'less ill' patients. There is some evidence for the latter here but not the former. However, in this pragmatic trial, it was essential to recruit patients whom GPs considered to be depressed, and were interested in managing actively. In primary care research, recruiting against an external assessment always runs the risk of introducing anxiety/hostility in GPs, but we now concede that methods of recruitment that do not rely solely on the GP may be necessary to achieve a sufficient recruitment rate.

We asked doctors to recruit patients who had been depressed for <6 months because we wanted to avoid long-term depression, which is more difficult to treat. There is no evidence that training GPs would have particularly helped this group of patients – indeed, more research into effective interventions for this group of patients is needed. We decided to seek recruitment of both patients who were already in continuing treatment as well as those where there was a decision to treat because there is no evidence that treatment as usual by GPs has an impact on clinical outcome (Goldberg *et al.* 1998) and equally no evidence that it is more difficult for training to have an impact on outcome if some management decisions have already been made. Limiting patients to younger than the age of 65 was probably, in retrospect, ageist – there is no clinical justification for this, however, we did not attempt to provide any specific advice on the pharmacological management of depression in older people during the training sessions.

Three GPs had minimal training (two or fewer sessions): we did not exclude their patients from the trial, as we considered this variation in attendance accurately reflected realities of postgraduate general practice education. The GPs were self-selected, and may, therefore, have been particularly interested in mental health: this too is a feature of those attending postgraduate education sessions across the world. Although it would be desirable to recruit on the basis of educational need, this was not practically possible. Two doctors dropped out early in the study after completion of the baseline measures and (in one case) after training, because he left the area. All the others continued to express willingness to recruit patients to the study until the end although two failed to recruit any patients.

Our use of the GHQ might be criticized on the grounds of its possible insensitivity to change. However, we also used the HAM-D, which is undoubtedly sensitive to change, and the SF-36. We have not reported caseness data for the GHQ, HAM-D or SF-36 because we were concerned with symptom reduction, not change in case status.

The attrition rate was 14% at 3 months and 37% at 12 months, and slightly higher for control than intervention patients. Attrition was not related to GP recruitment. The use of LOCF limits bias in unknown directions although when we analysed the 'completers' only or used available-case analysis, we generally obtained smaller effect estimates than when we used LOCF. Furthermore, the use of LOCF tends to lead to artificially narrowed confidence intervals so, apart from any bias due to drop-outs, the confidence limits are likely to be wider than quoted. However, drop-outs were at a level to be expected in a study where patient recruitment depends upon recognition and negotiation by GPs in inner-city practices where the number of people who leave the practice or change of address within the practice area can be as high as 30% per year or more (personal communication: Rusholme Health Centre, Manchester) and our dropout rate compares favourably with that of Thompson *et al.* (2000). It appears unlikely that any bias not accounted for using LOCF would be of sufficient magnitude to alter our substantive conclusions although, combined with a reduction in precision due to under-recruitment and dropout, we are unable to rule out a moderate effect of training.

The differential training effect between Manchester and Liverpool could reflect the different experience and training of the doctors in these cities. An alternative explanation is that patients recruited in Liverpool were more deprived than in Manchester, and hence less amenable to intervention: however, the magnitude of difference was small and the level of deprivation was high in both cities. Recruitment of doctors was more difficult in Manchester than Liverpool, where postgraduate training in mental health and communication skills is better established. Thus, training interventions may be more likely to succeed when the target group of doctors have received less previous training (but recruitment may be more difficult). However, the trial was not powered for subgroup investigations and there was substantial under-recruitment of patients, so no firm conclusions should be drawn from this inter-city analysis. It does generate hypotheses for future investigation, including a subsequent qualitative study of GPs' views of the relevance and impact of the training.

Process measures, other than satisfaction, did not reveal possible explanation for the inter-city differences. The study did not include measures – such as videotape – sufficiently sensitive to detect changes at the level of the consultation throughout the follow-up period.

Relevance to the published literature

There are limited published data with which to compare our findings. The previous relevant trials by Rutz *et al.* (1992); Lin *et al.* (1997, 2001) and Tiemens *et al.* (1999) were all pre-post training evaluations, while Thompson *et al.* (2000) was a randomized controlled trial of training primary care staff in implementing clinical guidelines. King *et al.* (2002), in a cluster randomized controlled trial, showed no effect of brief training of GPs in cognitive behaviour therapy.

Our design differed from the UK study by Thompson *et al.* (2000) in that our intervention has previously been demonstrated in a pilot study to bring about observable acquisition of new skills (Gask *et al.* 1998). However, this does not necessarily mean than doctors who have acquired such skills can apply them in practice with a resulting improvement in clinical outcome, and we shall be reporting separately on a qualitative study of the difficulties that the GPs in our study experienced in applying what they had learned. Three other important differences between these two studies are: our training focused on acquisition of skills, rather than on adherence to guidelines; we followed up a single group of patients for the duration of the study, instead of separate cohorts at three time points; and in general we used clinical face-to-face measures, rather than postal questionnaires. The studies by Lin *et al.* (1997 and 2001) also differ from our study in that training was considerably more 'skills' focused than that provided by the US investigators.

The study by King *et al.* (2002) is particularly pertinent to our study as this group attempted to train GPs specifically in cognitive-behavioural skills and found no impact on patient outcome. They found that their training course had little effect on GPs' attitudes to the identification and treatment of depression – which is contrary to our own findings (Gask *et al.* 1998). We decided to drop the cognitive element from our training package following our 1998 pilot study of training as we found no evidence that such skills

could be acquired by GPs in our (albeit much briefer) training module. Thus, the content of our training differs from that of King *et al.* in covering a broader range of skills, including those required for negotiating drug treatment and assessing suicide risk.

The Dutch study (Tiemens *et al.* 1999) suggested a training effect at 3 months, which disappeared by 1-year, and concluded that the effect was to promote faster recovery for patients with shorter duration of illness. The effect of training was weaker in the subgroup of physicians who had been previously involved in research projects. Our findings are consistent with this: although our overall result was negative, positive training effects were observed in the centre where study doctors were less experienced, and had less of an 'education' culture.

Finally, when considered against the recommendations provided by Hodges *et al.* (2001) in a recent review of methods to improve the knowledge, attitudes and skills of primary care physicians, our training did have some limitations. We were not able to provide ongoing support and supervision over time to participants. We have found it difficult to get course participants to return to follow-up sessions and in practice we suspect this can only be achieved by tenaciously following up doctors individually at their place of work rather than expecting more than a few of them to continue to attend a central educational meeting. This has implications in terms of funding and training personnel.

Policy implications

It is current policy in the UK to promote training for primary care in the management of 'minor' emotional disorders (Secretary of State for Health, 1999). Our findings suggest that a blanket training strategy will not lead to substantially better management of depression in primary care. Where doctors already have good access to postgraduate training in mental health, the impact of providing further training may be limited, unless targeted at those who most need it. A recent review of the impact of continuing medical education (Davis *et al.* 1999) emphasizes the importance of attention to contextual factors in practice, for example, combining training with office practice re-structuring, as in collaborative approaches to care (Simon *et al.*

2000; Wells *et al.* 2000). Although our approach to training differs from that of Lin *et al.* (2001), we would support their conclusions: 'our results do not imply that physician education has no role in improving quality of primary care. Physician education is a necessary but insufficient strategy'. Training, which should be aimed at providing participants with new skills, should form only part of a wider strategy for improving the quality of depression in primary care.

APPENDIX: THE ASSESSMENT AND MANAGEMENT OF DEPRESSION IN PRIMARY CARE, A 10 HOUR COURSE

Aims of the course

To improve the assessment and management of depression in general practice by effective utilization of pharmacological, physical and social interventions which are realistic within the confines of the consultation.

Training methods

The key focus of the course in on acquisition of appropriate clinical skills. Each of the five 2 h sessions consists of:

A ***brief presentation/lecture*** on each topic.
Viewing specially developed ***videotaped material***.
The opportunity for each GP to ***role-play*** consultations as both the GP and a depressed patient in order to develop specific ***microskills***.
These role-plays are videotaped and used for ***videofeedback*** in small groups.
Each GP receives ***written material*** to support each session.

Course timetable

Week 1 Assessing depression.
Week 2 Negotiating the treatment contract and drug treatment of depression.
Week 3 Problem-solving therapy and social interventions.
Week 4 The question of suicide.
Week 5 Cognitive and behavioural skills.

REFERENCES

Botega, B. N., Blizard, R., Wilkinson, G. & Mann, A. (1992). General practitioners and depression: first use of the Depression Attitude Questionnaire. *International Journal of Methods in Psychiatric Research* **2**, 169–180.

Davis, D., Thomson O'Brien, M. A., Wolf, F. M. & Freemantle, N. (1999). Impact of formal continuing medical education. *Journal of the American Medical Association* **282**, 867–874.

Dowrick, C. & Buchan, I. (1995). Twelve-month outcome of depression in general practice: does detection or disclosure make a difference? *British Medical Journal* **311**, 1274–1276.

Dowrick, C., Gask, L., Perry, R., Dixon, C. & Usherwood, T. (2000). Do general practitioners' attitudes towards depression predict their clinical behaviour? *Psychological Medicine* **30**, 413–419.

Efron, B. & Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall: New York.

Garratt, A., Ruta, D., Abdalla, M. I., Buckingham, J. K. & Russell, I. T. (1993). The SF-36 health survey questionnaire: an outcome measure for routine use within the NHS. *British Medical Journal* **306**, 1440–1444.

Gask, L., Thomson, H., Usherwood, T. & Williams, B. (1998). Evaluation of a training package in the assessment and management of depression in primary care. *Medical Education* **32**, 190–198.

Goldberg, D. & Williams, P. (1988). *A User's Guide to the General Health Questionnaire*. NFER-Nelson: Windsor.

Goldberg, D., Privett, M., Üstün, B., Simon, G. & Linden, M. (1998). The effects of detection and treatment on the outcome of major depression in primary care: a naturalistic study in 15 cities. *British Journal of General Practice* **38**, 1840–1844.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery and Psychiatry* **23**, 56–61.

Hodges, B., Inch, C. & Silver, I. (2001). Improving the psychiatric knowledge, skills and attitudes of primary care physicians, 1950–2000: a review. *American Journal of Psychiatry* **158**, 1579–1586.

King, M., Davidson, O., Taylor, F., Haines, A., Sharp, D. & Turner, R. (2002). Effectiveness of teaching general practitioners skills in brief cognitive behaviour therapy to treat patients with depression: randomised controlled trial. *British Medical Journal* **324**, 947–950.

Lin, E., Katon, W., Simon, G., Von Korff, M., Bush, T. M., Rutter, C. M., Saunders, K. W. & Walker, E. A. (1997). Achieving guidelines for the treatment of depression in primary care: is physician training enough? *Medical Care* **135**, 831–842.

Lin, E., Simon, G., Katzelnick, D. J. & Pearson, S. D. (2001). Does physician education on depression management improve treatment in primary care? *Journal of General Internal Medicine* **16**, 614–619.

Netten, A. & Dennet, J. (1997). *Unit Costs of Health and Social Care*. PSSRU University of Kent at Canterbury: Canterbury.

Ormel, J., von Korff, M., Üstün, T., Pini, S., Korten, A. & Oldehinkel, T. (1994). Common mental disorders and disability across cultures. *Journal of the American Medical Association* **272**, 1741–1748.

Oxman, A., Thomson, M., Davis, D. & Haynes, R. B. (1995). No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Canadian Medical Association Journal* **153**, 1423–1431.

Paykel, E., Hollyman, J., Freeling, P. & Sedgwick, P. (1998). Predictors of therapeutic benefit from amitriptyline in mild depression: a general practice placebo-controlled trial. *Journal of Affective Disorders* **14**, 83–95.

Rutz, W., von Knorring, L. & Walinder, J. (1989*a*). Frequency of suicide on Gotland after systematic postgraduate education of GPs. *Acta Psychiatrica Scandinavica* **80**, 151–154.

Rutz, W., Walinder, J., Eberhard, G., Holmberg, G., von Knorring, A. L., von Knorring, L., Wistedt, B. & Aberg-Wistedt, A. (1989*b*). An educational program on depressive disorders for GPs on Gotland: background and evaluation. *Acta Psychiatrica Scandinavica* **79**, 19–26.

Rutz, W., von Knorring, L. & Walinder, J. (1992). Long-term effects of an educational program for the prevention and treatment of depression. *Acta Psychiatrica Scandinavica* **85**, 83–88.

Secretary of State for Health (1998). *Our Healthier Nation*. Stationery Office: London.

Secretary of State for Health (1999). *National Service Frameworks: Mental Health*. Stationery Office: London.

Simon, G., VonKorff, M., Rutter, C. & Wagner, E. (2000). Randomised trial of monitoring, feedback, and management of care by telephone to improve treatment of depression in primary care. *British Medical Journal* **320**, 550–554.

SPSS Inc. (1997). *SPSS for Windows Version 8.0.0*. SPSS: Chicago, IL.

StataCorp (1999). *Stata Statistical Software: Release 6.0*. Statacorp: College Station, TX.

Thompson, C., Kinmonth, A., Stevens, L., Peveler, R. C., Stevens, A., Ostler, K. J., Pickering, R. M., Baker, N. G., Henson, A., Preece, J., Cooper, D. & Campbell, M. J. (2000). Effects of a clinical practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire depression project randomised controlled trial. *Lancet* **355**, 185–191.

Tiemens, B., Ormel, J., Jenner, J., van der Meer, K., Van Os, T. W., van den Brink, R. H., Smit, A. & van den Brink, W. (1999). Training primary-care physicians to recognize, diagnose and manage depression: does it improve patient outcomes? *Psychological Medicine* **29**, 833–845.

Townsend, P., Phillimore, P. & Beattie, A. (1988). *Health and Deprivation: Inequalities and the North*. Croom Helm: London.

Van Os, T., Ormel, J. & van den Brink, R. (1999). Training primary care physicians improves the management of depression. *General Hospital Psychiatry* **21**, 168–176.

Ward, E., King, M., Lloyd, M., Bower, P. & Friedli, K. (1999). Conducting randomized trials in general practice: methodological issues and practical issues. *British Journal of General Practice* **49**, 919–922.

Wells, K., Sherbourne, C., Schoenbaum, M., Duan, N., Meredith, L., Unutzer, J., Miranda, J., Carney, M. F. & Rubenstein, L. V. (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.

Wolf, M., Putnam, S., James, S. & Stiles, W. B. (1978). The Medical Interview Satisfaction Scale: development of a scale to measure patient perceptions of physician behaviour. *Journal of Behavioural Medicine* **1**, 391–401.