

Impacts of Evidence-Based Quality Improvement on Depression in Primary Care

A Randomized Experiment

Lisa V. Rubenstein, MD, MSPH,^{1,2,3} Lisa S. Meredith, PhD,² Louise E. Parker, PhD,²
Nancy P. Gordon, PhD,⁴ Scot C. Hickey, MA,² Carole Oken, MA,² Martin L. Lee, PhD^{1,5}

¹VA Greater Los Angeles, Los Angeles, CA, USA; ²RAND Health Program, Santa Monica, CA, USA; ³David Geffen School of Medicine and Department of Medicine at UCLA, Los Angeles, CA, USA; ⁴Kaiser Permanente Division of Research, Oakland, CA, USA; ⁵UCLA School of Public Health, Los Angeles, CA, USA.

CONTEXT: Previous studies testing continuous quality improvement (CQI) for depression showed no effects. Methods for practices to self-improve depression care performance are needed. We assessed the impacts of evidence-based quality improvement (EBQI), a modification of CQI, as carried out by 2 different health care systems, and collected qualitative data on the design and implementation process.

OBJECTIVE: Evaluate impacts of EBQI on practice-wide depression care and outcomes.

DESIGN: Practice-level randomized experiment comparing EBQI with usual care.

SETTING: Six Kaiser Permanente of Northern California and 3 Veterans Administration primary care practices randomly assigned to EBQI teams (6 practices) or usual care (3 practices). Practices included 245 primary care clinicians and 250,000 patients.

INTERVENTION: Researchers assisted system senior leaders to identify priorities for EBQI teams; initiated the manual-based EBQI process; and provided references and tools.

EVALUATION PARTICIPANTS: Five hundred and sixty-seven representative patients with major depression.

MAIN OUTCOME MEASURES: Appropriate treatment, depression, functional status, and satisfaction.

RESULTS: Depressed patients in EBQI practices showed a trend toward more appropriate treatment compared with those in usual care (46.0% vs 39.9% at 6 months, $P=.07$), but no significant improvement in 12-month depression symptom outcomes (27.0% vs 36.1% poor depression outcome, $P=.18$). Social functioning improved significantly (mean score 65.0 vs 56.8 at 12 months, $P=.02$); physical functioning did not.

CONCLUSION: Evidence-based quality improvement had perceptible, but modest, effects on practice performance for patients with depression. The modest improvements, along with qualitative data, identify potential future directions for improving CQI research and practice.

KEY WORDS: quality improvement; depression; continuous quality management; social function.

DOI: 10.1111/j.1525-1497.2006.00549.x

J GEN INTERN MED 2006; 21:1027-1035.

The current quality crisis highlights the gap between what we know, based on research evidence, and what primary care practices actually deliver.^{1,2} In the case of depression, appropriate treatment (antidepressants or psychotherapy) improves outcomes.³⁻⁵ Only a minority of primary care

patients, however, complete a minimally adequate course of depression treatment,^{6,7} despite a large national investment in depression care.⁸ Dissemination of clinical guidelines with or without additional clinician education is ineffective in rectifying this situation.^{4,5,9-11} New care models for depression¹²⁻¹⁷ that change practice structure¹⁸ to facilitate high-quality depression care are cost-effective and relatively affordable,¹⁹⁻²¹ but are difficult for health care systems and practices to implement, in part because they require significant organizational change. This study tests the impacts of a modified version of continuous quality improvement (CQI) when used to help health care systems design and implement evidence-based care models for depression.

Continuous quality improvement and related methods are among the few systematic approaches available to help health care practices plan and implement organizational change.²²⁻²⁴ Originally CQI focused on finding problems and solutions within individual practice settings rather than seeking outside evidence.²⁴ Later, researchers and others introduced an evidence focus on the *process* of care¹⁸ into CQI by charging teams with implementing evidence-based clinical treatment guidelines. Studies of depression guideline implementation using this form of CQI, however, showed no effects on performance of appropriate depression care at the practice level or depression outcomes.²⁵⁻²⁷

A possible reason why CQI was not effective in these studies is that although teams accessed evidence on the *process* of care, they were left to their own devices in terms of modifying the *structure* of care¹⁸ (e.g., by redesigning the care model). We modified CQI to encourage QI teams to focus on increasing appropriate depression treatment *and* on using evidence-based care models to do so. We included a focus on effective provider behavior change strategies that QI teams could incorporate into their models.²⁸⁻³⁰ We termed the modification evidence-based quality improvement (EBQI), and evaluated its impacts by assessing the performance of experimental and usual care practices on measures of depression-related care and outcomes.

Continuous quality improvement trials are complex. They test both the effectiveness of the CQI method in helping practices achieve effective organizational changes, and the effectiveness of the changes themselves. Practices, not researchers, control the change process initiated by CQI. In this paper, in addition to assessing the effects of EBQI, we aim to support future improvements in CQI practice and research by

No conflicts of interest.

Address correspondence and requests for reprints to Dr. Rubenstein: VA Greater Los Angeles, 16111 Plummer (152), North Hills, CA 91343 (e-mail: lisa_rubenstein@rand.org).

Manuscript received February 02, 2005

Initial editorial decision April 25, 2005

Final acceptance April 18, 2006

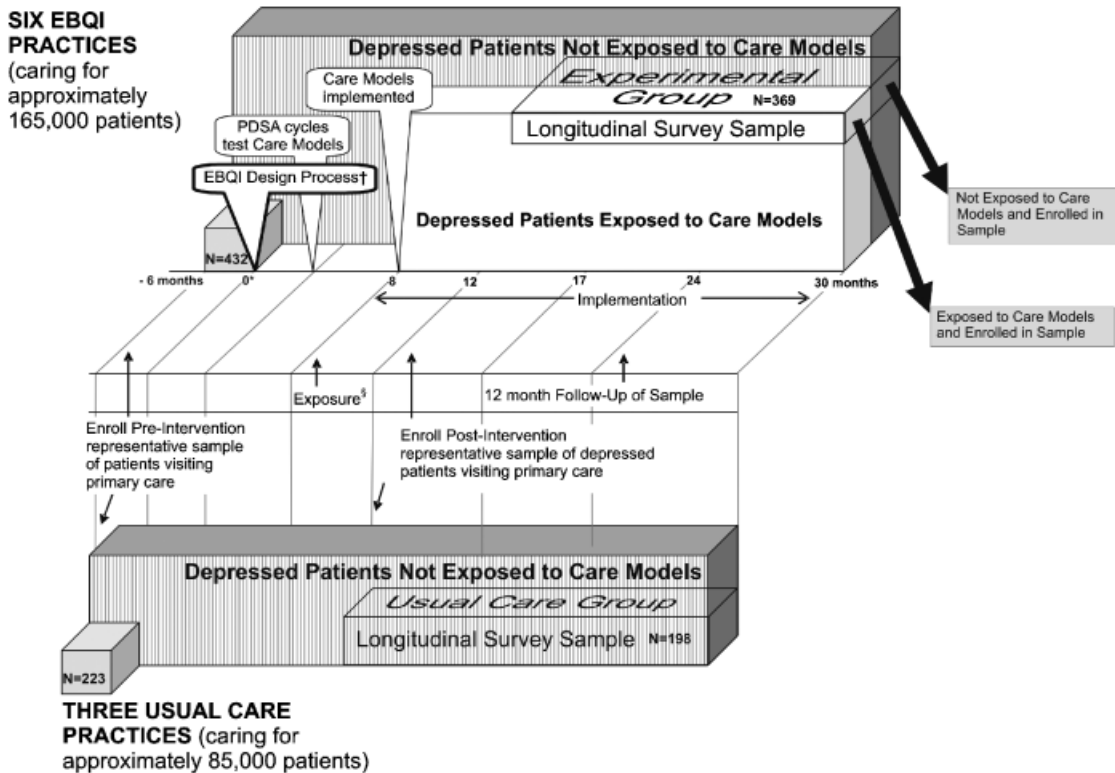


FIGURE 1. MHP intervention and evaluation design and timeline as implemented.[‡]

*Zero indicates the start of the intervention. The overall timeline ranges from – 6 months (start of the preintervention survey) to 30 months.

[†]The callouts at the left hand side of the figure identify the design, Plan-Do-Study-Act (PDSA) and implementation phases of evidence-based quality improvement (EBQI). The EBQI design call-out represents the study intervention (researchers initiated and structured the EBQI design process in the experimental practices).

[‡]The three-dimensional boxes in the diagram represent patient populations. Numbers (n) of patients for the pre and postintervention samples are listed on the appropriate boxes.

- In the experimental practices, the white box labeled “depressed patients exposed to care model” represents patients exposed to the new depression care models developed by QI teams (e.g., seen by a depression care manager), whether or not they were enrolled in the postintervention survey sample, and contrasts with all unexposed depressed patients visiting experimental practices (vertical stripes). Note: we do not know what proportion was exposed.

- The boxes labeled “longitudinal survey sample” represent patients enrolled in the postintervention evaluation. In experimental practices, the enrolled patients exposed to new care models (white area) contrast with the proportion not exposed (vertically striped area).

[§]The time period labeled “exposure” indicates the mean duration of the window (i.e., between 8 and 12 months) during which patients who visited an intervention practice before their enrollment in the longitudinal survey sample might have been exposed to the new care models.

providing qualitative detail on our intervention and evaluation process.

and method of accessing locally available mental health staff. No practices or clinicians refused to participate.

METHODS

Protocol

Our study was a cluster randomized experiment^{31,32} comparing EBQI with usual care conducted both within a nonprofit community organization (Kaiser Permanente of Northern California, or Kaiser) and a government-funded health care system (Veterans Administration, or VA). Institutional review boards at Kaiser, VA, and the RAND Corporation approved the study. We approached 9 large managed primary care practices in California (6 from Kaiser and 3 from VA) identified by regional organizational leaders as not having previously initiated or participated in formal depression care improvement. Each practice had its own leadership team, staffing, patients,

Assignment and Practice Characteristics

We matched the 9 enrolled practices into 3 triplets (2 Kaiser triplets and 1 VA triplet) based on patient ethnicity and urban or suburban location. Within each triplet, the statistician randomly assigned 2 practices as experimental and 1 as usual care.

Interventions

In both usual care and experimental groups, we mailed clinicians copies of clinical practice guidelines for depression.³³ We previously published qualitative information on the EBQI intervention.^{34,35} Figure 1 summarizes the major phases and timeline, including the QI design process, care model start-up, and maintenance (care model in place). The

entire timeline as pictured in Figure 1 spans the period from 6/96 to 12/98.

Only practices in the experimental group participated in EBQI. The major EBQI activities are summarized in Appendix 1, available on the web. Researchers provided QI team leaders with (1) an EBQI manual outlining how teams should carry out the design process, including guidelines for 16 hours of meeting time; (2) a depression care tool kit^{12,36,37}; (3) articles on effective care models for improving depression care or changing provider behavior; and (4) the top 5 guideline-based goals organizational senior leaders identified through an expert panel process (Appendix 2, available on the web).^{34,38} Researchers additionally provided 2 brief orientation sessions to each QI team, but had no ongoing involvement in the QI process; participating organizations and their QI teams, not researchers, were the decision makers. Researchers also provided participating organizations with limited funds (about \$25,000 total) that could be used to pay for release time for QI participants during the EBQI design phase.³⁴ Participating organizations provided all other resources that EBQI teams and practices used. Existing quality improvement committees at each organization reviewed and approved each EBQI team's proposed depression care model. Trained qualitative observers documented EBQI team care models and meetings and conducted semi-structured interviews.^{34,35}

Assessment of Preintervention Equivalence of Groups

We approached consecutive patients in each clinic, asked for their consent, assessed demographic and health status information by survey, and accessed their computer medical records for the succeeding 6 months for mental health specialty visit codes or depression diagnosis codes.

Outcomes

We began outcome assessment after EBQI structural changes were in place in each participating practice (Fig. 1). We enrolled a representative cross section of patients with current major depression by systematically approaching consecutive patients attending primary care appointments over a 5- to 8-month period in each clinic, screening them for depression symptoms, and testing them for major depression diagnosis. We administered follow-up surveys 6 and 12 months after enrollment. Implemented care models served enrolled and non-enrolled patients alike; enrolled patients were not informed about their depression or depression care options and clinicians did not know which patients were enrolled. Sample size calculations, accounting for cluster, showed 80% power to detect a 20% difference in appropriate treatment (from 40% to 60%) with a sample of 56 patients per practice, or 504 patients overall.

The research team identified appropriate depression treatment and recovery from depression as primary outcomes. We also targeted impacts on social functioning, which is critically important to emotional health³⁹ and more responsive to primary care depression interventions than physical functioning.^{14,40} In both study organizations, senior leaders identified improved patient education and participation in care as 1 of the 5 key goals for EBQI (Appendix 2); we evaluated patient

exposure to depression education and patient satisfaction with participation in care as corresponding measures.

We expected maximum increases in completion of appropriate treatment at 6 months and outcome effects at 12 months. During each survey wave, we queried patients about treatment during the previous 6 months. If EBQI practices were more proactive, we expected they would find and begin to treat our enrolled depressed patients sooner. Treated cases should accrue rapidly during the 6 months after enrollment, with the completed patients beginning to discontinue treatment thereafter. Cases of remission following treatment should start accruing in the initial 6-month window and continue to accrue through 12 months after enrollment. We expected usual care practices to be less proactive, and thus to find and treat enrolled patients at a delayed and relatively low but continuous rate, and to have relatively high rates of prolonged, partial treatment.

Enrollment began with an initial 10-minute self-administered survey that included depression screening questions based on the Composite International Diagnostic Interview (CIDI).⁴¹ Patients with current, frequent lack of pleasure, or depressed mood were eligible to consent to follow-up surveys and record review. After consent, we administered the full CIDI.⁴¹ We followed only patients with a structured diagnosis of major depression (Appendix 3, available on the web, shows enrollment yield at each step).

As dependent variables, we measured appropriate treatment with antidepressants or psychotherapy¹⁴ and patient exposure to depression education³⁶ based on survey responses at baseline, 6, and 12 months, using previously validated indicators. We measured patient satisfaction with participation in care on a 0 to 100 scale constructed from 9 items on satisfaction with quality of communication with the patient's clinician, explanation of tests and treatments, involvement in decisions about care, and ease of getting help. Items are scored from 1 (low satisfaction) to 5 (high satisfaction). This scale was administered only at 6 months, was modified from a previous scale,⁴² and had a Cronbach's α of 0.94. A 20% improvement on the single item "involvement in decisions about care" from this scale has been found to be associated with a 4% to 5% increased probability of receiving guideline-concordant care and a 2% to 3% increased probability of depression resolution over the following 18 months.⁴³ At baseline, we measured patient satisfaction with care for personal problems using a single item scored from 1 to 5.⁴⁴ We measured poor depression outcome based on a previously tested summary measure⁴⁴ of whether the patient remained depressed (scored below cutoffs) on all of the following: (1) current major depression symptoms with functional impairment based on CIDI items⁴¹; (2) the Center for Epidemiological Studies Depression (CESD) Scale^{44,45}; and (3) the Mental-health Composite Score (MCS) from the SF-12.^{44,46} We assessed the effects of physical health or emotional problems on social functioning using a 0 to 100 scale (higher is better) from the SF-36.⁴⁷ We assessed each variable, other than satisfaction, at baseline, 6, and 12 months.

As covariates, we measured age, sex, completion of high school, ethnicity, count of chronic diseases, marriage, alcohol use (3 questions),⁴⁸ dysthymia,^{14,41} and household wealth.⁴⁹ We also assessed the timing of enrollment relative to any prior visits that the patient had made while EBQI practices' new care models were already in place, thus reflecting potential

Table 1. Evidence-Based Quality Improvement (EBQI) Team Design and Implementation of Depression Care Model Components

| Depression Care Model Components | Major Care Model Components Planned or Implemented by Each Practice | | | | | |
|--|---|----|----|----|--------------|----|
| | KP Practices | | | | VA Practices | |
| | #1 | #2 | #3 | #4 | #1 | #2 |
| Provider education and decision support | | | | | | |
| Presentations, seminars, written materials | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Face-to-face detailing on depression care | — | — | — | ✓ | ✓ | — |
| Individual provider feedback on performance | — | — | — | — | ✓ | — |
| Patient education | | | | | | |
| Patient education classes | ✓ | ✓ | — | ✓ | — | — |
| Written materials | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Screening/detection | | | | | | |
| Nurses flag charts for suspected depression | — | — | P | P | — | P |
| Annual screening policy with computer reminders | — | — | — | — | ✓ | ✓ |
| Monitoring/enforcement activities carried out | — | — | — | — | ✓ | — |
| Assessment | | | | | | |
| Provider depression assessment worksheet | ✓ | ✓ | — | — | ✓ | — |
| Provider assessment reminders | — | — | ✓ | — | ✓ | — |
| Care management | | | | | | |
| By a non-MD clinician | ✓ | ✓ | ✓ | — | — | — |
| Collaboration with mental health specialists | | | | | | |
| Improved referral process to mental health specialty | — | — | — | — | I | I |
| Psychiatrists give feedback to primary care | P | P | — | — | — | P |

P, planned but not implemented; ✓, planned and implemented; I, implemented but not planned.

prior exposure to improved depression care among experimental group patients (see Fig. 1).

Blinding

Independent telephone interviewers blinded to treatment group conducted outcome assessments of patients and practices. Practices and clinicians were blinded as to which patients participated in the evaluation, except in the case of 7 patients deemed to be at substantial risk of suicide. Patients were not told whether they attended an experimental or a usual care practice.

Data Analysis

We compared patients in practices assigned to the experimental group with those in practices assigned to usual care. In our final models, we used analysis of covariance⁵⁰ including the baseline value of the tested outcome with logistic regression for categorical and ordinary least squares regression for continuous measures, using SAS software, Version 6.12. We used the sandwich estimator, also known as the robust variance estimator,^{51,52} to adjust standard errors for hierarchical sampling with clustering of patients in practices. We used regression parameters to generate a predicted percentage or mean for each dependent variable at 6 or 12 months based on all covariates. We weighted data for the probability of enrollment and attrition at each step. We carried out 1 set of subgroup regression analyses by classifying EBQI practices into 2 subgroups based on the theoretical strength of their care models and comparing them with usual care practices (as the reference group).

RESULTS

We assessed equivalency of practice characteristics across groups before any changes in care had been initiated based on a sample of 655 patients (432 in practices assigned to EBQI and 223 in usual care) (Fig. 1), and on practice characteristics. We found no significant differences (not shown) between experimental and usual care practices in mean self-reported patient age or distribution across race, gender, marital status, education, or health status. There were also no significant differences in the percent of patients having a mental health specialty visit or in the percent diagnosed with depression.

In terms of other practice characteristics, 1 of the 3 usual care practices was larger in size than the remaining 8 practices, with 111 physicians compared with between 22 and 45 clinicians for all remaining practices (mean 33). Two experimental and 2 usual care practices included resident physicians. Ratios of support staff and of mental health specialists per primary care clinician were similar for experimental and usual care groups, but different across organizations with VA practices having lower levels of support staff and higher levels of mental health specialists.³⁵

Through our qualitative analyses, we examined the effect of EBQI on team implementation of potentially effective plans. We found that teams implemented most planned elements as well as some elements that had not been planned during the design phase (Table 1). Based on this information, and before any quantitative data analysis, we judged KP Teams #1, #2, and #4 and VA Team #1 to have implemented care models that were sufficiently adherent to the depression care and provider behavior literature to have the potential to affect depression outcomes. For example, the adherent KP teams implemented care management by a non-MD, a key component of evidence-based models. Veterans Administration Team #1, in view of

Table 2. Demographic and Functional Status Characteristics of Experimental and Usual Care Practices at Survey Baseline*

| Characteristic | Experimental (n=369) | | | Usual Care (n=198) | | | P Value |
|--|----------------------|---------------------|----------------------|--------------------|---------------------|----------------------|---------|
| | Raw No. Responding | Weighted Mean | Weighted % (Raw No.) | Raw No. Responding | Weighted Mean | Weighted % (Raw No.) | |
| % (n) white vs not white [†] | 366 | — | 75.3 (288) | 195 | — | 75.5 (147) | .95 |
| % (n) American Indian | — | — | 1.8 (6) | — | — | 1.6 (3) | — |
| % (n) Asian | — | — | 6.9 (21) | — | — | 4.2 (8) | — |
| % (n) African American | — | — | 5.5 (18) | — | — | 7.9 (16) | — |
| % (n) hispanic | — | — | 10.6 (33) | — | — | 10.9 (21) | — |
| % male | 369 | — | 46.9 (163) | 198 | — | 46.3 (71) | .90 |
| % married | 365 | — | 42.1 (139) | 197 | — | 47.7 (99) | .21 |
| % working | 366 | — | 63.1 (237) | 197 | — | 58.1 (118) | .26 |
| % poor health status | 360 | — | 7.2 (26) | 196 | — | 9.1 (19) | .45 |
| % 3 or more chronic diseases | 360 | — | 44.3 (166) | 197 | — | 46.9 (93) | .58 |
| % less than high school education | 366 | — | 31.7 (103) | 197 | — | 30.5 (57) | .78 |
| % poor depression outcome | 369 | — | 59.0 (218) | 198 | — | 67.6 (133) | .05 |
| % baseline appropriate treatment | 369 | — | 39.4 (145) | 198 | — | 46.1 (89) | .14 |
| % satisfied or very satisfied with care for personal problems | 321 | — | 47.6 (150) | 190 | — | 40.0 (75) | .10 |
| % with dysthymia | 369 | — | 18.3 (68) | 198 | — | 13.1 (28) | .12 |
| % with no alcohol use | 366 | — | 38.9 (138) | 198 | — | 39.1 (79) | .92 |
| Age (y), mean ± SD | 369 | 48.0 ± 14.6 | — | 198 | 47.4 ± 12.7 | — | .87 |
| Highest level of education (y), mean ± SD | 366 | 13.9 ± 2.2 | — | 197 | 13.9 ± 1.8 | — | .91 |
| Household income, mean ± SD | 333 | 42,871.4 ± 50,202.4 | — | 176 | 46,607.5 ± 51,540.3 | — | .44 |
| Mental health composite of the SF-12, mean ± SD [‡] | 353 | 33.1 ± 10.3 | — | 190 | 31.8 ± 9.3 | — | .17 |
| Physical health composite of the SF-12, mean ± SD [‡] | 353 | 45.3 ± 11.2 | — | 190 | 43.3 ± 11.2 | — | .06 |
| Social functioning, mean ± SD [‡] | 369 | 45.4 ± 21.4 | — | 198 | 46.3 ± 18.5 | — | .61 |
| Alcohol use, mean ± SD [§] | 366 | 118.2 ± 341.8 | — | 198 | 95.7 ± 259.7 | — | .43 |
| Window before baseline survey and after date new care models were in place in experimental practices, in days, mean ± SD | 369 | 135.4 ± 113.8 | — | 198 | 131.1 ± 97.2 | — | .66 |

*Because means and percents are weighted to represent the full practice populations, percents will not exactly represent the raw number in parentheses over the analytic n (raw number responding).

[†]The distribution of ethnicities across white and all others is also not significant by χ^2 ($P = .08$).

[‡]Higher scores indicate better health on these scales.

[§]Alcohol use is calculated by multiplying the quantity times the frequency of drinking alcohol in the past 12 mo.

opposition to care management by VA leadership, used screening, computer clinical reminders that mandated a follow-up action before the visit could be closed, and improved mental health specialty access.⁵³ The remaining 2 teams (KP Team #3

and VA Team #2) implemented educational strategies known not to affect outcomes.^{10,11} VA Team #2 implemented reminders with no enforcement, a strategy previously shown to be ineffective for major depression.^{9,40}

Table 3. Depression Treatment and Outcomes for Patients in Experimental Versus Usual Care Practices

| Process or Outcome/Survey Point | Analytic N | Percentage or Mean [†] (95% CI) | | Significance: P Value | Effect Size |
|--|------------|--|----------------------|-----------------------|-------------|
| | | Experimental Practices | Usual Care Practices | | |
| % completing appropriate treatment | | | | | |
| 6 mo | 434 | 46.0 (41.3 to 50.6) | 39.9 (34.7 to 45.2) | .07 | 0.17 |
| 12 mo | 400 | 45.6 (37.8 to 53.5) | 47.0 (42.7 to 51.3) | .77 | 0.03 |
| Mean satisfaction with participation in care | | | | | |
| 6 mo | 412 | 57.4 (54.6 to 60.3) | 54.3 (50.9 to 57.6) | .02* | 0.14 |
| Mean social functioning (SF) | | | | | |
| 6 mo | 434 | 61.3 (55.9 to 66.7) | 58.6 (54.7 to 62.5) | .18 | 0.07 |
| 12 mo | 399 | 65.0 (59.1 to 70.8) | 56.8 (52.7 to 60.9) | .02* | 0.23 |
| Percentage poor depression outcome | | | | | |
| 6 mo | 434 | 33.7 (25.9 to 41.4) | 36.4 (30.5 to 42.2) | .50 | 0.06 |
| 12 mo | 400 | 27.0 (18.1 to 35.8) | 36.1 (24.0 to 48.4) | .18 | 0.12 |

* $P < .05$.

[†]All regressions controlled for covariates (age, sex, completion of high school, ethnicity, count of chronic diseases, marriage, alcohol use, dysthymia, household wealth, timing of enrollment) and baseline values of the dependent variable. Satisfaction with participation in care is controlled for baseline satisfaction with care for personal problems.

The evaluation sample, enrolled after new depression care models were in place in experimental practices, included 567 patients (369 experimental and 198 usual care) with major depression. As shown in Table 2, there were no significant differences between patients in EBQI and usual care practices at baseline in demographics, physical or mental health-related quality of life, or in the timing of the baseline survey in relationship with experimental group patients' potential exposure to new care models. Evidence-based quality improvement practices had a 9% lower proportion of patients who met criteria for our depression outcome at baseline than did usual care practices ($P=.05$).

As shown in Table 3, controlling for baseline values, by 6 months after baseline 46.0% of patients in experimental practices had completed appropriate depression treatment compared with 39.9% in usual care practices. This difference of 6.0% (confidence interval [CI] 0.00, 0.12, $P=.07$, effect size 0.17) was at a trend level of significance. Also, by 6 months, patient satisfaction with participation in care was 3 points higher for patients in experimental practices compared with usual care, and this difference was statistically significant (CI 0.94, 5.40, $P=.02$, effect size 0.14).

By 12 months, also shown in Table 3, depressed patients in EBQI practices had significantly higher levels of social functioning, scoring 8.2 points higher (better) on the social functioning scale of the SF-36 than did those in usual care practices (CI 2.8, 13.6; $P=.02$, effect size 0.23). Also, by 12 months, 27.0% of patients in EBQI practices were experiencing a poor depression outcome, compared with 36.1% of patients in usual care practices. This difference of 9.1% (CI 21.1, 2.6, $P=.18$, effect size 0.12) was not significant. Results for the mean number of depression symptoms on the CESD and the mean score on the mental health component of the SF-12, both of which are components of our measure of poor depression outcomes, similarly favored the experimental group but were not significant. There were also no significant differences in exposure to or helpfulness of educational classes or materials.

In our subgroup analysis, at 6-month follow-up 46.4% of 247 depressed patients in the 4 EBQI practices we had judged as implementing theoretically stronger care models had completed appropriate depression treatment, a significant difference of 6.4% from usual care (CI 0.01, 0.12, $P=.05$, effect size 0.26). At 12 months, 21.9% of patients in practices with more theory-based care models were experiencing a poor depression outcome, compared with 36.1% of patients in usual care, a significant difference of 14.2% from usual care (CI 23.8, -4.7, $P=.03$, effect size 0.31).

DISCUSSION

As is the case for other chronic conditions,^{1,2,54} research indicates that national performance improvement for depression is unlikely to occur without widespread implementation of new depression care models.⁵⁵ Effective depression care models are relatively low cost once implemented, but require organizational change.⁵⁶ Continuous quality improvement-based methods help organizations and practices self-design and implement improvements. In this study, in contrast to previous CQI studies, practice-based QI teams received EBQI manuals, 5 key senior leader goals for depression care, a depression care model tool kit, and relevant literature. Organizational quality improvement leaders also reviewed and approved team pro-

posals before implementation. Our results signal both encouragement and caution to those interested in using CQI-based methods for moving the abundant knowledge about effective depression care models into routine use in primary care settings. Our study also points toward ways to improve current QI and QI evaluation approaches.

The study signals encouragement because it shows a perceptible effect of EBQI on practice performance and 2 outcomes, patient participation and social functioning, whereas previous studies on CQI for depression²⁵⁻²⁷ have been entirely negative. These results suggest that the types of improvements we made to CQI are on the right track. Conversely, these findings caution against assuming that current CQI-based methods will produce large impacts at the practice population level within short time frames. Effects on depression outcomes were absent or minimal.

We consider the low-level effects we observed on social functioning and satisfaction with participation in care to be early sign posts on the path toward improving depression outcomes through routine QI. These markers are relevant to improving depression outcomes, as shown in other studies^{39,43} but may be achievable through nonspecific care model changes such as patient and clinician education⁴⁰ in the absence of specific changes such as care management that directly support completion of major depression treatments.

Our results are most applicable to managers in large practices or health care systems who intend to improve performance across diverse clinicians and practices. We aimed to test a quality improvement method under routine conditions. By design, no practices we approached had participated in depression quality improvement, and although all agreed to participate, they were not best-case examples. For example, the mental health leadership for the 2 VA practices openly opposed primary care treatment of depression,^{34,53} a realistic impediment likely to be encountered in some practices. Some KP primary care leaders were not supportive of depression care improvement.³⁴ Nevertheless, all QI teams developed and implemented improvement plans. Some team plans appeared appropriate for bringing a practice from precontemplation into contemplation of depression care improvement.^{57,58} As we showed in previous work, the practices implementing theoretically weaker care models had lower levels of team leader expertise and/or leadership support.³⁴ A stepped QI model that adjusts a practice's goals to its readiness for change may improve return on QI investment among low readiness practices.⁵⁹

We provided EBQI teams with senior organizational leader priorities, identified using expert panel methods^{34,35,38} as guidance. Teams responded to these priorities, with no KP teams implementing screening and no VA teams implementing primary care-based treatment support through care management. Both organizations prioritized improving patient knowledge and engagement, and nearly all teams addressed this goal, possibly accounting for the improvements we saw in patient participation. Both organizations also prioritized improving clinician assessment of depression. A previously published analysis of suicide assessment rates based on this and a companion study suggests that our experimental practices improved clinician assessment.⁶⁰

The results of this study, and our presentation of these results, should challenge researchers to consider which evaluation designs and reporting standards should be used for QI

interventions. In our design, like other randomized evaluations of CQI,^{22,25-27,61,62} enrolled patients were not told whether they were in the experimental group and study clinicians did not know which patients were enrolled, reflecting the goal of testing impacts on practice performance. Other cluster randomized studies of care models test outcomes on patients enrolled based on willingness to be referred to the study care model, which may only be available to research patients.^{14,15} Some care model studies are randomized at the patient level.¹² Some are nonrandomized.⁶³ These types of studies serve different purposes; yet, researchers or managers may not distinguish between them in drawing conclusions. In addition, features often not reported in QI studies may affect how the information should be used, such as the degree to which researchers control care model design, the quantity of research resources provided to teams and organizations, and the characteristics of participating practices and organizations. Consort Criteria,⁶⁴ the current standard reporting criteria for randomized trials, address some but not all relevant design issues. More comprehensive theory and guidelines for using, reporting, and interpreting intervention and evaluation designs for QI studies are needed.⁶⁵

Given our design, even the minimal changes we observed may be a cause for cautious optimism because the depression care models in our study practices were no more available to the representative depressed patients enrolled in our evaluation than to any other depressed patient in the practice. Experimental practices were visited by a total of about 165,000 patients during the year post-implementation. If we assume that about 9% of primary care patients have major depression, a 9% lower rate of poor depression outcomes such as we observed in experimental compared with usual care practices represents about 1,200 additional treatment successes in the experimental group. Estimating a 60% recovery rate from treatment, these additional treatment successes represent about 2,000 treated patients, and as many as 3,500 identified for possible treatment through screening or case-finding. Reaching a significant difference at a practice level is challenging, yet is what we must aim for if we wish to show true impact on routine care.

We carried out 1 hypothesis-generating subgroup analysis. Evidence-based quality improvement's lack of effect on depression-specific outcomes could have been due to ineffectiveness in stimulating practices to adopt evidence-based improvements or due to ineffectiveness of evidence-based improvements as implemented by practices. Our results, showing that EBQI practices implementing theoretically stronger care models yielded better depression outcomes, provide support for future efforts to improve the link between participating in EBQI and implementing evidence-based care model elements. For example, initial and possibly periodic review of teams' care models to assess fidelity to evidence shows promise.^{34,59} More centralized, expert-supported QI methods also appear promising for producing models with higher fidelity to evidence.^{15,34}

Our study has limitations. Our power to detect differences in outcomes was limited by the small number of practices involved, and the conservative adjustment for hierarchical sampling that we used to account for this.^{51,52} The organizations in our study were large, not-for-profit, staff-model health maintenance-type organizations, which have different depression-related characteristics than other types of managed care or fee-for-service practices.⁶⁶ Findings may not be generalizable across

practice types. Also, some QI approaches purposely target practices that are ready for change. In contrast, in this study organizations selected practices for not having demonstrated prior interest in depression. Our subgroup analysis is hypothesis generating only, does not address causality, and is not relevant to conclusions regarding EBQI effectiveness. Finally, we report only research support dollars provided for EBQI, not costs.

In summary, this study is the first randomized experiment showing that practices can self-design and self-implement improved depression care models using a CQI-based approach and gain a perceptible impact on depression-related practice outcomes. Looking beyond its outcomes, our study shows the complexity of both real-world QI and the evaluation designs needed for studying it. By exposing readers to this complexity, we hope to stimulate the development of new intervention approaches and evaluation standards that more fully take into account the needs and challenges of translating research evidence on effective care models into routine care solutions.

L.V.R. was the principal investigator and is guarantor. She designed the study along with L.S.M. and L.E.P. C.O. led data collection. Data analysis was carried out by L.V.R., M.L.L., and S.C.H. with support from all authors. All authors collaborated on interpretation of the data and writing the manuscript. The authors acknowledge the invaluable contributions of QI team leaders, as well as Christy Klein, BA, Mary Abdun-Nur, MA, Bernadette Benjamin, MS, Cathy Sherbourne, PhD, Becky Mazel, PhD, Chantal Avila, MA, and James Chiesa, MS. Robert Brook, MD, ScD, and Laurence Rubenstein, MD, MPH, reviewed and critiqued the manuscript. The study could not have taken place without the support of Kaiser Permanente in Northern California and of the VA Greater Los Angeles Healthcare System, Los Angeles, California.

Funding sources: The National Institute for Mental Health, the MacArthur Foundation, and the VA Health Services Research and Development Center of Excellence for the Study of Healthcare Provider Behavior funded the Mental Health Awareness Project upon which this paper is based. The study funders had no involvement in this publication.

REFERENCES

1. **Institute of Medicine.** Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academy Press; 2001.
2. **Institute for the Future.** Health and Health Care 2010: The Forecast, the Challenge. 2nd edn. New York: John Wiley & Sons; 2002.
3. **Schulberg HC, Katon W, Simon GE, Rush AJ.** Treating major depression in primary care practice: an update of the agency for health care policy and research practice guidelines. *Arch Gen Psychiatry.* 1998;55:1121-7.
4. **Depression Guideline Panel.** Clinical Practice Guideline, Depression in Primary Care: Vol. 1. Detection and Diagnosis (AHCPR Publication No. 93-0550). Rockville, MD: U.S. Department of Health and Human Services, Public Health Service Agency for Health Care Policy and Research; 1993.
5. **Depression Guideline Panel.** Clinical Practice Guideline, Depression in Primary Care: Vol. 2. Treatment of Major Depression (AHCPR Publication No. 93-0551). Rockville, MD: U.S. Department of Health and Human Services, Public Health Service Agency for Health Care Policy and Research; 1993.
6. **Charbonneau A, Rosen AK, Ash AS, et al.** Measuring the quality of depression care in a large integrated health system. *Med Care.* 2003;41:669-80.
7. **Kerr EA, McGlynn EA, Van Vorst KA, Wickstrom SL.** Measuring antidepressant prescribing practice in a health care system using administrative data: implications for quality measurement and improvement. *Jt Comm J Qual Improv.* 2000;26:203-16.

8. **Olfson M, Marcus SC, Druss B, Elinson L, Tanielian T, Pincus HA.** National trends in the outpatient treatment of depression. *JAMA*. 2002;287:203–9.
9. **Rollman BL, Hanusa BH, Lowe HJ, Gilbert T, Kapoor WN, Schulberg HC.** A randomized trial using computerized decision support to improve treatment of major depression in primary care. *J Gen Intern Med*. 2002;17:493–503.
10. **Dobscha SK, Gerrity MS, Ward MF.** Effectiveness of an intervention to improve primary care provider recognition of depression. *Eff Clin Pract*. 2001;4:163–71.
11. **Thompson C, Kinmonth AL, Stevens L, et al.** Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: hamshire depression project randomised controlled trial. *Lancet*. 2000;355:185–91.
12. **Katon WJ, Von Korff M, Lin E, et al.** Collaborative management to achieve treatment guidelines: impact on depression in primary care. *JAMA*. 1995;273:1026–31.
13. **Rost K, Nutting P, Smith J, Werner J, Duan N.** Improving depression outcomes in community primary care practice: a randomized trial of the quEST intervention. Quality enhancement by strategic teaming. *J Gen Intern Med*. 2001;16:143–9.
14. **Wells KB, Sherbourne C, Schoenbaum M, et al.** Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA*. 2000;283:212–20.
15. **Dietrich AJ, Oxman TE, Williams JW Jr., et al.** Re-engineering systems for the treatment of depression in primary care: cluster randomised controlled trial. *BMJ*. 2004;329:602.
16. **Unutzer J, Katon W, Callahan CM, et al.** Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA*. 2002;288:2836–45.
17. **Hunkeler EM, Meresman JF, Hargreaves WA, et al.** Efficacy of nurse telehealth care and peer support in augmenting treatment of depression in primary care. *Arch Fam Med*. 2000;9:700–8.
18. **Donabedian A.** Explorations in Quality Assessment and Monitoring. Vol. I: The Definition of Quality and Approaches to its Assessment. Ann Arbor, MI: Health Administration Press; 1980.
19. **Schoenbaum M, Unutzer J, Sherbourne C, et al.** Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. *JAMA*. 2001;286:1325–30.
20. **Liu CF, Hedrick SC, Chaney EF, et al.** Cost-effectiveness of collaborative care for depression in a primary care veteran population. *Psychiatr Serv*. 2003;54:698–704.
21. **Simon GE, Katon WJ, VonKorff M, et al.** Cost-effectiveness of a collaborative care program for primary care patients with persistent depression. *Am J Psychiatry*. 2001;158:1638–44.
22. **Shortell SM, Bennett CL, Byck GR.** Assessing the impact of continuous quality improvement on clinical practice: what it will take to accelerate progress. *Milbank Q*. 1998;76:593–624.
23. **McLaughlin CP, Kaluzny CA.** Continuous Quality Improvement in Health Care. 2nd edn. Gaithersburg, MD: Aspen; 1999.
24. **Berwick DM.** Continuous improvement as an ideal in health care. *N Engl J Med*. 1989;320:53–6.
25. **Brown Jr. BW, Shye D, McFarland BH, Nicholas GA, Mullooly JP, Johnson R.** Controlled trials of CQI and academic detailing to implement a clinical practice guideline for depression. *Jt Comm J Qual Improv*. 2000;26:39–54.
26. **Horowitz CR, Goldberg HI, Martin DP, et al.** Conducting a randomized controlled trial of CQI and academic detailing to implement clinical guidelines. *Jt Comm J Qual Improv*. 1996;22:734–50.
27. **Solberg LI, Fischer LR, Wei F, et al.** A CQI intervention to change the care of depression: a controlled study. *Eff Clin Pract*. 2001;4:239–49.
28. **Oxman AD, Thomson MA, Davis DA, Haynes RB.** No magic bullets: a systematic review of 102 trials of interventions to improve professional practice. *Can Med Assoc J*. 1995;153:1423–31.
29. **Davis DD, Thomson MA, Oxman AD, Haynes RB.** Evidence for the effectiveness of CME: a review of 50 randomized controlled trials. *JAMA*. 1992;268:1111–7.
30. **Grimshaw JM, Shirran L, Thomas R, et al.** Changing provider behavior: an overview of systematic reviews of interventions. *Med Care*. 2001;39(Suppl):II–45.
31. **Cook TDCDT.** Quasi-Experimentation: Design and Analysis Issues for Field Settings. Boston: Houghton Mifflin Company; 1979.
32. **Grimshaw J, Campbell M, Eccles M, Steen N.** Experimental and quasi-experimental designs for evaluating guideline implementation strategies. *Fam Pract*. 2000;(Suppl 1):11–6.
33. **Agency for Health Care Policy and Research.** Clinical Practice Guideline: Quick Reference Guide for Clinicians. AHCPR Publication No. 93-0552. Rockville, MD: Agency for Health Care Policy and Research; 1993.
34. **Rubenstein LV, Parker LE, Meredith LS, et al.** Understanding team-based quality improvement for depression in primary care. *Health Serv Res*. 2002;37:1009–29.
35. **Rubenstein LV, Meredith L, Parker LE, et al.** Design and Implementation of Evidence-based Quality Improvement for Depression in the Mental Health Awareness Project: Background, Methods, Evaluation, and Manuals, MR1699-NIMH. Santa Monica: RAND Corporation; 2003.
36. **Rubenstein LV, Jackson-Triche M, Unutzer J, et al.** Evidence-based care for depression in managed primary care practices. *Health Aff (Millwood)*. 1999;18:89–105.
37. **Depression Patient Outcomes Research Team (PORT-II).** Partners in care (PIC) tool for an integrated approach to improving care for depression in primary care. Available at: <http://www.rand.org/health/pic.products/>. Accessed February 3, 2006. 2003.
38. **Rubenstein LV, Fink A, Yano EM, Simon B, Chernof B, Robbins AS.** Increasing the impact of quality improvement on health: an expert panel method for setting institutional priorities. *Jt Comm J Qual Improv*. 1995;21:420–32.
39. **Sherbourne C, Schoenbaum M, Wells KB, Croghan TW.** Characteristics, treatment patterns, and outcomes of persistent depression despite treatment in primary care. *Gen Hosp Psychiatry*. 2004;26:106–14.
40. **Rubenstein LV, McCoy JM, Cope DW, et al.** Improving patient quality of life with feedback to physicians about functional status. *J Gen Intern Med*. 1995;10:607–14.
41. **World Health Organization.** Composite International Diagnostic Interview (CIDI) Core Version 2.1 Interviewer's Manual. Geneva: World Health Organization; 1997.
42. **Meredith LS, Orlando M, Humphrey N, Camp P, Sherbourne CD.** Are better ratings of the patient-provider relationship associated with higher quality care for depression? *Med Care*. 2001;39:349–60.
43. **Clever S, Ford DE, Rubenstein LV, et al.** Primary care patients involvement in decision making is associated with improvements in depression. *Med Care*. 2006; 44:398–405.
44. **Sherbourne CD, Wells KB, Duan N, et al.** Long-term effectiveness of disseminating quality improvement for depression in primary care. *Arch Gen Psychiatry*. 2001;58:696–703.
45. **Radloff J.** The CES-D scale. A self-report depression scale for research in the general population. *Appl Psychiatric Measures*. 1977;1:385–401.
46. **Ware JE Jr, Kosinski M, Keller S.** SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Boston, MA: The Health Institute, New England Medical Center; 1995.
47. **Ware JE Jr, Sherbourne CD.** The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473–83.
48. **Bohn MJ, Babor TF, Kranzler HR.** The alcohol use disorders identification test (AUDIT): validation of a screening instrument for use in medical settings. *J Stud Alcohol*. 1995;56:423–32.
49. **Smith JP.** Racial and ethnic differences in wealth in the health and retirement study. *J Hum Resour*. 1995;30(suppl):S158–83.
50. **van Belle G, Fisher LD, Heagerty PJ, Lumley T.** Biostatistics: A Methodology for the Health Sciences. 2nd edn. Hoboken, NJ: Wiley-Interscience; 2004.
51. **McCaffrey DF, Bell RM.** Bias reduction in standard errors for linear regression with multi-stage samples. AT&T Technical Report TD-4S9H9T; 2000.
52. **Williams RL.** A note on robust variance estimation for cluster-correlated data. *Biometrics*. 2000;56:645–6.
53. **Sherman SE, Chapman A, Garcia D, Braslow JT.** Improving recognition of depression in primary care: a study of evidence-based quality improvement. *Jt Comm J Qual Safety*. 2004;30:80–88.
54. **Wagner EH, Austin BT, Von Korff M.** Organizing care for patients with chronic illness. *Milbank Q*. 1996;74:511–44.
55. **Katon WJ.** The Institute of Medicine "Chasm" report: implications for depression collaborative care models. *Gen Hosp Psychiatry*. 2003;25:222–9.
56. **Rosenheck RA.** Organizational process: a missing link between research and practice. *Psychiatr Serv*. 2001;52:1607–12.
57. **Lehman WE, Greener JM, Simpson DD.** Assessing organizational readiness for change. *J Subst Abuse Treat*. 2002;22:197–209.

58. **Weeks B, Helms MM, Ettkin LP.** A physical examination of health care's readiness for a total quality management program: a case study 1. *Hosp Mater Manage Q.* 1995;17:68-74.
59. **Gustafson DH, Sainfort F, Eichler M, Adams L, Bisognano M, Steudel H.** Developing and testing a model to predict outcomes of organizational change. *Health Serv Res.* 2003;38:751-76.
60. **Nutting PA, Dickinson LM, Rubenstein LV, Keeley RD, Smith JL, Elliot CE.** Improving detection of suicidal ideation among depressed patients in primary care. *Ann Fam Med.* 2005;3:529-36.
61. **Solberg LI, Kottake TE, Brekke ML, et al.** Failure of a continuous quality improvement intervention to increase the delivery of preventive services: a randomised trial. *Eff Clin Pract.* 2000;3:105.
62. **Flottorp S, Havelsrud K, Oxman AD.** Process evaluation of a cluster randomized trial of tailored interventions to implement guidelines in primary care—why is it so hard to change practice? *Fam Pract.* 2003;20:333-9.
63. **Wagner EH, Glasgow RE, Davis C, et al.** Quality improvement in chronic illness care: a collaborative approach. *Jt Comm J Qual Improv.* 2001;27:63-80.
64. **Campbell MK, Elbourne DR, Altman DG.** CONSORT statement: extension to cluster randomised trials. *BMJ.* 2004;328:702-8.
65. **Davidoff F, Batalden P.** Toward stronger evidence on quality improvement. Draft publication guidelines: the beginning of a consensus project. *Qual Saf Health Care.* 2005;14:319-25.
66. **Meredith LS, Rubenstein LV, Rost K, et al.** Treating depression in staff-model versus network-model managed care organizations. *J Gen Intern Med.* 1999;14:39-48.

Supplementary Material

The following supplementary material is available for this article online at www.blackwell-synergy.com

Appendix 1. Timing of Evidence-Based Quality Improvement (EBQI) and Research Evaluation Activities.

Appendix 2. The Top Five Priorities for Improving Depression Care at Kaiser and at the VA.

Appendix 3. MHAP Patient Sample Flow.