Cost-effectiveness of Practice-Initiated Quality Improvement for Depression
Results of a Randomized Controlled Trial

Michael Schoenbaum, PhD
Jürgen Unützer, MD, MPH
Cathy Sherbourne, PhD
Naihua Duan, PhD
Lisa V. Rubenstein, MD, MSHS
Jeanne Miranda, PhD
Lisa S. Meredith, PhD
Maureen F. Carney, MS
Kenneth Wells, MD, MPH

Depression is a leading cause of disability worldwide, but treatment rates in primary care are low. Improving quality of care for depression in primary care could potentially increase well-being for patients, their families, and society at large.11,12

Studies suggest that practice-based interventions to improve quality of treatment for depression in primary care can improve short-term clinical outcomes relative to usual care, at modest cost.13-15 One such study of high utilizers found that a practice-based quality improvement (QI) program for depression was cost-effective relative to usual care, with longer-term improvements that included physical functioning.10 In addition, less intensive interventions, such as nurse telephone contact, also improve clinical outcomes, but those relying on clinician training alone may have limited benefits.17,18 Most prior studies have focused on interventions implemented under research protocols in academically affiliated, organized care settings, so the extent to which results can be generalized to more naturalistic practice conditions and diverse types of practices is unclear.

Findings from Partners in Care (PIC) suggest that diverse managed primary care practices can implement such in-
PRACTICE-INITIATED QUALITY IMPROVEMENT FOR DEPRESSION

METHODS

Experimental Design and Sample

PIC is a group-level, randomized controlled trial of practice-initiated QI programs for depression.19 PIC was fielded in 6 nonacademic managed care organizations. Forty-six of 48 primary care practices (clinics) and 181 of 183 clinicians participated. Within organizations, practices were matched into blocks of 3 clusters based on factors expected to affect outcomes (specialty mix, patient socioeconomic and demographic factors, and having on-site mental health specialists). Within blocks, practices were randomized to usual care or 1 of 2 QI programs (QI-meds or QI-therapy).

Study staff screened 27332 consecutive patients in participating practices over a 5- to 7-month period between June 1996 and March 1997 (FIGURE). Patients were eligible if they intended to use the practice over the next 12 months and screened positive for depression, based on stem items from the World Health Organization’s 12-month Composite International Diagnostic Interview (CIDI).20 Patients were considered positive for depression if they reported at least 1 week of depression in the last 30 days, plus 2 or more weeks of depressed mood or loss of interest in pleasurable activities in the last year or persistent depression over the year. This indicator has a 55% positive predictive value for 12-month major depressive or dysthymic disorder by the full CIDI.9 Patients were ineligible if younger than 18 years, not fluent in English or Spanish, or lacking in insurance coverage for intervention therapists.

Of those completing the screener, 3918 were eligible, 2417 were available to confirm insurance eligibility, and 241 were ineligible. Of those reading the informed consent, 1356 (70%) enrolled: 443 in usual care, 424 in QI-meds, and 489 in QI-therapy practices. The study was approved by the institutional review boards of RAND and the practices.

Interventions

Intervention design and implementation are described elsewhere, and all QI materials are available from RAND (www.rand.org/organization/health/pic/products/order.html).19,21 Prior to implementation, practices committed to implementing the programs and the study provided a payment of up to half of the estimated practice participation costs ($35000-$70000). Usual care clinics received depression practice guidelines by mail. The interventions provided practices with training and resources to initiate and monitor QI programs according to local practice goals and resources. Patients and clinicians retained choice of treatment, and their use of intervention resources was optional. In effect, the study served as an external disease management firm, designing the materials, hosting initial training, and offering limited support during implementation.

For both QI-meds and QI-therapy interventions, local practice teams were trained in a 2-day workshop to provide clinician education through lectures, academic detailing, or audit and feedback, and to supervise intervention staff and conduct team oversight. Practice nurses were trained as depression specialists, following a written protocol, to assist in initial patient assessment, education, and motivation for treatment. Practice teams were given patient education pamphlets and videotapes, patient tracking forms, and clinician manuals and pocket reminder cards and were encouraged to distribute them. The materials described guideline-concordant care for depression and presented psychotherapy and antidepressant medication as equally effective.

In the QI-meds program, nurse specialists were trained to support medication adherence through monthly telephone contacts or visits for 6 or 12 months, randomized at the patient level. In the QI-therapy program, practice therapists were trained to provide individual and group cognitive behavioral therapy, following a protocol developed at the San Francisco General Hospital Depression Clinic.22,23 This therapy was available at the primary care co-payment rate (usually $5-$10) for 6 months after enrollment. All patients could have other therapy at the usual co-payment rate. Clinical supervision was provided by local experts assisted by study experts in cognitive behavioral therapy. In all conditions,
patients could have medication, therapy, or both. However, the extra QI-meds or QI-therapy resources made it easier to obtain appropriate medication or therapy, respectively.

Data Collection
Patients were asked to complete the screener, the telephone CIDI for depression, a detailed economic survey by telephone at baseline and 24 months of follow-up, and mail surveys at baseline and 6, 12, 18, and 24 months. Completion rates were 95% and 85% for the baseline and 24-month telephone surveys and 90%, 86%, 84%, 83%, and 85%, respectively, for the mail surveys. Subjects were eligible for follow-up unless they disenrolled from the study. Claims and encounter data from the practices were only consistently available for the first 6 months of patient follow-up, and most practices did not have pharmacy data.

Cost Measures
Intervention Costs. These included screening, intervention materials, initial nurse specialist assessments, and 20 minutes of supervision of nurses and therapists per enrolled patient. We assigned costs to intervention activities based on data from the practices about the average cost of clinic staff. Research-specific costs were excluded. For main analyses, we assumed that follow-up visits to intervention staff were included in patient reports of outpatient visits. In sensitivity analyses, we used data from intervention logs to include such visits as intervention costs (which double counts them if they were also reported by patients directly); our results did not change substantively.

Health Care Costs. We assigned costs to patient-reported counts of emergency department visits, medical and mental health visits, and psychotropic medications used, for each follow-up. Patient report was selected due to the limitations in the available claims and encounter data. In addition, the number of outpatient visits was higher for patient surveys than claims data over the first 6 months, probably due to out-of-practice use or incomplete claims data. We excluded inpatient costs because we did not expect or observe intervention effects on them and had limited precision to analyze them.24

Average costs in 1998 dollars were assigned to each component of patient-reported health care use using a national database of about 1.8 million privately insured individuals (provided by Ingenix, a benefits consulting firm in New Haven, Conn). The Ingenix data included information on provider reimbursements (ie, patient and plan payments, plus coordination of benefits), which we used as a proxy for health care costs. Specifically, we calculated the mean cost per outpatient medical visit ($46), mental health visits ($96), and emergency department visits ($450), respectively, for adults in the Ingenix data; these costs include facility charges, professional fees, and ancillary services associated with the visits, as applicable. We then multiplied the visit counts reported by PIC patients by these mean costs.

For psychotropic medications, we matched patient-reported data of medication names, daily dosages, and months of use to average costs for that combination from the Ingenix data, pooling data on generic and brand names for the same medication according to their relative proportion in the Ingenix data and summing all medications used (for reference, 20 mg of fluoxetine cost $2.20 per pill, on average).

Indirect costs of treatment include patient time costs for obtaining health care.25 We assumed an average time of 30 and 45 minutes for outpatient medical and mental health visits, respectively, and added average travel and waiting times reported by patients at baseline. In addition, we assumed 3 hours for emergency department visits and 1.5 hours to fill prescriptions in a month of use. We priced patients’ time using reported hourly wage at baseline and sex-specific mean wage for those not working at baseline (this may slightly overstate the value of time for nonworking patients).

Outcomes
Quality-Adjusted Life-Years. To measure quality-adjusted life-years (QALYs), we calculated a health utility index from the Short-Form, 12-Item Health Survey (SF-12) items collapsed into 6 health states that had been identified through cluster analysis of SF-12 physical and mental component scores.20,22 Utility weights from this index were derived from a convenience sample of primary care patients with symptoms of depression using a standard gamble approach.26 QALY weights were calculated for each 6-month follow-up time period, and we analyzed patterns over time. We call this measure “QALY-SF.”

In addition, following an approach developed by Lave et al.,28 we developed a measure of depression-burden days and assigned utility scores from the literature to estimate QALYs.28,29 We call this measure “QALY-DB.” Specifically, for each survey from baseline through 24 months, we developed a count of positive scores on the following 3 measures: probable major depressive disorder, based on a repeat of the baseline screener19; significant depressive symptoms, based on a modified Center for Epidemiologic Studies Depression Studies (CES-D) scale19,30; and poor mental health-related quality of life (HRQOL), based on being more than 1 SD below the population mean on the mental health subscale of the SF-12.3

We averaged the count for the beginning and end of each 6-month follow-up period and multiplied by 182 to estimate depression-burden days during the period. We summed across periods to get the 24-month total. We then used findings from the literature that a year of depression is associated with losses of 0.2 to 0.4 QALYs to convert the intervention effect on depression-burden days into the QALY-DB estimates.9,31-34

Employment. We created a measure of days worked in each 6-month follow-up by taking the average of employment status at the start and end of each period and multiplying by 116 (the number of workdays in 6 months). We
summed across periods to calculate the 24-month total. We also examined days missed from work due to illness, which patients reported for the 4 weeks preceding each follow-up survey.

Covariates
All multivariate models controlled for baseline measures of patient age, sex, marital status, education, rank in the distribution of household wealth, employment status, ethnicity, medical comorbidity, depressive disorder status, the SF-12 aggregate HRQOL measures, presence of comorbid anxiety disorder, and practice randomization block.

Data Analysis
To estimate the effects of practice-initiated QI on patients, we conducted patient-level intent-to-treat analyses, controlling for baseline patient differences that could remain after group-level randomization. We examined intervention effects on health care costs using 2-part models, due to the skewed distribution of costs. The first is the probability of positive costs, using logistic regression. The second is the log of costs given any, using ordinary least squares. We used the smearing estimate for retransformation, applying separate factors for each intervention group to ensure consistent estimates. We did not adjust cost models for clustering by clinic because we know of no existing software to do so for 2-part models. We expected the interventions to increase health care costs, relative to usual care; not accounting for clustering is thus conservative from a policy perspective, since not adjusting for clustering is likely to overstate the statistical significance of cost differences.

For the QALY-SF measure, we specified 3-level (repeated measurements nested within patients, and patients nested within clinics) mixed effects linear time-trend regression models, controlling for the baseline utility value in addition to the covariates listed above (except HRQOL). We calculated the area under the curve to derive values for each wave. For outcomes, we averaged predictions from 5 randomly imputed data sets and adjusted SEs for uncertainty due to imputation.

Because many tests are in the same direction as hypothesized, a formal Bonferroni correction for multiple statistical comparisons is too conservative, so we report actual P values and interpret results with multiple comparisons in mind.

RESULTS
At baseline, the percentage of patients with a college education was lower for usual care (15.0%) than QI-meds (22.9%, P = .004) or QI-therapy (21.5%, P = .01). Usual care patients were more likely to have current symptoms with lifetime disorder (26.1%) compared with QI-meds (18.5%, P = .02) and QI-therapy (19.4%, P = .03) vs current disorder or symptoms but no lifetime disorder. QI-therapy patients were slightly older (by 3 years on average) (P = .02). Intervention and control patients did not differ with respect to the other baseline covariates listed above at P = .05.

The Table reports average per patient costs and outcomes over 24 months (including patient time costs, but not inpatient care and nonpsychotropic medications). Average total costs for usual care patients were estimated to be $3835, increasing by $419 (11%) among QI-meds participants and by $485 (13%) among QI-therapy participants. Neither intervention effect on total costs is statistically significant. Patient time costs represented 22% of the average total cost under usual care. Increases in time costs represented 3% of the incremental increase due to QI-meds and 25% of the incremental increase due to QI-therapy. The intervention effects on patient time costs were also statistically nonsignificant (details available from authors on request).

For the QALY-SF measure, the incremental increase due to QI-meds was 0.0115 QALYs over 24 months (P = .15), while the increase due to QI-therapy was 0.0226 (P = .006). Combining these point estimates with our point estimates of the incremental intervention costs yields an estimated cost per QALY of $36,467 for QI-meds and $21,478 for QI-therapy.

For the QALY-DB measure, we assumed that depression reduces the value of a life-year by 0.2 to 0.4 QALYs. Compared with usual care, QI-meds reduced the number of depression-burden days by 25 (P = .19), or 0.0137 to 0.0274 QALYs. QI-therapy yielded 47 fewer depression-burden days over 24 months (P = .01), or 0.0258 to 0.0515 QALYs. These point estimates yield a cost-per-QALY range of $15,331 to $30,663 for QI-meds and $9,478 to $18,953 for QI-therapy.

As shown in the Table, participants from QI-meds clinics had 17.9 more employed days relative to usual care over 24 months (P = .07), on average, and QI-therapy participants had 20.9
more employed days ($P = .03$). Intervention and usual care patients who were working did not differ substantively or statistically with respect to sick days at any follow-up period. For instance, intervention patients were significantly more likely to be working at the 12-month follow-up survey (65.7% in QI vs 60.8% in usual care; 95% confidence interval [CI] for difference: 0.01-0.09; $P = .03$). Among employed patients, however, the number of reported sick days in the previous 4 weeks was virtually identical (1.2 days in QI vs 1.1 in usual care; 95% CI for difference: −0.5 to 0.6; $P = .81$). Results for other follow-up periods were similar.

**COMMENT**

We found that practice-initiated, locally implemented programs that encourage guideline-concordant care for depression can substantially reduce the individual suffering and economic consequences of depression. The point estimates for incremental costs per QALY relative to usual care were within the range of many accepted medical interventions and substantially below the estimated value of a year of life. Our data suggested that QI-therapy may have a better overall value in terms of cost per QALY than QI-meds, suggesting that there may be particular value to improving access to structured psychotherapy such as cognitive behavioral therapy for depressed primary care patients.

We found significant intervention effects on patients’ labor supply, on the order of 1 additional month of employment over 2 years. In addition to its importance to patients, this result could suggest broader economic benefits of the intervention to families and purchasers—benefits that may not be fully captured in standard measures of QALYs, which are based on patients’ HRQOL. If so, the true societal cost-effectiveness of the interventions may be more favorable than what we report.

This study has important limitations. For instance, we studied 6 practice networks, although they were chosen to be diverse. We relied on patient self-report for most measures, which would bias intent-to-treat analyses if the interventions affected patient reports. We had a relatively low enrollment rate, which we partially account for by weighting back to the eligible population. Despite a large sample size relative to most clinical trials, our cost estimates lacked precision.

Our findings suggest that practice-initiated interventions to improve quality of care for depression can substantially increase patients’ and societal welfare, even when implemented locally and under flexible, naturalistic practice conditions that support patient and clinician treatment choices. These interventions increase costs for clinicians and insurers, suggesting that their widespread adoption may require increases in consumer demand or public policy initiatives that provide incentives for implementing them. But the gains observed in such an applied and real-world context suggest that improved medical care has much to offer depressed patients and their families and communities if we can create the conditions necessary to put such programs in place.

**Author Contributions:** Study concept and design: Schoenbaum, Unützer, Sherbourne, Rubenstein, Miranda, Carney, Wells. Acquisition of data: Schoenbaum, Unützer, Miranda, Meredith, Carney, Wells. Analysis and interpretation of data: Schoenbaum, Unützer, Sherbourne, Duan, Rubenstein, Wells. Drafting of the manuscript: Schoenbaum, Unützer, Sherbourne, Miranda, Carney, Wells. Critical revision of the manuscript for important intellectual content: Schoenbaum, Unützer, Sherbourne, Duan, Rubenstein, Meredith, Wells. Statistical expertise: Schoenbaum, Duan. Obtained funding: Sherbourne, Meredith, Wells. Administrative, technical, or material support: Unützer, Sherbourne, Carney, Wells. Study supervision: Unützer, Sherbourne, Miranda, Wells.

**Funding/Support:** This work was funded by Agency for Healthcare Research and Quality grant R01-HS08349; grants SR01 MH57992 and P50 MH54623 and Research Scientist Award MH01170-05 from the National Institute of Mental Health; and grant 96-42901A-HE from The John D. and Catherine T. MacArthur Foundation.

**Acknowledgment:** We are grateful to Daniel McCaffrey, PhD, Robert Bell, PhD, and Tom Belin, PhD, for statistical advice; Roland Sturm, PhD, for development of labor force participation measures; Wayne Katon, MD, and Maga Jackson-Triche, MD, MSHS; for assistance with the design and implementation of the interventions; Bernadette Benjamin, MS, and Martha Manewicz, MS, for their meticulous programming support; and Susan Ettner, Katherine Harris, Emmett Keeler, and Willard Manning for helpful comments on the manuscript. We are grateful to the managed care organizations participating in this study, who provided access to their expertise and patients, implemented interventions, and provided in-kind resources: Allina Medical Group (Minn), Patuxent Medi-

**Table.** Average Costs and Outcomes per Patient of Quality Improvement (QI) Interventions Relative to Usual Care Over 24 Months*

<table>
<thead>
<tr>
<th>Outcome Measure</th>
<th>Total Under Usual Care</th>
<th>Incremental Effect of QI-Meds</th>
<th>Incremental Effect of QI-Therapy</th>
<th>Incremental Effect of Pooled QI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Health care costs (including patient time), $</td>
<td>3835</td>
<td>419</td>
<td>485</td>
<td>454</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>3282 to 4389</td>
<td>−467 to 1306</td>
<td>−393 to 1363</td>
<td>−305 to 1214</td>
</tr>
<tr>
<td>$P$ (value)</td>
<td>$\ldots$</td>
<td>0.928 (35)</td>
<td>1.082 (28)</td>
<td>1.173 (24)</td>
</tr>
<tr>
<td>Quality-adjusted life-years (QALY-5F)</td>
<td>1.6624</td>
<td>0.0115</td>
<td>0.0226</td>
<td>0.0173</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>1.628 to 1.697</td>
<td>−0.004 to 0.027</td>
<td>0.008 to 0.038</td>
<td>0.004 to 0.030</td>
</tr>
<tr>
<td>$P$ (value)</td>
<td>$\ldots$</td>
<td>1.49 (15)</td>
<td>2.94 (006)</td>
<td>2.61 (007)</td>
</tr>
<tr>
<td>Days of depression burden</td>
<td>419.9</td>
<td>−25.0</td>
<td>−46.7</td>
<td>−36.5</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>398.9 to 441.0</td>
<td>−63.1 to 13.2</td>
<td>−83.1 to −10.3</td>
<td>−68.9 to −4.1</td>
</tr>
<tr>
<td>$P$ (value)</td>
<td>$\ldots$</td>
<td>−1.33 (19)</td>
<td>−2.61 (01)</td>
<td>−2.29 (03)</td>
</tr>
<tr>
<td>Days of employment</td>
<td>279.2</td>
<td>17.9</td>
<td>20.9</td>
<td>19.5</td>
</tr>
<tr>
<td>95% confidence interval</td>
<td>270.2 to 288.1</td>
<td>−1.6 to 37.4</td>
<td>2.4 to 39.3</td>
<td>3.5 to 35.5</td>
</tr>
<tr>
<td>$P$ (value)</td>
<td>$\ldots$</td>
<td>1.87 (07)</td>
<td>2.30 (03)</td>
<td>2.48 (02)</td>
</tr>
</tbody>
</table>

*All tests compare outcomes in respective intervention arm with outcomes under usual care (2-sided test, $d f = \infty$). Ellipses indicate data not applicable.
REFERENCES


7. Unutzer J, Patrick D, Diehr P, et al. Collaborative Agreement to Test Depression Practice Guidelines (Lisa Rubenstein, MD, MSHS, Kathryn Rost, PhD, and Daniel Ford, MD, MPH, principal investigators), and investigators from that project helped design the quality of care and health measures.


18. JAMA. September 19, 2001—Vol 286, No. 11 (Reprinted) ©2001 American Medical Association. All rights reserved.