



## HEALTH

THE ARTS  
CHILD POLICY  
CIVIL JUSTICE  
EDUCATION  
ENERGY AND ENVIRONMENT  
HEALTH AND HEALTH CARE  
INTERNATIONAL AFFAIRS  
NATIONAL SECURITY  
POPULATION AND AGING  
PUBLIC SAFETY  
SCIENCE AND TECHNOLOGY  
SUBSTANCE ABUSE  
TERRORISM AND  
HOMELAND SECURITY  
TRANSPORTATION AND  
INFRASTRUCTURE  
WORKFORCE AND WORKPLACE

This PDF document was made available from [www.rand.org](http://www.rand.org) as a public service of the RAND Corporation.

[Jump down to document](#) ▼

The RAND Corporation is a nonprofit research organization providing objective analysis and effective solutions that address the challenges facing the public and private sectors around the world.

### Support RAND

[Browse Books & Publications](#)

[Make a charitable contribution](#)

### For More Information

Visit RAND at [www.rand.org](http://www.rand.org)

Explore [RAND Health](#)

View [document details](#)

This product is part of the RAND Corporation reprint series. RAND reprints present previously published journal articles, book chapters, and reports with the permission of the publisher. RAND reprints have been formally reviewed in accordance with the publisher's editorial policy, and are compliant with RAND's rigorous quality assurance standards for quality and objectivity.

# Cost-Effectiveness of Quality Improvement Programs for Patients With Subthreshold Depression or Depressive Disorder

Kenneth B. Wells, M.D., M.P.H.

Michael Schoenbaum, Ph.D.

Naihua Duan, Ph.D.

Jeanne Miranda, Ph.D.

Lingqi Tang, Ph.D.

Cathy Sherbourne, Ph.D.

**Objective:** This study explored the cost-effectiveness of quality-improvement interventions for depression in primary care, relative to usual care, among patients with subthreshold depression or depressive disorder.

**Methods:** A total of 746 primary care patients in managed care organizations with 12-month depressive disorder and 502 with current depressive symptoms but no disorder (subthreshold depression) participated in a group-level randomized controlled trial initiated between June 1996 and March 1997. Matched clinics were randomly assigned to enhanced usual care or one of two quality improvement interventions that provided education to manage depression over time and resources to facilitate access to medication management or psychotherapy for six to 12 months.

**Results:** The cost-effectiveness ratio for the pooled intervention groups versus usual care was \$2,028 for patients with subthreshold depression (95% confidence interval [CI]=-\$17,225 to \$21,282) and \$53,716 for those with depressive disorder (CI=\$14,194 to \$93,238), by using a measure of quality-adjusted life years (QALY) based on the 12-Item Short Form Health Survey. Similar results were obtained when alternative QALY measures were used. **Conclusions:** Although precision was limited, even the upper limit of the 95% CIs suggests that such interventions are as cost-effective for patients with subthreshold depression as are many widely used medical therapies. Despite lack of evidence for efficacy of treatments for subthreshold depression, disease management programs that support clinical care decisions over time for patients with subthreshold depression or depressive disorder can yield cost-effectiveness ratios comparable to those of widely adopted medical therapies. Achieving greater certainty about average cost-effectiveness would require a much larger study. (*Psychiatric Services* 58:1269–1278, 2007)

Depressive disorders are an important focus of practice-based efforts to improve quality of primary care (1–11), owing to their high prevalence and impact on disability and evidence of low to moderate rates of use of evidence-based treatments in primary care (12–16). Short-term quality improvement programs for primary care patients with depression can improve clinical and functional outcomes for six to 28 months and for up to five years (1–11,17). Most such programs have focused on patients with major depressive and dysthymic disorders. Although subthreshold depression, or depressive symptoms below the threshold for depressive disorder, is common and associated with morbidity, the efficacy of treatments for this condition is uncertain (18,19).

Patients with recent depressive disorder or subthreshold depression were included in Partners in Care (PIC), a multisite group-level randomized trial of two quality improvement programs for depression in primary care versus enhanced usual care (10). In that study two quality improvement interventions were implemented that provided information and resources to facilitate decisions about treatment over time and that facilitated use of evidence-based treatments for depression, as warranted, for six to 12 months. Analyses pooling data on intervention groups suggested similar intervention effects

---

*Dr. Wells and Dr. Sherbourne are affiliated with the Health Program, RAND Corporation, 1776 Main St., P.O. Box 2138, Santa Monica, CA 90407-2138 (e-mail: kwells@ucla.edu). At the time the study was conducted, Dr. Schoenbaum was principally affiliated with the RAND Corporation, Washington, D.C. He is now affiliated with the National Institute of Mental Health, Bethesda, Maryland. Dr. Duan is with the Division of Biostatistics, New York State Psychiatric Institute, New York, although the work was principally conducted when he was with the Semel Institute Health Services Research Center, University of California, Los Angeles. Dr. Miranda and Dr. Tang are with the Health Services Research Center, Semel Institute for Neuroscience, University of California, Los Angeles, with which Dr. Wells is also affiliated.*

on health at one-year follow-up among patients with either depressive disorder or subthreshold depression at baseline (10). In a time-trend analysis for functioning and well-being outcomes over the course of the first two follow-up years, we found no significant differences in intervention effects relative to usual care, by initial depressive disorder status (20).

Although we demonstrated that societal cost-effectiveness of the PIC interventions was comparable with that of accepted medical interventions over two years of follow-up, we did not examine cost-effectiveness separately for persons who initially had a likely depressive disorder or subthreshold symptoms. Such data would inform programmatic decisions about including or excluding patients with subthreshold depression in intervention dissemination efforts. Although the data from PIC that support cost-effectiveness analyses (that is, the first two follow-up years) are from the late 1990s, the study remains a unique opportunity to explore cost-effectiveness of quality improvement interventions for patients with subthreshold depression and for those with depressive disorders, because other studies have not had large samples of both groups.

Studies have shown that some quality improvement programs for depression in primary care increase health care costs over one to two years but can achieve cost-effectiveness ratios relative to enhanced usual care that fall within the range of accepted medical therapies, especially for sicker patients or high utilizers of medical services (21,22). These programs may also offer cost saving for subgroups, such as elderly persons with diabetes and depression (23). Some suggest that routing primary care patients with subthreshold or minor depression immediately to treatments may be neither effective nor cost-effective (24). Given PIC findings of similar outcomes by initial disorder status over two follow-up years, however, we thought that the PIC intervention approach that promotes monitoring the course of illness over time and supports care decisions based on need could be cost-effective for patients with initial sub-

threshold depression. We have not previously reported cost-effectiveness ratios or costs or quality-adjusted life year (QALY) measures, which are integral to cost-effectiveness analysis, by intervention and disorder status.

We previously reported outcomes of the PIC interventions at the five-year follow-up for patients by initial disorder status, but we do not have the necessary longitudinal data on outcomes or costs between two and five years to support cost-effectiveness analyses over that full interval. At five years when we compared the intervention groups to patients who received usual care, we found improved clinical outcomes among patients with initial subthreshold depression who were in the intervention involving therapy quality improvement, which provided supplemental resources to facilitate access to evidence-based psychotherapy, and among patients with initial depressive disorder who were in the intervention involving medication quality improvement, which provided supplemental resources to facilitate access to education and resources for medication management. Further, each intervention reduced the use of mental health services at five years among patients with subthreshold depression (25). These findings stimulated our interest in exploring whether the interventions were cost-effective for patients with subthreshold depression within the first two years.

## Methods

### *Experimental design and implementation*

The data are from PIC, a group-level, randomized controlled trial of practice-implemented quality improvement programs for depression (10,26,27). Participating in the trial were six managed care organizations, 46 of 48 eligible primary care practices, and 181 of 183 eligible primary care clinicians. Within organizations, practices were matched into blocks of three clusters on the basis of specialty mix, patient socioeconomic and demographic factors, and presence of mental health specialists on site. Practice clusters were randomized within blocks to enhanced usual care (mailing of written practice guide-

lines to medical directors) or to the medication quality improvement intervention or to the therapy quality improvement intervention. Within the medication quality improvement arm only, half of the patients were randomly assigned to receive an additional six months of contacts from a nurse.

Study staff screened 27,332 consecutive patients between June 1996 and March 1997. Patients were eligible if they intended to use the practice for 12 months and screened positive for current depressive symptoms plus probable major depressive or dysthymic disorder in the past year according to lead-in items of the World Health Organization's 12-month Composite International Diagnostic Interview (CIDI) (28). Patients were ineligible if they were younger than 18 years, if they were not fluent in English or Spanish, or if their insurance did not cover either their practice providers or the services encouraged by the interventions. The study was approved by the institutional review boards of RAND and the practices.

Among patients completing the screener, 3,918 were potentially eligible for the study, but many left the clinic before insurance status could be checked; 2,417 were available for confirming insurance, and 241 (10%) did not have insurance that could guarantee access to the treatments facilitated by the interventions, and thus they were ineligible. Of those who completed the informed consent forms, 1,356 of 1,485 enrolled in the study: 443 in the usual care group, 424 in the medication quality improvement intervention, and 489 in the therapy quality improvement intervention.

### *Interventions*

The interventions are described in detail elsewhere (29); all intervention materials are posted at [www.rand.org/health/projects/pic/order.html](http://www.rand.org/health/projects/pic/order.html).

We estimated each organization's participation costs and provided half that amount (\$35,000–\$70,000 per organization). The interventions provided practices with training and resources to initiate and monitor quality improvement programs, adapted to local goals and resources. Patients

and clinicians retained choice of treatment and use of intervention materials; the randomization was to resources for improved care, not mandated treatment

For both interventions, local teams (a primary care practitioner, practice nurse, practice administrator, and a psychiatrist or psychologist) were trained in a two-day workshop to educate primary care clinicians and to supervise staff and conduct team oversight. Practice nurses were trained to help in patient assessment, education, and activation for treatment. Practice teams were given patient education pamphlets, videotapes, tracking forms, clinician manuals, lecture slides, and pocket reminder cards. The materials described guideline-concordant care for depression—for example, presented psychotherapy and antidepressant medication as equally effective for most patients with the disorder, encouraged attention to patient preferences, and advised adjusting treatment plans to patient need and preferences (30,31).

In the medication quality improvement intervention, nurse specialists were trained to support medication adherence through monthly visits or telephone contacts for six or 12 months. In the therapy quality improvement intervention, practice therapists were trained to provide individual and group cognitive-behavioral therapy (32,33), which was available to participants for the cost of the primary care copayment (about \$5–\$10) for six months after enrollment. All patients could receive other therapy at the cost of the usual copayment (about \$20–\$35). In all conditions, patients could receive medications, therapy, both, or neither. For example, in the first and second six months of the study, 40% and 35% of patients, respectively, in the therapy quality improvement intervention received an antidepressant and 38% and 34%, respectively, received at least four psychotherapy sessions. Fifty-two percent and 43%, respectively, of patients in the medication quality improvement intervention received an antidepressant; 30% and 29%, respectively, received at least four psychotherapy sessions (27,34).

The interventions were designed to encourage providers and care managers to review the patient's initial clinical status and consider education, treatment, and management strategies appropriate to the patient's clinical status and course of illness over time. Intervention practices were provided with lists of participating patients, indicating which met CIDI criteria for 12-month depressive disorder. Providers were encouraged to watch for early signs of depressive disorder among patients who initially did not have the disorder and to initiate treatment as needed. In the therapy quality improvement intervention, a four-session form of cognitive-behavioral therapy was available for patients with subthreshold depression. The provider training materials noted that there was little evidence for effectiveness of antidepressant medication for patients with subthreshold depression, particularly in the absence of lifetime disorder (30,31).

#### *Data collection*

At baseline patients were asked to complete the Patient Screening Questionnaire, which gathered information on demographic characteristics and health status; a telephone interview on economic variables; and a mailed survey—the Patient Assessment Questionnaire (PAQ)—on depression and health outcomes. We mailed follow-up PAQ surveys at six, 12, 18, and 24 months. A telephone survey was also conducted at 24 months. Outcomes data at 57 months are reported elsewhere (17,25). Data completion rates of having either mail or telephone surveys relative to all initial enrollees (N=1,356) were 95% and 85%, respectively, for the baseline and 24-month surveys.

#### *Outcome measures*

*Quality-adjusted life years based on SF-12.* A health utility index from the 12-Item Short-Form Health Survey (SF-12) was developed specifically for the overall study to measure quality-adjusted life years (QALYs) (35, 36). Six health states were identified through cluster analyses of SF-12 physical and mental component scores. Utility weights from this index

were derived from a convenience sample of primary care patients with symptoms of depression by using a standard gamble approach. QALY weights were calculated for each six-month follow-up time period, and patterns were analyzed over time. This measure is called the QALY-SF.

*Days of depression burden.* Following an approach developed by Lave and colleagues (21), we developed a measure of depression-burden days and assigned utility scores from the literature to estimate QALYs. For each survey from baseline through 24 months, we developed a count of positive scores (possible scores of .00, .33, .67, 1.00) based on the following three dichotomous measures: probable major depressive disorder, based on a repeat of the baseline screener (10); significant depressive symptoms, based on a modified Center for Epidemiologic Studies Depression Scale cutoff score (CES-D) (10,20, 37); and poor mental health-related quality of life, based on being more than one standard deviation below the population mean on the mental health subscale of the SF-12 (35). We averaged the count for the beginning and end of each six-month follow-up period and multiplied by 182 to estimate number of days fully burdened during six months. We summed across periods to get the 24-month total. We used findings from the literature stating that a year of depression is associated with losses of .2 to .4 QALYs to convert the intervention effect on depression-burden days into the QALY-DB estimates (34,36).

*Employment.* A measure of days worked in each six-month follow-up was developed by taking the average of employment status (scored as 1 if employed and as 0 otherwise) at the start and end of each period and multiplying by 116 (the number of workdays in six months). Total days worked in 24 months were obtained by summing across the periods. Days missed from work as a result of illness, which patients reported for the four weeks preceding each follow-up survey, were also examined.

*Intervention costs.* We assigned costs to intervention activities (screening, intervention materials, nurse assessments, and supervision of nurses

and therapists) per enrolled patient on the basis of data from practices about the average costs of clinic staff (excluding research costs). Follow-up visits to intervention staff—for example, for psychotherapy in the therapy quality improvement arm—were included in outpatient visits, described in the next section.

**Health care costs.** Costs were assigned to patient-reported counts of emergency department visits, medical and mental health visits, psychotropic medications used, and inpatient days during each follow-up period. Patient report was selected because of limitations in the available claims and encounter data. In addition, the number of outpatient visits was higher for patient surveys than for claims data over the first six months, probably because of out-of-practice use or incomplete claims data. Inpatient costs were excluded from our main analyses because the interventions were not expected to change these costs and because of limited sample size.

Average costs in 1998 dollars were assigned to each component of patient-reported health care use by using a national database of about 1.8 million privately insured individuals (provided by Ingenix, a benefits consulting firm in New Haven, Connecticut). The Ingenix data included information on provider reimbursements, which were used as a proxy for health care costs. By using these techniques, the mean costs were \$46 for each outpatient medical visit, \$96 for each mental health visit, and \$450 for each emergency department visit. These costs include facility charges, professional fees, and ancillary services associated with the visits, as applicable. The visit counts reported by PIC patients were multiplied by these mean costs to estimate the total visit costs.

For psychotropic medications, patient-reported data of medication names, daily dosages, and months of use were matched in the Ingenix data to obtain average costs for that combination. 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 to obtain costs (for

reference, 20 mg of fluoxetine costs \$2.20 per pill, on average).

Indirect costs of treatment include patient time costs for obtaining health care (38). An average time for outpatient medical (30 minutes) and mental health (45 minutes) visits was assumed. Travel and waiting times were reported by patients at baseline. In addition, we assumed three hours for emergency department visits and 1.5 hours to fill prescriptions in a month of use. Patients' time was priced by using reported hourly wage at baseline and gender-specific mean wage for those not working at baseline.

We calculated two total cost measures, one including and one excluding inpatient costs—that is, we added costs for outpatient services for mental and physical health care, emergency room services, medications, patient time, and intervention services. We did not expect the intervention to impact inpatient costs, which are highly variable. We consider the measure including inpatient costs as a sensitivity analysis to the main focus on outpatient costs.

#### **Measures: independent variables**

**Intervention status.** We used indicators for the medication quality improvement intervention and the therapy quality improvement intervention, each compared with enhanced usual care. And in separate analyses we used an indicator of the pooled intervention groups compared with enhanced usual care.

**Baseline disorder status.** We used data from the screener and baseline CIDI (26) to categorize patients as having either recent depression (that is, 12-month major depressive or dysthymic disorder plus having 30-day depressive symptoms) versus having “subthreshold depression,” defined as not having a recent disorder but having 30-day symptoms plus having a history of either two weeks of depressed mood or loss of interest in usual activities in the last 12 months or depressed mood or loss of interest in usual activities most days over the past two years. Thirty-day symptoms are defined as five or more days of depressed mood or loss of interest in usual activities in the past 30 days.

Among persons with subthreshold depression, we assessed probable lifetime disorder using two items derived from the lifetime CIDI (28).

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

#### **Data analysis**

We extended the methods used in our previous PIC analysis of two-year cost effectiveness (34) to estimate the intervention effect on each health and cost outcome separately for patients with 12-month depressive disorder or subthreshold depression at baseline. To do so we conducted the analyses on the overall study sample and included intervention status, baseline disorder status, and their interactions in the model. We tested whether the intervention effect differed by baseline disorder status, by testing for the interaction between intervention status and baseline disorder status, but we had poor precision for such tests and focused instead on the separate estimates within each group. Sample sizes were relatively small for cost comparisons, and we focused on the pooled intervention groups compared with usual care. We considered analyses of the intervention effects on costs for each disorder group, relative to usual care, as exploratory.

We examined baseline imbalance in patient characteristics for the overall sample and by baseline disorder status. Baseline imbalance for the overall sample was controlled for by including the main effects for the covariates in the models. Differential baseline imbalance by disorder status was controlled for by including the interaction between disorder status and the covariates manifesting differential imbalance.

We examined intervention effects on total health care costs by using a one-part model for log (cost + 1) instead of the widely used two-part model, because there were only four patients with zero total cost.

In reaching the decision to take the logarithmic transformation for the one-part model, we conducted residual analyses both with and without the logarithmic transformation. Without the logarithmic transformation, the residuals are highly skewed for total costs (skewness=4.09). After the logarithmic transformation, the skewness was reduced to -1.28. Tukey's one-degree-of-freedom test was insignificant, suggesting adequate model fit (39).

We used a smearing estimate for retransformation, applying separate factors for each intervention group to ensure consistent estimates (40,41). We adjusted the standard errors in the fitted models for clustering by clinic using the bias-reduced linearization method to overcome bias problems in the usual linearization method (also known as the Huber-White method or robust standard errors) when the number of clusters is small (42).

For the QALY-SF measure, we specified three-level (repeated measurements nested within patients and patients nested within clinics) mixed-effects linear time-trend regression models. We calculated the area under the curve for the trajectory of the intervention effect on QALY to derive the aggregate intervention effect over 24 months. For days of depression burden and employment we specified two-level (patients nested within clinics) mixed-effects linear regression models to account for patient clustering at the practice level. For these outcomes, we examined the 24-month value directly.

Significance of comparisons across intervention groups for each health outcome is based on the regression coefficients. We illustrated average intervention effects relative to usual care, adjusted for patient characteristics, using standardized predictions. Specifically, we used the regression coefficients and each individual's actual values for all covariates other than intervention status to derive three predicted outcomes, one for each intervention condition (usual care or either intervention), assuming that the patient had been assigned to that intervention condition. We then calculated the mean prediction under

each intervention condition, averaged across all patients in the study.

We analyzed data for patients completing at least one follow-up (92% of the enrolled sample; N=1,248). The data were weighted for the probability of study enrollment at screening and follow-up response. Multiple imputation (43,44) was used to deal with item nonresponse, using an extended hot deck technique that modifies the predictive mean matching method. The analysis results were summarized across the five imputed data sets by multiple imputation inference methods: the point estimates were averaged across the five imputed data sets; the standard errors within the imputed data sets were combined with the variation of the point estimates across the five imputed data sets to form standard errors that reflect both within-imputation variability and between-imputation variability (44).

We considered the analyses as exploratory, given limited precision for cost comparisons. We developed cost-effectiveness ratios for the pooled interventions relative to usual care to maximize precision. We calculated the ratio of incremental costs to the incremental outcome (QALY-SF and QALY-DB separately) over two years on the basis of the regression models described above. To develop the 95% confidence intervals (CIs), we used Taylor's series approximation (or delta method) (45,46) for the variance of the ratio estimator, where the means, variances, and covariance between the numerator and denominator were estimated from the bootstrap method for a clustered randomized trial with 10,000 replicates (45,46), which allowed us to take into account the group-level randomized design and multiple imputation. We tried several other methods (45): Fieller's method (47), nonparametric bootstrap (percentile method), bootstrap bias correction (48), and O'Brien and colleagues' confidence box (49).

In the bootstrap approach, the problem of undefined intervals arose in the subgroup analyses for subthreshold depression at baseline because the bootstrap replicates of ratio estimates were observed in all four quadrants of cost-effectiveness plane

(49). For example, in estimating the ratio estimate of incremental cost to the incremental QALYs, we found 59.84% of bootstrap replicates falling in the first quadrant in the cost-effectiveness plane (more effective, more costly), 40.09% in the fourth quadrant (more effective, less costly), leaving .03% in the second and .04% in the third quadrants. Therefore, it is problematic to interpret the percentile bootstrap confidence interval. Similar problems were found in the use of other methods. Therefore, we chose the Taylor series approximation for the variance of the ratio estimator, using the bootstrap estimates of the means, variances, and covariance between the numerator and denominator. We note, however, that we have a higher coefficient of variation in health-related quality of life than that recommended for this method. The recommended maximum is either less than .05 (50) or less than .10 for large samples (51). In PIC the coefficient of variation for QALY-SF was .29 for the pooled group of persons in the intervention groups, both for those with subthreshold depression and those with 12-month depressive disorder.

Given the exploratory nature of this study, we supplemented the conventional significance threshold of  $\alpha=5\%$  with an exploratory significance threshold of  $\alpha=10\%$ , referring to results based on this exploratory threshold as "weak evidence" in the Results section. Among the outcomes, we focus primarily on the main measure of QALYs and health care costs (excluding inpatient costs). We report actual p values and interpret results with multiple comparisons in mind (52).

## Results

Tables 1 and 2 provide baseline characteristics of the sample by intervention status, stratified by initial disorder status. There were no significant differences in baseline characteristics by intervention status among patients with depressive disorder. Among patients with subthreshold depression, the three intervention groups differed significantly in gender distribution, mental health-related quality of life, age distribution, and probable lifetime depressive dis-

**Table 1**Baseline characteristics of patients (N=746) with recent depressive disorder in Partners in Care<sup>a</sup>

Characteristic	Enhanced usual care (N=214)		Quality improvement intervention					
			Medication (N=249)		Therapy (N=283)		Pooled (N=532)	
	N	%	N	%	N	%	N	%
Married	110	51	131	53	151	53	282	53
Female	152	71	191	77	216	76	407	77
Education								
Less than high school	43	20	38	15	51	18	89	17
High school graduate	72	34	69	28	78	28	147	28
Some college	66	31	81	33	94	33	175	33
College graduate	33	15	61	24	59	21	120	23
Employed	138	65	158	63	181	64	339	64
Ethnicity								
White	129	60	165	66	159	56	324	61
Latino	58	27	55	22	94	33	149	28
Black	16	7	13	5	14	5	27	5
Other	11	5	16	7	16	6	32	6
12-month depressive disorder status								
Current major depressive and dysthymic disorders	44	20	40	16	71	25	111	21
Current major depressive disorder only	163	76	190	77	198	70	388	73
Current dysthymic disorder only	8	4	19	8	14	5	33	6
Current anxiety	119	55	134	54	156	55	290	55
Chronic health condition count								
0	44	21	58	23	61	22	119	22
1	54	25	60	24	74	26	134	25
2	39	18	44	18	60	21	104	20
3 or more	77	36	88	35	88	31	176	33
Any antidepressant medication or counseling for depression in the past 6 months	109	51	127	57	163	59	290	58
Age (mean±SD)	43±14		43±13		44±15		44±14	
Health-related quality of life (mean±SD score)								
MCS-12 <sup>b</sup>	33±10		32±10		33±10		32±10	
PCS-12 <sup>c</sup>	44±12		45±11		44±11		44±11	

<sup>a</sup> Ns across rows may not total to the whole sample because of either rounding of weighted data or missing data for a given variable. The quality improvement groups and the usual care group did not differ from one another on any characteristic.

<sup>b</sup> Mental health composite score from the 12-item Short-Form Health Survey. Possible scores range from 0 to 100, with higher scores indicating better mental health. A score of 50 is equivalent to the general population mean.

<sup>c</sup> Physical health composite score from the 12-item Short-Form Health Survey. Possible scores range from 0 to 100, with higher scores indicating better physical health. A score of 50 is equivalent to the general population mean.

order status. Therefore, we included in the main models interactions between baseline depressive disorder status and age, gender, mental health–related quality of life, and probable lifetime disorder status. Conclusions were similar with and without these interactions, however.

Among patients with 12-month depressive disorder (Table 3), the pooled interventions reduced depression burden days by 46 days on average ( $p=.02$ ) and increased days of employment by 23 days on average ( $p=.01$ ). The corresponding effects for patients with subthreshold depression (Table 3) were somewhat smaller—that is, 31 fewer burden days, which was not statistically sig-

nificant, and 15 more employed days, for which there was weak evidence ( $p=.07$ ).

There was weak evidence ( $p\leq.10$ ) for gains in QALYs both for patients with 12-month depressive disorder (average gain=.017,  $p=.10$ ) and for patients with subthreshold depression (average gain=.018,  $p=.06$ ). The intervention effects on QALYs, days of depression burden, and days of employment did not differ significantly by disorder status—that is, the interaction terms for intervention  $\times$  disorder status were not significant for these outcomes.

There was weak evidence that the pooled intervention groups had higher total health care costs (ex-

cluding inpatient costs) than the usual care group (difference=\$912,  $p=.1$ ) among persons with 12-month depressive disorder. The separate intervention results provided weak evidence of higher total health care cost for the medication quality improvement intervention ( $p=.09$ ) but not for the therapy quality improvement intervention.

We calculated that the costs of the intervention per se—as distinct from intervention effects on use of services and medication—were \$86 per patient in the medication quality improvement intervention and \$79 per patient in the therapy quality improvement intervention. These did not vary by disorder status, so the di-

**Table 2**Baseline characteristics of patients with subthreshold depression (N=502) in Partners in Care<sup>a</sup>

Characteristic	Enhanced usual care (N=208)		Quality improvement intervention					
			Medication (N=144)		Therapy (N=150)		Pooled (N=294)	
	N	%	N	%	N	%	N	%
Married	116	56	85	59	88	58	173	59
Female	139	67	77	54	111	74	188	64
Education								
Less than high school	42	20	25	18	31	21	56	19
High school graduate	69	33	45	31	38	25	83	28
Some college	66	32	43	30	47	32	90	31
College graduate	31	15	30	21	34	22	64	22
Employed	141	68	89	62	96	64	185	63
Ethnicity								
White	103	50	80	56	79	53	159	54
Latino	72	35	44	31	46	30	90	31
Black	21	10	11	7	14	9	25	8
Other	12	6	9	7	12	8	21	7
Lifetime depressive disorder	110	53	61	43	73	48	134	46
Current anxiety	65	31	45	31	41	28	86	29
Chronic health conditions								
0	44	21	29	20	37	25	66	22
1	53	25	39	27	29	20	68	23
2	44	21	31	21	31	21	62	21
3 or more	68	33	45	31	52	35	97	33
Any antidepressant medication or counseling for depression in the past 6 months	56	33	39	26	43	28	82	27
Age (mean±SD)	41±12		45±17		47±18		46±17	
Health-related quality of life (mean±SD score)								
MCS-12 <sup>b</sup>	40±11		42±12		38±11		40±12	
PCS-12 <sup>c</sup>	44±12		44±12		44±12		44±12	

<sup>a</sup> Ns across rows may not total to the whole sample because of either rounding of weighted data or missing data for a given variable. The F test for three-group comparisons indicated significant differences in the percentage of women (F=4.10, df = 2 and 41, p=.02) and in mean mental health composite scores from the 12-item Short-Form Health Survey (MCS-12) score (F=4.00, df=2 and 41, p=.03). The t test for the pooled two-group comparison indicated significant differences for mean age (t=2.28, df= 42, p=.03) and lifetime depressive disorder (t=2.13, df =42, p=.04). The degrees of freedom were adjusted for clustering; weights were used to adjust differential probability of enrollment.

<sup>b</sup> Mental health composite score from the 12-item Short-Form Health Survey. Possible scores range from 0 to 100, with higher scores indicating better mental health. A score of 50 is equivalent to the general population mean.

<sup>c</sup> Physical health composite score from the 12-item Short-Form Health Survey. Possible scores range from 0 to 100, with higher scores indicating better physical health. A score of 50 is equivalent to the general population mean.

rect intervention costs represent around 10% of the overall intervention effect on costs for patients with depressive disorder and a much more substantial part for patients with subthreshold depression.

The estimated cost increases were much smaller and statistically insignificant for patients with subthreshold depression (Table 3). For example, the average cost increase for the pooled interventions was estimated to be only \$37 for patients with subthreshold depression. The interactions between intervention and disorder status were statistically insignificant.

The cost-effectiveness ratio for pooled intervention groups versus usual care was \$2,028 (CI=-\$17,225

to \$21,282) for those with subthreshold depression and \$53,716 (CI=\$14,194 to \$93,238) for those with depressive disorder, using the QALY-SF measure.

The comparable results for QALY-DB for patients with subthreshold depression ranged from \$2,180 (CI=-\$18,668 to \$23,028) with a QALY weight of -.2 for a depression burden day to \$1,090 (CI=-\$9,334 to \$11,514) with a QALY weight of -.4. For those with depressive disorder, the QALY-DB results ranged from \$36,204 (CI=\$17,575 to \$54,832) to \$18,102 (CI=\$8788 to \$27,416) for these two weights, respectively.

## Discussion

In this exploratory analysis we found

that implementing quality improvement interventions for depression across a sample that included patients with subthreshold depression and those with depressive disorder yielded cost-effectiveness ratios comparable to those of a widely used medical therapy among those with subthreshold depression. For this group, even the upper limit of the CI for the pooled interventions relative to usual care was within the range of that for widely used medical therapies (that is, \$11,514-\$23,028, depending on the QALY measure) (38,53). For patients with depressive disorder, the upper limits of the confidence interval were higher and not always within the range of a widely used medical therapy (that is, \$18,668-\$93,238).



**Table 3**

Average costs and outcomes over 24 months per patient with recent depressive disorder or subthreshold depression in Partners in Care (N=1,248)<sup>a</sup>

	Total under enhanced usual care		Incremental effect of quality improvement interventions											
			Medications <sup>b</sup>				Therapy <sup>b</sup>				Pooled			
	Total	95% CI	Total	95% CI	t <sup>c</sup>	p	Total	95% CI	t <sup>c</sup>	p	Total	95% CI	t <sup>c</sup>	p
<b>Depression</b>														
<b>12 months (N=746)</b>														
Health care costs	\$4,098	\$3,277 to \$4,921	\$1,051	-\$151 to \$2,253	1.71	.09	\$802	-\$442 to \$2,049	1.26	.21	\$912	-\$174 to \$1,999	1.64	.10
QALY-SF <sup>d</sup>	1.62	1.59 to 1.65	.01	-.02 to .03	.64	.52	.03	.00 to .05	2.27	.02	.02	-.00 to .04	1.65	.10
Days of depression burden over 24 months <sup>e</sup>	479	445 to 512	-.24	-67 to 19	1.11	.27	-64	-106 to -21	-2.94	<.01	-46	-84 to -8	-2.37	.02
Days of employment over 24 months <sup>f</sup>	275	250 to 300	20	0 to 40	1.92	.05	25	3 to 48	2.19	.03	23	5 to 41	2.48	.01
<b>Subthreshold (N=502)</b>														
Health care costs	\$3,419	\$2,724 to \$4,114	-\$130	-\$1,014 to \$754	-.29	.77	\$205	-\$978 to \$1,388	.34	.73	\$37	-\$826 to \$903	.08	.93
QALY-SF <sup>d</sup>	1.69	1.67 to 1.71	.02	-.00 to .04	1.57	.12	.02	-.00 to .04	1.6	.11	.02	-.00 to .04	1.92	.06
Days of depression burden over 24 months <sup>e</sup>	348	308 to 376	-.37	-87 to 14	-1.42	.16	-25	-70 to 21	-1.07	.28	-31	-71 to 9	-1.51	.13
Days of employment over 24 months <sup>f</sup>	284	252 to 316	17	-6 to 41	1.44	.15	13	-7 to 33	1.24	.22	15	-1 to 31	1.18	.07

<sup>a</sup> All tests compare outcomes in the respective intervention arm with outcomes under usual care by using a two-sided test. All models control for patient age, gender, marital status, education, rank in the distribution of household wealth, employment status, medical comorbidity, depressive disorder status, the 12-Item Short-Form Health Survey (SF-12) aggregate health-related quality-of-life measures, presence of comorbid anxiety disorder, practice randomization block, and interactions between disorder status and age, gender, the SF-12 mental components score, and probable lifetime depression.  
<sup>b</sup> The level for each outcome measure expected for each intervention group can be derived by summing the level for the enhanced usual care group and the level for the specific quality improvement intervention.  
<sup>c</sup> df=42  
<sup>d</sup> As measured by using a quality-adjusted life years measure based on the SF-12. Possible scores over two years range from 0 to 2, with higher scores indicating a higher number of quality-adjusted life years.  
<sup>e</sup> Possible scores range from 0 to 728, with higher scores indicating more burden days.  
<sup>f</sup> Possible scores range from 0 to 464, with higher scores indicating more employed days.

Thus it appears that implementing quality improvement for depression, using the PIC approach to intervention that emphasized adjusting treatment decisions to changing patient needs over time, may yield cost-effectiveness ratios, relative to usual care, that are comparable to widely used medical therapies among those with minor depression. Findings were similar among those with depressive disorder, if not as confidently within this range under all estimation scenarios.

We speculate that the PIC interventions may have been cost-effective

for patients with subthreshold depression, despite inconclusive evidence for the efficacy of acute treatment among such patients, because the interventions emphasized symptom monitoring and adjusting treatments as symptoms changed, rather than necessarily routing such patients directly to treatments. We emphasize that this finding should not be interpreted as evidence regarding the cost-effectiveness of active treatment for patients with subthreshold depression but rather as support for the cost-effectiveness of broader disease

management for such patients when they are part of a larger pool that includes patients with depressive disorder. Such a strategy might lead to active treatment for some, for example, if such patients developed a depressive disorder. Identifying the mechanisms underlying the present findings—and of course their replicability—requires further research.

It can be practically difficult or expensive to confirm a diagnosis of depressive disorder in order to route only patients with a depressive disorder into a quality improvement inter-

vention. The PIC approach to managing an at-risk group over time can offer practices an alternative quality improvement program that achieves cost-effectiveness ratios comparable to those in widely used therapies, even in patient groups including persons with subthreshold depression. Although statistical precision was limited for cost estimates and CIs were wide for cost-effectiveness ratios, we note that even the upper limit of the CI for cost-effectiveness for patients with subthreshold depression would represent a favorable cost-effectiveness ratio. Thus we are somewhat more confident of our main conclusion than is typical for an exploratory study. The main consequence of the limited precision in the study is that we cannot determine whether the interventions differed in their effectiveness or costs by initial disorder status, relative to usual care. To achieve greater confidence in anticipating average cost-effectiveness ratios or in estimating differential effectiveness or costs by patient subgroup would require much larger studies than this one and would involve larger samples than those found in prior trials of depression quality improvement programs (23,54,55).

We note that current standard estimates of cost-effectiveness are primarily meant for application to broad populations and not for comparisons of subgroups. It is quite likely that other interventions that are cost-effective overall may be less so for some subgroups, including primary targets (such as sicker patients), of disease management programs. The challenge to the field is to determine the best standards and methods for estimating cost-effectiveness of interventions for subgroups, particularly given the very large samples needed to do so from primary data, as well as to determine how cost-effectiveness estimates for subgroups should be used in health care planning. Meta-analyses and use of large descriptive databases may be strategies to overcome the precision challenges.

Limitations of this analysis include the reliance on particular practice locations, self-report data, cost estimates from 1998, reliance on pooled intervention groups for cost-effec-

tiveness estimates, and the limited precision for analyses of patients with subthreshold depression. Nevertheless, this preliminary analysis points out the potential importance of constructing interventions that are designed to manage a broad cross-section of patients at high risk of an illness and that promote a range of treatment, monitoring, and symptom management capacities.

### Conclusions

Our findings suggest that the PIC interventions, in the first two years of follow-up, were cost-effective within the range expected for a widely used medical therapy among patients with subthreshold depression and under most assumptions about QALYs for patients with depressive disorder. It may be possible to include patients with subthreshold depression in such disease management programs for broad at-risk groups and to achieve a cost-effective intervention strategy overall and for this at-risk subgroup.

### Acknowledgments and disclosures

This work was funded by grants 5RO1-MH-57992 and P30-MH-068639 from the National Institute of Mental Health (NIMH) and grant R01-HS-08349 from the Agency for Healthcare Research and Quality. The opinions expressed in the article are not necessarily the opinions of the NIMH, the National Institutes of Health, or the federal government. The authors are grateful to the practice organizations participating in this study, which provided access to their expertise and patients, implemented interventions, and provided in-kind resources: Allina Medical Group (Twin Cities, Minnesota), Patuxent Medical Group (Columbia, Maryland), Humana Health Care Plans (San Antonio), MedPartners (Los Angeles), PacificCare of Texas (San Antonio), and Valley-Wide Health Services (San Luis Valley, Colorado). The authors also thank their associated behavioral health organizations: Alamo Mental Health Group (San Antonio), San Luis Valley Mental Health/Colorado Health Networks (San Luis Valley, Colorado), and Magellan/Green-Spring Behavioral Health (Green Spring, Maryland). The authors are grateful to the clinicians and patients who contributed their time and efforts.

The authors report no competing interests.

### References

1. Katon W, Von Korff M, Lin E, et al: Collaborative management to achieve treatment guidelines: impact on depression in primary care. *JAMA* 273:1026–1031, 1995

2. Katon W, Robinson P, Von Korff M, et al: A multifaceted intervention to improve treatment of depression in primary care. *Archives of General Psychiatry* 53:924–932, 1996
3. Simon GE, Katon WJ, VonKorff M, et al: Cost-effectiveness of a collaborative care program for primary care patients with persistent depression. *American Journal of Psychiatry* 158:1638–1644, 2001
4. Hunkeler EM, Meresman J, Hargreaves WA, et al: Efficacy of nurse telehealth care and peer support in augmenting treatment of depression in primary care. *Archives of Family Medicine* 9:700–708, 2000
5. Rost K, Nutting P, Smith J, et al: Improving depression outcomes in community primary care practice: a randomized trial of the quEST intervention: Quality Enhancement by Strategic Teaming. *Journal of General Internal Medicine* 16:143–149, 2001
6. Katon W, Russo J, Von Korff M, et al: Long-term effects of a collaborative care intervention in persistently depressed primary care patients. *Journal of General Internal Medicine* 17:741–748, 2002
7. Rost K, Nutting P, Smith JL, et al: Managing depression as a chronic disease: a randomized trial of ongoing treatment in primary care. *British Medical Journal* 325:934–940, 2002
8. Liu CF, Hedrick SC, Chaney EF, et al: Cost-effectiveness of collaborative care for depression in a primary care veteran population. *Psychiatric Services* 54:698–704, 2003
9. Barrett JE, Williams JW, Oxman TE, et al: Treatment of dysthymia and minor depression in primary care: a randomized trial in patients aged 18–59 years. *Journal of Family Practice* 50:405–412, 2001
10. 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* 283:212–220, 2000
11. Unützer J, Katon W, Callahan C, et al: Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA* 288:2836–2845, 2002
12. Young AS, Klap R, Sherbourne CD, et al: The quality of care for depressive and anxiety disorders in the United States. *Archives of General Psychiatry* 58:55–61, 2001
13. Wang AS, Berglund P, Kessler RC: Recent care of common mental disorders in the United States: prevalence and conformance with evidence-based recommendations. *Journal of General Internal Medicine* 15:284–292, 2000
14. Hays RD, Wells KB, Sherbourne CD, et al: Functioning and well-being outcomes of patients with depression compared with chronic medical illness. *Archives of General Psychiatry* 52:11–19, 1995
15. Murray CJ, Lopez AD: The Global Burden of Disease: A Comprehensive Assessment of Mortality and Disability From Disease,

- Injuries, and Risk Factors in 1990 and Projected to 2020. Boston, Harvard School of Public Health on behalf of the World Health Organization and the World Bank, 1996
16. Kessler RC, Berglund P, Demler O, et al: The epidemiology of major depression disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 289:3095–3105, 2003
  17. Wells K, Sherbourne C, Schoenbaum M, et al: Five-year impact of quality improvement for depression: results of a group-level randomized controlled trial. *Archives of General Psychiatry* 61:378–386, 2004
  18. Judd LL, Schettler PJ, Akiskal HS: The prevalence, clinical relevance, and public health significance of subthreshold depression. *Psychiatric Clinics of North America* 25:685–698, 2002
  19. Oxman TE, Sengupta A: Treatment of minor depression. *American Journal of Geriatric Psychiatry* 10:256–264, 2002
  20. Sherbourne CD, Wells KB, Duan N, et al: Long-term effectiveness of disseminating quality improvement for depression in primary care. *Archives of General Psychiatry* 58:696–703, 2001
  21. Lave JR, Frank RG, Schulberg HC, et al: Cost-effectiveness of treatments for major depression in primary care practice. *Archives of General Psychiatry* 55:645–651, 1998
  22. Von Korff M, Katon W, Bush T, et al: Treatment costs, cost offset, and cost-effectiveness of collaborative management of depression. *Psychosomatic Medicine* 60:143–149, 1998
  23. Katon W, Unützer J, Fan MY, et al: Cost-effectiveness and net benefit of enhanced treatment of depression for older adults with diabetes and depression. *Diabetes Care* 29:265–270, 2006
  24. Williams JW Jr, Barrett J, Oxman T, et al: Treatment of dysthymia and minor depression in primary care: a randomized controlled trial in older adults. *JAMA* 284:1519–1526, 2000
  25. Wells K, Sherbourne C, Duan N, et al: Quality improvement for depression in primary care: do patients with subthreshold depression benefit in the long run? *American Journal of Psychiatry* 162:1149–1157, 2005
  26. Wells KB: The design of Partners in Care: evaluating the cost-effectiveness of improving care for depression in primary care. *Social Psychiatry and Psychiatric Epidemiology* 34:20–29, 1999
  27. Unützer J, Rubenstein L, Katon WJ, et al: Two-year effects of quality improvement programs on medication management for depression. *Archives of General Psychiatry* 58:935–9342., 2001
  28. Composite International Diagnostic Interview (CIDI). Geneva, World Health Organization, 1995
  29. Rubenstein LV, Jackson-Triche M, Unützer J, et al: Evidence-based care for depression in managed primary care practices. *Health Affairs* 18(5):89–105, 1999
  30. Depression Guidelines Panel: Depression in Primary Care: I. Detection and Diagnosis. Rockville, Md, US Department of Health and Human Services, 1993
  31. Depression Guidelines Panel: Depression in Primary Care: II. Treatment of Major Depression. Rockville, Md, US Department of Health and Human Services, 1993
  32. Muñoz RJ, Miranda J: Group Therapy for Cognitive Behavioral Treatment of Depression: San Francisco General Hospital Clinic, 1986. Document no MR01198/4. Santa Monica, Calif, RAND, 2000
  33. Muñoz RJ, Aguilar-Gaxiola S, Guzmán J: Group Therapy for Cognitive Behavioral Treatment of Depression: San Francisco General Hospital Clinic, 1986 [in Spanish]. Santa Monica, Calif, RAND, 2000
  34. Schoenbaum M, Unützer J, Sherbourne C, et al: Cost-effectiveness of practice-initiated quality improvement for depression: results of a randomized controlled trial. *JAMA* 286:1325–1330, 2001
  35. Ware JEJ, Kosinski M, Keller SD: SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales. Boston, Health Institute, New England Medical Center, 1995
  36. Lenert LA, Sherbourne CD, Sugar C, et al: Estimation of utilities for the effects of depression from the SF-12. *Medical Care* 38:763–770, 2000
  37. Radloff LS: The CES-D Scale: a self-report depression scale for research in the general population. *Applied Psychological Measurement* 1:385–401, 1977
  38. Gold M, Siegel J, Russel L, et al: Cost-Effectiveness in Health and Medicine. New York, Oxford University Press, 1996
  39. Tukey J: One degree of freedom for non-additivity. *Biometrics* 5:232–242, 1949
  40. Duan N: Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association* 78:605–610, 1983
  41. Manning WG: The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics* 17:283–295, 1998
  42. Bell RM, McCaffrey DF: Bias Reduction in Standard Errors for Linear Regression With Multi-Stage Samples. Florham Park, NJ, AT&T Labs, 2002
  43. Little RJA: Pattern-mixture models for multivariate incomplete data. *Journal of the American Statistical Association* 88:125–134, 1993
  44. Schafer J: Analysis of Incomplete Multivariate Data. London, England, Chapman and Hall, 1997
  45. Briggs A, Fenn P: Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. *Health Economics* 7:723–740, 1998
  46. Davison A: Bootstrap Methods and Their Application. Cambridge, United Kingdom, Cambridge University Press, 1997
  47. Fieller E: Some problems in interval estimation. *Journal of the Royal Statistical Society* 16:175, 1954
  48. Efron B: Better bootstrap confidence intervals. *Journal of the American Statistical Association* 82:172–200, 1987
  49. O'Brien BJ, Drummond MF, Labelle RJ, et al: In search of power and significance: issues in the design and analysis of stochastic cost-effectiveness studies in health care. *Medical Care* 32:150–163, 1994
  50. Hansen MH, Hurwitz W, Madow WG: Sample Survey Methods and Theory. Vol 1. New York, Wiley, 1953
  51. Cochran W: Sampling Techniques. New York, Wiley, 1977
  52. Miller RGJ: Simultaneous Statistical Inference. New York, Springer-Verlag, 1981
  53. Tengs TO, Adams ME, Pliskin JS, et al: Five-hundred life-saving interventions and their cost effectiveness. *Risk Analysis* 15:369–390, 1995
  54. Katon WJ, Schoenbaum M, Fan MY, et al: Cost-effectiveness of improving primary care treatment of late-life depression. *Archives of General Psychiatry* 62:1313–1320, 2005
  55. Sturm R, Unützer J, Katon W: Effectiveness research and implications for study design: sample size and statistical power. *General Hospital Psychiatry* 21:274–283, 1999