**ORIGINAL INVESTIGATION**

**Collaborative Care for Depression**

*A Cumulative Meta-analysis and Review of Longer-term Outcomes*

Simon Gilbody, MBChB, MRCPsych, DPhil; Peter Bower, PhD; Janine Fletcher, MSc; David Richards, PhD; Alex J. Sutton, PhD

**Background:** Depression is common in primary care but is suboptimally managed. Collaborative care, that is, structured care involving a greater role of nonmedical specialists to augment primary care, has emerged as a potentially effective candidate intervention to improve quality of primary care and patient outcomes.

**Methods:** To quantify the short-term and longer-term effectiveness of collaborative care compared with standard care and to understand mechanisms of action by exploring between-study heterogeneity, we conducted a systematic review of randomized controlled trials that compared collaborative care with usual primary care in patients with depression. We searched MEDLINE (from the beginning of 1966), EMBASE (from the beginning of 1980), CINAHL (from the beginning of 1980), PsycINFO (from the beginning of 1980), the Cochrane Library (from the beginning of 1966), and DARE (Database of Abstracts of Reviews of Effectiveness) (from the beginning of 1985) databases from study inception to February 6, 2006.

**Results:** We found 37 randomized studies including 12,355 patients with depression receiving primary care. Random effects meta-analysis showed that depression outcomes were improved at 6 months (standardized mean difference, 0.25; 95% confidence interval, 0.18-0.32), and evidence of longer-term benefit was found for up to 5 years (standardized mean difference, 0.15; 95% confidence interval, 0.001-0.31). When exploring determinants of effectiveness, effect size was directly related to medication compliance and to the professional background and method of supervision of case managers. The addition of brief psychotherapy did not substantially improve outcome, nor did increased numbers of sessions. Cumulative meta-analysis showed that sufficient evidence had emerged by 2000 to demonstrate the statistically significant benefit of collaborative care.

**Conclusions:** Collaborative care is more effective than standard care in improving depression outcomes in the short and longer terms. Future research needs to address the implementation of collaborative care, particularly in settings other than the United States.

Arch Intern Med. 2006;166:2314-2321

**DEPRESSION IS SOON TO BECOME the second leading cause of disability worldwide.** It affects between 5% and 10% of individuals and is the third most common reason for consultation in primary care. Management falls below accepted evidence-based standards, and the enhanced management of depression in primary care is central to the World Health Organization strategy for mental health. Many organizational and educational strategies targeted at health care professionals have been proposed to improve the recognition and management of depression in primary care. These include the following: educational strategies, such as guidelines, targeted at health care professionals; consultation-liaison, with an educative role for practitioners working more closely with non-nurse practitioners working in conjunction with the primary care physician and a mental health specialist.

Previous reviews of management of depression have identified collaborative care as the most effective of these approaches. However, there are important limitations to these published reviews. Some reviews have pooled heterogeneous studies with only limited exploration of important causes of heterogeneity and others have omitted important studies because of the inadequacy of search.
strategies. In addition, substantial evidence has emerged in recent years capturing the longer-term outcome of these interventions. Collaborative care is an active area of research, and previous reviews are substantially out of date.

Collaborative care captures a range of interventions of varying intensity, ranging from simple telephone interventions to encourage compliance with medication to more complex interventions that involve intensive follow-up and incorporate a form of structured psychosocial intervention. Such study-level design variables might be related to the overall effectiveness of a collaborative care program. Similarly, collaborative care has generally been developed in the United States within managed health care settings, and the overall effectiveness of collaborative care programs might vary when it is implemented and evaluated in non-US settings. It remains unclear, therefore, just how effective collaborative care is and what the important determinants of effectiveness are. Identifying the magnitude of clinical effectiveness and the important determinants of effect is vital for those who plan services and might implement collaborative care.

Our purpose was to explore the totality of randomized research into collaborative care in more detail and with more rigor than has been done previously, to establish the clinical effectiveness of collaborative care during both the short- and longer-terms, important determinants of effectiveness of collaborative care, and how research has evolved with time and the totality of research evidence in this area.

**METHODS**

**LITERATURE SEARCH AND INCLUSION CRITERIA**

We searched a variety of biomedical, nursing, and psychological databases from study inception to February 6, 2006, including MEDLINE (from the beginning of 1966), EMBASE (from the beginning of 1980), CINAHL (from the beginning of 1980), PsycINFO (from the beginning of 1980), the Cochrane Library (from the beginning of 1986), and DARE (Database of Abstracts of Reviews of Effectiveness) (from the beginning of 1980). We also scrutinized reference lists of studies and used citation searching for all studies that met our inclusion criteria.

We included randomized controlled trials with patients with depression being managed in primary care settings using a collaborative care approach. For this review, collaborative care was broadly defined as a multifaceted intervention involving combinations of 3 distinct professionals working collaboratively within the primary care setting: a case manager, a primary care practitioner, and a mental health specialist. To be included, studies had to involve 2 of these 3 components of collaborative care.

**OUTCOMES**

We analyzed short-term (6 months) and longer-term (12, 18, and 24 months, and 5 years) outcomes in both collaborative care and standard care groups. We sought to standardize outcomes between studies, specifically by seeking data on depression outcomes and antidepressant medication concordance. Collaborative care interventions often seek to improve concordance with antidepressant medication, and we analyzed changes in measures of antidepressant use (eg, the percentage of patients taking antidepressant medications or meeting standardized guidelines for antidepressant medication use). Where multiple outcomes were reported, we chose any identified primary outcome first, then prioritized observer-rated scales over self-report measures. We translated continuous measures to a standardized effect size (ie, mean of intervention group minus mean of control group divided by the pooled standard deviation). We translated outcomes reported as dichotomous variables to standardized effect size using the logit transformation.

**META-ANALYTIC POOLING AND MEASUREMENT OF HETEROGENEITY**

We performed a random-effects meta-analysis. Between-study heterogeneity was assessed using the I² statistic, which describes the percentage of total variation across studies that is the result of heterogeneity rather than chance. Publication bias was examined by constructing Begg funnel plots and by testing for funnel plot asymmetry using the Egger weighted regression test. All statistical pooling was conducted using “meta” and “metabias” user-written commands in STATA version 8 (StataCorp, College Station, Tex).

**CORRECTION FOR UNIT OF ANALYSIS ERROR**

We identified all studies using cluster randomization and, where necessary, adjusted the precision of these studies in the meta-analysis using a sample size or variation inflation method and assuming an intraclass correlation of 0.02, in line with published estimates.

**EXPLORATION OF CAUSES OF HETEROGENEITY**

We anticipated several sources of heterogeneity relating to the content of the intervention and fidelity to a collaborative care model; the health care setting, and the degree to which patients were concordant with medication within a collaborative care program. Predictive variables included fidelity to the collaborative care model, as defined by Katon et al; study setting (US vs non-US); recruitment method (screening vs referral by clinicians); study population (unselected depressed patients vs depressed patients identified as willing to take medication); use of primary care physician training; case manager professional background; case manager supervision; addition of psychotherapy to standard care management; and number of case management sessions.

Where these were reported, in most studies, we explored the effect of these study-level variables on the overall effectiveness of collaborative care using sensitivity analyses and meta-regression techniques. Clinical outcomes closest to 6 months were analyzed using meta-regression, with a permutation test (using 1000 Monte Carlo simulations, StataCorp) to calculate P values and to reduce the chance of spurious false-positive findings. The amount of heterogeneity explained by the use of predictive covariates was examined by reductions in the I² inconsistency statistic within our model. Analyses were conducted using the “meta” and “metareg” commands in STATA version 8. To examine the relationship between the use of antidepressant medication and depression outcomes, we fitted a weighted Bayesian regression model, allowing for measurement error in both variables, using WinBUGS (Medical Research Council Biostatistics Unit, Cambridge, England).

**CUMULATIVE META-ANALYSIS**

We explored the evolution of evidence of the effectiveness of collaborative care over time using cumulative meta-analysis.
RESULTS

From 15,633 citations, 37 randomized studies met our incision criteria. Our review included 12,355 patients with depression randomized to receive either collaborative care or usual primary care.

EFFECT OF COLLABORATIVE CARE ON DEPRESSION OUTCOMES AT 6 MONTHS

Collaborative care had a clearly positive effect on standardized depression outcomes at 6 months compared with standard care (standardized mean difference [SMD], 0.25; 95% confidence interval [CI], 0.18-0.32). As anticipated, there was a moderate level of heterogeneity between studies ($I^2=52.8\%$) (Figure 1). There was no evidence of small study or publication bias ($P=.14$) for these studies (Figure 2).

LONGER-TERM OUTCOMES

Eleven studies (31,36,39,44,45,61-66) provided longer-term outcomes of up to 57 months with collaborative care compared with standard care. The overall trend was for clinical improvement to be maintained at 12 months (SMD, 0.31; 95% CI, 0.01 to 0.53), 18 months (SMD, 0.25; 95% CI, 0.03 to 0.46), 24 months (SMD, 0.15; 95% CI, -0.03 to 0.34), and 5 years (SMD, 0.15; 95% CI, 0.001 to 0.30), although this failed to reach statistical significance at 24 months. There was substantial between-study heterogeneity ($I^2=84\%$ at 24 months) (Figure 3), but there were insufficient individual studies to explore the overall sources of this heterogeneity.

Since only a subset of studies reported longer-term outcome, we also tested for publication bias using the Egger test. None was evident at 12 months ($P=.09$), 18 months...
or 24 months ($P = .26$). Insufficient studies were available at 36 months or 5 years to exclude publication bias.

**EXPLORATION OF CAUSES OF HETEROGENEITY**

Sufficient study-level data were also available from most of the studies to allow the effect of the predictor variables to be entered into a meta-regression analysis.

**Study Setting**

Studies from the United States showed a strongly positive and statistically significant effect for collaborative care in improving depression outcomes at 6 months (SMD$_{US}$ studies, 0.27; 95% CI, 0.22-0.33) and were subject to minimal between-study heterogeneity ($I^2_{US}$ studies, 5.4%), whereas non-US studies were nonsignificant in their pooled effect size (SMD$_{non-US}$ studies, 0.24; 95% CI, −0.06 to 0.55) and were subject to substantial between-study heterogeneity ($I^2_{non-US}$ studies, 85.7%) (Figure 4). However, the pooled point estimate was essentially similar in US and non-US studies (meta-regression $\beta$, 0.01; 95% CI, −0.19 to 0.21; $P = .91$).

**Content of Intervention**

When we examined fidelity to the collaborative care model,$^9$ all studies had a case manager, but several studies deviated from the model in that they did not have access to specialist input. These studies with lower fidelity showed a lower pooled effect size and were more heterogeneous (SMD$_{low fidelity}$, 0.187; $I^2_{low fidelity}$, 73.3%; SMD$_{high fidelity}$, 0.30; $I^2_{high fidelity}$, 4.6%), although this difference was not significant (meta-regression $\beta$, 0.09; 95% CI, −0.08 to 0.25; $P = .29$; $I^2 = 50.7$%). Two study-level variables, regular supervision and the mental health background of case managers, were significantly related to study effect size. The use of regular and planned supervision of the case manager, usually by a psychiatrist, was related to a more positive clinical outcome (SMD$_{usual supervision}$, 0.29; SMD$_{unplanned and ad hoc supervision}$, 0.14; meta-regression $\beta$, 0.15; 95% CI, −0.02 to 0.31; $P = .07$; $I^2 = 49.3$%). Case managers with a specific mental health background also achieved better outcomes (SMD$_{CM mental health background}$, 0.34; SMD$_{CM non–mental health background}$, −0.164; meta-regression $\beta$, 0.18; 95% CI, 0.04-0.32; $P = .02$; $I^2 = 42.4$%). However, the addition of a specific form of psychotherapy to medication management in collaborative care was not associated with any significantly increased effect size (SMD$_{psychotherapy + medication management}$, 0.30; SMD$_{medication management only}$, 0.21; meta-regression $\beta$, 0.10; 95% CI, −0.05 to 0.25; $P = .20$; $I^2 = 49.3$%). Similarly, studies in which antidepressant medication was prescribed at entry to the trial were no more effective (SMD$_{antidepressants at entry}$, 0.21; SMD$_{antidepressants not consistently prescribed}$, 0.30; meta-regression $\beta$, −0.09; 95% CI, −0.24 to 0.06; $P = .23$; $I^2 = 50.7$%). The number of case management sessions ranged from 2 to 14, but the number of sessions was not related to outcome (meta-regression $\beta$, 0.02; 95% CI, −0.008 to 0.04; $P = .19$; $I^2 = 50.9$%) (Figure 5).
Compliance With Medication

From the weighted Bayesian regression model, compliance with medication predicted depression outcomes with credible certainty (slope coefficient, 0.19; 95% credible interval, 0.08-0.30) \((\text{Figure 6})\).

Cumulative Meta-analysis of Outcome at 6 Months

By plotting the emergence of collaborative care with time \((\text{Figure 7})\), it was clear that earlier trials of collaborative care fitting our inclusion criteria conducted in the late 1980s and early 1990s demonstrated a high degree of heteroge-
neity and a high percentage of negative results. More positive studies emerged in the mid 1990s, and a sufficient body of randomized evidence had accumulated by 2000 to demonstrate a reliable, consistent, and statistically significant benefit of collaborative care over standard care. Since 2000, the overall effect size for collaborative care has remained relatively stable within an effect size between SMD 0.20 and 0.29.

Our results confirm that collaborative care is effective in improving short-term outcomes in depression and, to our knowledge, summarize for the first time the emerging evidence of longer-term benefit. We believe ours is the most comprehensive review of this area to date and builds on previous work by correcting common methodological limitations and in exploring important sources of heterogeneity. The totality of evidence, when given using cumulative meta-analysis, shows that further trials are unlikely to overturn this positive result. Several areas deserve further consideration.

This is a substantial randomized evidence base and should help form a baseline for planning and delivering services. The magnitude of effect that can be expected in practice is moderate but comparable to other more intensive forms of face-to-face psychotherapy and is likely to be cost-effective. As a population-level strategy to improve the management of depression in a greater number of patients, collaborative care has the potential to substantially reduce the global burden of illness associated with depression. In addition, a sustained benefit over the longer-term, even if this were of small to moderate magnitude, would also improve population well-being by reducing the number of days with depression and disability.

Collaborative care can be designed with varying levels of intensity and requires careful consideration in its implementation. Our review found important between-study heterogeneity, and we have used this variability to explore some of the issues in design and implementation that affect the magnitude of effectiveness in individual trials. This meta-regression analysis is exploratory and involves making observational associations within a randomized literature, and loses the power of causal inference. The alternative approach is to plan prospective factorial trials with many arms to test all possible permutations of important aspects of the intervention. Such an approach is costly, time consuming, and unlikely to be undertaken, and in the absence of such trials meta-regression is a viable and efficient approach.

When we undertook our exploration of heterogeneity using meta-regression, we found that collaborative care facilitates improved concordance and that there was a dose-response relationship between medication use and improved depression outcomes. Other positive relationships included the use of regular and planned supervision, and case managers with a specific mental health background. These factors may serve to enhance thera-

Figure 7. Cumulative random-effects meta-analysis of collaborative care vs standard care: depression outcomes at 6 months.
peutic engagement and to work within a biopsychosocial model, where medication concordance and therapeutic alliance are seen as important in improving patient outcomes. The duration of case management and number of sessions were unrelated to effect size, and even brief interventions, such as telephone follow-up, were effective. The addition of psychotherapy was not generally associated with improved outcome. This finding should be interpreted with care because some trials offered patients a choice of either psychotherapy or medication enhancement before enrollment. Improved outcomes in both interventions may in some way reflect a strong initial preference for one or the other of these treatments. Suggestive evidence emerged about the importance of fidelity to the collaborative care model offered by Katon et al. Studies with all 3 elements of collaborative care in place (a case manager, a primary care physician, and access to specialist input) tended to be more effective and were certainly more homogeneous than those studies with less model fidelity.

A further potential source of between-study heterogeneity might be the quality of usual care offered. Those studies with quality care in place might have less to gain from quality enhancements such as collaborative care, and vice versa. The quality of usual care is not described in any detail in the studies included in this review, and it was impossible to explore this further. This remains a limitation of the present review and is an unexplored alternative explanation for some of the associations we found.

A striking finding of our review was the presentation of this evidence, to our knowledge, for the first time, using cumulative meta-analysis. Sufficient randomized evidence had emerged by 2000 to demonstrate the effectiveness of collaborative care beyond conventional levels of statistical significance. Further and subsequent randomized trials have only sought to increase the precision of existing estimates of effectiveness, and it is unlikely that further randomized evidence will overturn this result. This begs the question of whether further trials of collaborative care are needed.

A recent editorial on the role of collaborative care concluded that “The evidence base is now sufficient for the emphasis to shift from research to dissemination and implementation.” Our review supports that assertion, but only with certainty in the United States. Collaborative care studies conducted outside of the United States yielded nonsignificant results and were subject to a much larger degree of between-study heterogeneity compared with US studies. We believe there is a need for further research in the form of randomized controlled trials to examine how best this intervention can be designed and implemented in well-funded European health care systems and in less-well-funded systems in the developing world. Important evidence in Europe and in the developing world is beginning to emerge. However, further research is needed to help clarify whether this system of care can be translated and implemented in settings other than US managed care.

Trials of collaborative care in the United States are targeted at high-risk groups such as patients with coexisting physical illness or other common psychiatric problems such as anxiety, representing a greater desire to understand how the effectiveness of collaborative care can be extended beyond depression. Compared with outcomes over the short term, there was still some uncertainty regarding the longer-term outcomes of collaborative care. Any further trials of collaborative care should also address the longer-term effects in addition to the longer-term cost-effectiveness of this approach.

Accepted for Publication: August 7, 2006.
Correspondence: Simon Gilbody, MBChB, MRCPsych, DPhil, Department of Health Sciences, University of York, York YO10 5DD, England (sg519@york.ac.uk).
Author Contributions: Study concept and design: Gilbody, Bower, and Richards. Acquisition of data: Gilbody, Bower, Fletcher, and Richards. Analysis and interpretation of data: Gilbody, Bower, Richards, and Sutton. Drafting of the manuscript: Gilbody and Sutton. Critical revision of the manuscript for important intellectual content: Gilbody, Bower, Fletcher, Richards, and Sutton. Statistical analysis: Gilbody and Sutton. Obtained funding: Gilbody and Richards. Administrative, technical, and material support: Gilbody, Bower, and Fletcher. Study supervision.
Financial Disclosure: None reported.
Acknowledgment: We thank the staff at the National Health Service Centre for Reviews and Dissemination for conducting the literature searches, and 2 anonymous peer reviewers for helpful comments.

REFERENCES
