

The Effects of Quality Improvement for Depression in Primary Care at Nine Years: Results from a Randomized, Controlled Group-Level Trial

Kenneth B. Wells, Lingqi Tang, Jeanne Miranda, Bernadette Benjamin, Naihua Duan, and Cathy D. Sherbourne

Objective. To examine 9-year outcomes of implementation of short-term quality improvement (QI) programs for depression in primary care.

Data Sources. Depressed primary care patients from six U.S. health care organizations.

Study Design. Group-level, randomized controlled trial.

Data Collection. Patients were randomly assigned to short-term QI programs supporting education and resources for medication management (QI-Meds) or access to evidence-based psychotherapy (QI-Therapy); and usual care (UC). Of 1,088 eligible patients, 805 (74 percent) completed 9-year follow-up; results were extrapolated to 1,269 initially enrolled and living. Outcomes were psychological well-being (Mental Health Inventory, five-item version [MHI5]), unmet need, services use, and intermediate outcomes.

Principal Findings. At 9 years, there were no overall intervention status effects on MHI5 or unmet need (largest $F(2,41) = 2.34, p = .11$), but relative to UC, QI-Meds worsened MHI5, reduced effectiveness of coping and among whites lowered tangible social support (smallest $t(42) = 2.02, p = .05$). The interventions reduced outpatient visits and increased perceived barriers to care among whites, but reduced attitudinal barriers due to racial discrimination and other factors among minorities (smallest $F(2,41) = 3.89, p = .03$).

Conclusions. Main intervention effects were over but the results suggest some unintended negative consequences at 9 years particularly for the medication-resource intervention and shifts to greater perceived barriers among whites yet reduced attitudinal barriers among minorities.

Key Words. Depression, quality improvement, long-term outcomes

Quality improvement (QI) programs for depression in primary care reduce symptoms and improve functioning among depressed patients (Katon et al.

1995, 1996, 2001, 2004; Katon, Von Korff, and Lin 1999; Hunkeler et al. 2000; Rost et al. 2001, 2002; Sherbourne et al. 2001; Simon, Katon et al. 2001; Simon, Von Korff et al. 2001; Unutzer et al. 2001, 2002; Schoenbaum et al. 2002; Araya et al. 2003; Gilbody et al. 2003; Hedrick et al. 2003; Miranda et al. 2003b; Bruce et al. 2004; Ciechanowski et al. 2004; Dietrich et al. 2004; Neumeyer-Gromen et al. 2004; Asarnow et al. 2005; Katzelnick et al. 2005). Programs based on a collaborative care model can improve patient quality of care and symptom outcomes as well as employment for 2 years or more (Neumeyer-Gromen et al. 2004; Katon and Unutzer 2006), with similar findings for employer-sponsored programs (Wang et al. 2007). Programs supporting more sustained implementation find improved outcomes over that full implementation period (Rost et al. 2002; Hunkeler et al. 2006).

The Partners-in-Care (PIC) study reported individual health outcome benefits at 5-year follow-up after implementation of 6–12 months QI interventions, relative to usual care (UC) (Wells et al. 2004). PIC evaluated two interventions, one that provided resources for antidepressant medication management for 6–12 months (QI-meds) and one that provided resources for use of Cognitive Behavioral Therapy (CBT) (QI-therapy) for 6 months at the client level. During the first year of follow-up and at 5 years, underserved minority groups benefited more than did whites in clinical outcomes and outcome disparities relative to UC were reduced, especially for QI-therapy relative to UC (Wells et al. 2004). Other studies have found intervention effects of QI or treatments for depression among underserved minorities (Miranda et al. 2003b; Arean et al. 2005).

We previously proposed potential mechanisms underlying long-term outcome effects of short-term QI programs for depression (Wells et al. 2004), including: (1) improved depression outcomes reduce the risk for subsequent episodes; (2) improved provider knowledge or skills could improve long-term management; (3) client learning could improve help seeking and treatment

Address correspondence to Kenneth B. Wells, M.D., M.P.H., The RAND Corporation, 1776 Main Street, Santa Monica, CA 90401; David Geffen School of Medicine, 10833 Le Conte Ave, Los Angeles, CA 90095; Semel Institute for Neuroscience and Human Health, 760 Westwood Plaza, Los Angeles, CA 90095; UCLA School of Public Health, 10911 Weyburn Ave, Los Angeles, CA 90024; e-mail: kwells@ucla.edu. Lingqi Tang, Ph.D., and Jeanne Mirande, Ph.D., are with the Semel Institute for Neuroscience and Human Health, Los Angeles, CA. Dr. Miranda is also with the David Geffen School of Medicine, Los Angeles, CA. Bernadette Benjamin, M.S., and Cathy D. Sherbourne, Ph.D., are with the The RAND Corporation, Santa Monica, CA. Naihua Duan, Ph.D., is with the Departments of Psychiatry and Biostatistics, Columbia University and the Division of Biostatistics, New York State Psychiatric Institute, New York, NY.

decisions over time; (4) client learning of coping skills such as avoidance of circumstances leading to stressful life events could reduce recurrence of depression; (5) structural changes in participating practices could facilitate access to treatments in the long run. We doubt that practice or provider changes could account for long-term client outcomes because many patients changed providers or practices by 2 years of follow-up (Orlando and Meredith 2002). Our anecdotal experience suggested that most practices discontinued intervention activities shortly after study completion. Unmet need for depression treatment was reduced by the interventions for minorities at 5 years, suggesting improved helpseeking when sick could be a factor (Wells et al. 2004). In a structural analysis of effects of the psychotherapy intervention over 9 years of follow-up, we found additional support for the first and fourth-listed mechanisms above. Specifically, we found that this intervention improved depression outcomes at 1 year and reduced the occurrence of stressful life events at 5 years, and both of these effects lead to improved depression outcomes at 9 years. This provides evidence of indirect long-term intervention benefits (Sherbourne et al. 2008).

These findings increased our interest in looking comprehensively at direct intervention effects at 9-year follow-up. Long-term patient learning effects might be expected for PIC interventions because they did not route patients to treatments but supported treatment according to patient preferences but based on evidence, and adjusted treatment decisions over time based on client progress (Wells et al. 2007, 2004; Sherbourne et al. 2008).

In the present study, we provide the first comprehensive snapshot of long-term outcomes of the short-term PIC QI interventions relative to UC, at 9-year client follow-up. There have been few studies of such long-term effects of either depression QI interventions or treatments. Initially, on seeking funding for this effort, we hypothesized that the interventions would continue to improve mental health status, especially the psychotherapy resource intervention for underserved minorities. However, we thought that either process of care outcomes or intermediate outcomes such as perceived barriers to care might reflect a mix of the beneficial effects of intervention exposure (Neu-meyer-Gromen et al. 2004; Katon and Unutzer 2006), and potential negative consequences of losing facilitated access to treatments found to be helpful, over time as patients changed practices. We also thought that the balance of such effects might differ for whites and underserved minorities because of different levels of care and experiences with health care for depression both before and over the course of the study. For example, we were concerned that subjects encouraged to enter treatment under the interventions might

experience long-term difficulty with insurance coverage based on prior-condition exclusion criteria when changing jobs. There are few empirical precedents for such hypotheses in this context so we considered our analyses exploratory. Our concerns about a mixture of consequences of exposure to and termination of interventions lead us to examine several intermediate outcomes such as perceived barriers to care, in addition to the main mental health and unmet need outcomes. The clear indication in prior follow-ups that minorities particularly benefited from the interventions also lead us to examine intervention effects overall and for whites compared with underserved minority groups (African Americans and Latinos combined).

METHODS

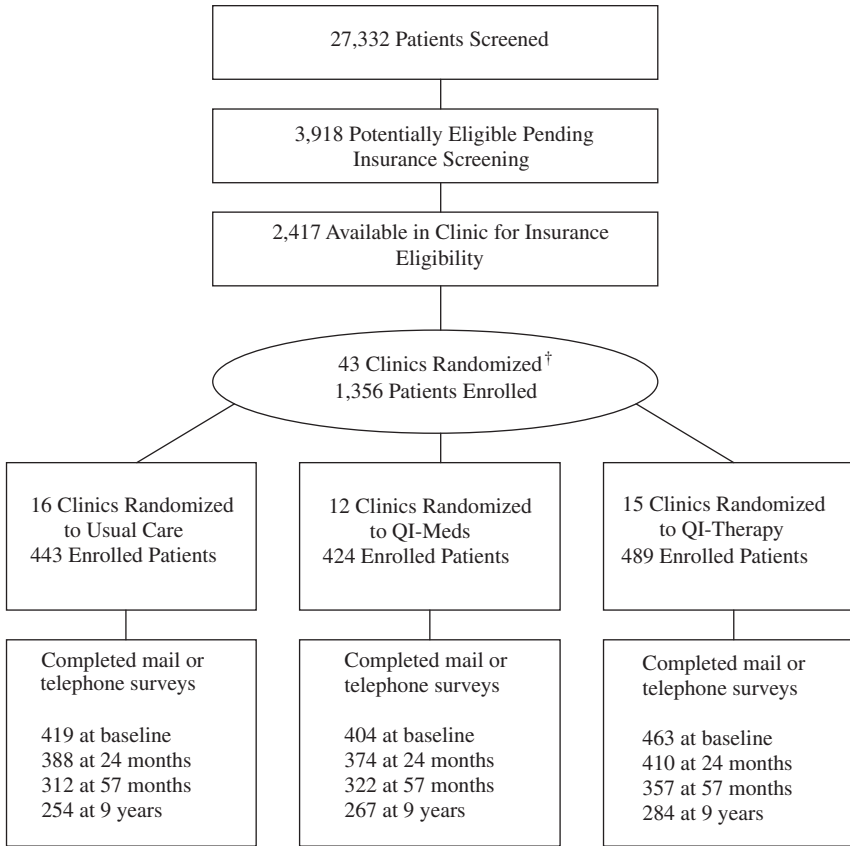
Experimental Design and Implementation

We examine 9-year follow-up from PIC, a group-level, randomized-controlled trial of practice-initiated QI programs for depression (Wells et al. 2000; Schoenbaum et al. 2001; Sherbourne et al. 2001; Unützer et al. 2001; Miranda et al. 2003b). Six nonacademic-managed care organizations participated, with 46 of 48 eligible primary care practices (clinics) and 181 of 183 primary care clinicians. Within organizations, practices were matched into blocks of three clusters based on factors expected to affect outcomes (specialty mix, patient socioeconomic and demographic factors, and having mental health specialists nearby). Within blocks, practices were randomized to enhance UC (mailing of written practice guidelines to medical directors), or to one of two interventions, which we refer to as QI-meds and QI-therapy (defined below).

Study staff screened 27,332 consecutive patients in these practices over a 5–7-month period for each practice, between June 1996 and March 1997. Patients were eligible if they intended to use the practice for the next 12 months and screened positive for current depressive symptoms plus probable depressive disorder in the last year, using a screener developed from stem items of the World Health Organization's 12-month Composite International Diagnostic Interview (CIDI) (World Health Organization 1995). Patients were ineligible if they were younger than 18 years, not fluent in English or Spanish, or lacked insurance coverage for the local therapists participating in the interventions. The 9-year follow-up was approved by the Institutional Review Boards of RAND and UCLA.

Of those completing the screener, 3,918 were eligible, 2,417 were available to confirm insurance eligibility, and 241 were ineligible. Of those reading

Figure 1: PIC Patient Screening, Enrollment, and 9-Year Follow-up*



*QI-meds indicates quality improvement plus medication management; QI-therapy, quality improvement plus psychotherapy.

informed consent, 1,356 (70 percent) enrolled, including 443 in UC, 424 in QI-meds, and 489 in QI-therapy (Figure 1).

Interventions

The QI interventions are described elsewhere (Wells 1999). All QI materials are available at <http://www.rand.org/organization/health/pic.products/order.html>. Before implementation, the study provided a payment of up to half the estimated practice participation costs (\$35,000–70,000). The interventions

provided practices with training and resources to initiate and monitor QI programs and adapt them to local practice goals and resources. Patients and clinicians retained choice of treatment and use of QI materials. The study provided training and offered limited support for implementation.

For both interventions, local practice teams were trained in a 2-day workshop to educate primary care clinicians through lectures, academic detailing, or audit and feedback, and to supervise intervention staff and conduct team oversight based on practice guidelines (Depression Guidelines Panel 1993a, b). Practice nurses were trained to help in patient assessment, education, and activation for treatment. Practice teams were given patient education pamphlets, videotapes, and tracking forms, and clinician manuals, lecture slides, and pocket reminder cards to distribute. The materials described guideline-concordant care for depression, encouraged attention to patient preferences, and advised adjusting treatment to patient characteristics and course of illness over time.

In the QI-meds program, nurse specialists were trained to support medication adherence through monthly visits or telephone contacts for 6 or 12 months, randomized at the patient level. In QI-therapy, practice therapists were trained to provide individual and group CBT (Muñoz, Aguilar-Gaxiola, and Guzmán 2000; Muñoz and Miranda 2000). This therapy was available at the primary care copay (about \$5–10) for 6 months. All patients could have other therapy at usual copay. Supervision was provided by local experts assisted by study experts. In all conditions, patients could have medications, therapy, both, or neither. Practices were given permission to modify the implementation plan. Intervention patients and clinicians were free to use or not the intervention resources. After the active intervention phase, implementation support by the study was terminated but practices retained their training manuals and extra resources of patients and clinicians. Based on informal telephone follow-up with site leaders, few practices continued implementation after the study.

Data Collection

Patients were asked to complete a screener at their physician visit, a telephone-administered CIDI for diagnosing depression, and questionnaire on economic status, and a mailed survey at baseline. We report data from mailed surveys at 12 and 24 months, a follow-up telephone survey at 57 months (limited to subjects who completed the 24-month survey), and a 9-year follow-up telephone survey, conducted March–December 2005. Completion rates relative to all initial enrollees (1,356) are 88 percent for the baseline survey ($N = 1,187$),

83 percent for the 12-month survey ($N=1,126$), 86 percent for the 24-month survey ($N=1,159$), 73 percent for the 57-month survey ($N=991$), and 59 percent for the 9-year survey ($N=805$), representing 63 percent of the 1,269 initial enrollees still alive at 9 years. We took this sample (1,269) as the main analysis sample. We examined predictors of nonresponse to the 9-year survey using logistic regressions conducted separately by intervention condition. Under UC, males, younger individuals, and minorities were significantly less likely to complete a survey ($\chi^2(1)$ ranges from 4.15 to 14.89, each $p < .05$). Under QI-Meds, Latinos and African Americans, compared with non-Hispanic whites were less likely to respond ($\chi^2(1)$ from 3.77 to 7.73, each $p = < .05$). Under QI-Therapy, males and those with less education were less likely to respond ($\chi^2(1)$ from 4.79 to 8.46, each $p < .05$). These predictors are included as covariates in analyses.

Measures

Intervention Status. We used indicators for each intervention (QI-meds and QI-therapy) versus UC.

Primary Outcomes. We selected the Mental Health Inventory, five-item version (MHI5) as the primary health outcome (Ware and Sherbourne 1992). It includes five items that assess symptoms of depression and anxiety, loss of behavioral or emotional control, and psychological well-being in the prior month.

Unmet Need for Depression Treatment. Over long-term follow-up, patients' needs for care may change. We developed an indicator of unmet need for depression treatment, that contrasts persons who have probable depressive disorder but are not receiving treatment for depression, versus others (Wells et al. 2000). The indicator of probable depressive disorder is based on a repeat of the screener measure for the prior 6 months, removing the dysthymia items. The indicator of treatment for this variable was having four or more specialty visits or use of an antidepressant medication for 2 months or more in the prior 6 months. This measure is an indicator of overall appropriateness of care given outcome status at follow-up.

Use of Services and Treatments. We assessed the presence or absence in the prior 6 months of any use of general medical outpatient visit, any use of mental health specialists. We assessed whether or not respondents used any antidepressant medication (and separately, any use for 2 months or more) in the prior 6 months.

Intermediate Outcomes. We selected as intermediate outcomes several measures that reflected potential explanatory pathways for long-term outcomes (Wells et al. 2004). We included a single-item measure of perceived effectiveness of attempts to cope with the most stressful event in the last year (a higher score is less-effective coping); we thought that difficulties coping could both be a consequence of depression and represent a risk factor for subsequent depression. We included items assessing whether respondents had delays or difficulties in obtaining care for a mental health problem in the last year due to 16 specific barriers, including: worrying about cost; the provider would not accept insurance; the health plan would not pay for treatment; the respondent could not find where to go; could not get an appointment; could not get to the provider's office, or it takes too long to get to the office; the respondent could not get through on the phone; did not think it would help; was embarrassed to discuss the problem or were afraid of what others would think; would lose pay from work; that the respondent needed someone to take care of their children; no one spoke the respondent's language at the clinic; the respondent felt discriminated against because of their race or cultural background; or the respondent felt they could get over their problems on their own. We developed an indicator of having any barrier. We conducted exploratory analyses of single items. We included a three-item measure of tangible social support, to reflect resources to support help seeking.

Covariates. From the patient screener, we measured self-reported age, sex, education (less than high school, completed high school, some college, completed college or more), race/ethnicity (white, nonwhite), physical and mental health composites from the SF-12 (Ware, Kosinski, and Keller 1995) and a count of having 0, 1, 2, or more than 2 chronic medical conditions out of 19. We used data from the screener and baseline CIDI to categorize patients as having 30-day symptoms plus having a depressive disorder sometime during the past year versus symptoms plus no 1-year disorder. Using items modeled after the Health and Retirement Survey, we developed a baseline household wealth variable, summing the net value of home and other assets. We used indicators for each randomization practice cluster. We conducted sensitivity analyses including when available a baseline measure of the dependent variable as an additional covariate, with no change in conclusions or substantive results.

Data Analysis

We conducted patient-level, intent-to-treat analyses of end status at 9 years. For each dependent variable, we estimated a multiple regression model with

QI-meds and QI-therapy, relative to UC, as the independent variables, with the covariates above. For dichotomous measures, we estimated logistic regression models. For continuous measures, we conducted linear regression on untransformed scores. In separate analyses, we interacted each intervention indicator with ethnicity (African American/Hispanics versus non-Hispanic whites), excluding persons with ethnicity other than white, African American, or Hispanic ($n = 81$).

Significance of comparisons by intervention status and tests of interactions were based on regression coefficients. We followed a two-level strategy to consider significance of findings. Level 1 (planned) is used to designate statistical significance of 0.05 or stronger for either (1) the overall test of the difference among the three intervention arms for the sample as a whole; (2) the overall test within a specific ethnic group; or (3) the overall interaction of intervention status with ethnicity. Level 2 (exploratory) is used when a pairwise comparison of two intervention arms (such as QI-meds versus UC) is significant at 0.05 or stronger but the overall main intervention test is not significant. To guard against multiple statistical comparisons, we focus our conclusions on Level 1 findings and report actual p -values.

We used weighting by baseline predictors of enrollment to reflect the characteristics of the eligible screener sample with probable depression. We adjusted for clustering of patients within providers and clinics using a modification of the usual sandwich variance estimator, the bias-reduced linearization method (BRL) developed by McCaffrey and Bell (2006). The degrees of freedom for t -tests and F -tests were based on the number of clusters. We illustrate average results for an intervention group adjusted for all covariates using standardized predictions generated from the fitted regression model: We used the regression parameters and each individual's actual values for all covariates other than intervention status to calculate the predicted outcome assuming the patient had been assigned to UC or to either intervention, respectively, and calculated the mean prediction under each scenario (Graubard and Korn 1999).

Attrition. In prior PIC analyses, we weighted the data for baseline predictors of attrition at each wave, and imputed item nonresponse using multiple imputation of five data sets (Little 1988; Bell 1999). But at 9-year follow-up, baseline variables are not necessarily good predictors, so we used the approximate Bayesian bootstrap method for unit nonresponse (Lavori, Dawson, and Shera 1995; Tang et al. 2005) to deal with wave nonresponse at year 9. We imputed five data sets for each of the five item-level imputed data

sets, for a total of 25 imputed data sets. For this two-stage nested multiple imputation, we used an extension of Rubin's conventional multiple imputation inference. The point estimates were averaged across the 25 data sets. The standard errors involve three components of variability: the estimated complete data variance, the between-nest variance, and the within-nest variance (Schafer 1997; Shen 2000). *F*-statistics are adjusted by the design effect resulting from two-stage imputation. The imputations were conducted across waves for all participants in the main analyses ($N = 1,269$ for the whole sample; $N = 1,188$ for the sample of whites and African Americans/Latinos). Unless otherwise specified, the imputed analytic N is 1,269 but the sample responding at 9 years is 805. Further details of the multiple imputation procedure can be found in Tang et al. (2007). We conducted sensitivity analyses using unweighted raw data without unit imputation (but including item imputation on covariates and ethnicity to permit consistent sample sizes) on the 805 completing the 9-year survey, with the same main intervention conclusions; a few secondary findings have less significance in unweighted analyses.

Table 1 shows baseline characteristics by intervention status for the 1,269 participants, weighted for differential probability of enrollment. The two significant differences are in percent female ($F(2,41) = 4.79$, $p = .01$; relatively more females in QI-therapy) and in the percent with depressive disorder rather than subthreshold depressive symptoms ($F(2,41) = 4.68$, $p = .01$, relatively more with disorder in QI-therapy). These variables are included as covariates in analyses.

RESULTS

Primary Outcomes (MHI5, Unmet Need for Treatment, Table 2)

Neither of the main overall intervention effects was statistically significant (largest $F(2,41) = 2.34$, $p = .11$). The effect of QI-Meds relative to UC on MHI5 was negative at $p = .05$. The overall interactions (minority versus white \times intervention status) were not significant (largest $F(2,41) = 1.79$, $p = .18$).

Services Use (Table 3)

No statistically significant overall intervention effects were found on services use measures for the whole sample (largest $F(2,41) = 2.05$, $p = .14$, results not

Table 1: Baseline Characteristics of Analytic Sample: Initially Enrolled Participants Alive at 9-Year Follow-up ($N = 1,269$)*

Characteristic	UC ($N = 421$)	QI-Meds Group ($N = 397$)	QI-Therapy Group ($N = 451$)	Difference across Groups [†]		
				F	df ₁ , df ₂	p
Age, mean (SD), y	41.27 (13.41)	42.33 (13.65)	42.63 (14.53)	0.30	2, 41	.74
Female (%)	69.51	67.95	78.70	4.79	2, 41	.01
Married (%)	53.43	55.06	55.59	0.12	2, 41	.89
Employed (%)	67.99	66.04	68.41	0.07	2, 41	.93
Education (%)				0.94	6, 37	.48
< 12 years	18.95	16.55	17.08			
High school graduate	33.65	29.10	26.85			
Some college	31.88	30.88	33.55			
College	15.52	23.47	22.52			
Ethnicity (%)				0.32	6, 37	.92
White	55.01	59.88	52.64			
Latino	31.08	27.36	34.18			
African American	8.65	6.00	6.49			
Other	5.26	6.76	6.69			
Depressive disorder (%)	49.89	57.49	62.14	4.68	2, 41	.01
Anxiety disorder (%)	42.58	45.81	44.07	0.46	2, 41	.63
PCS-12 score, mean (SD)	44.75 (11.58)	46.00 (11.32)	46.36 (11.02)	0.95	2, 41	.40
MCS-12 score, mean (SD)	36.39 (10.98)	35.65 (10.57)	34.83 (10.47)	1.41	2, 41	.26
Chronic disease count (%)				0.18	6, 37	.98
None	21.76	23.97	24.88			
One	25.14	25.47	24.98			
Two	19.89	20.08	20.19			
Three or more	33.21	30.48	29.95			

*Sample of 1,269 who were enrolled to PIC study and still living at 9-year follow-up.

[†]The statistics were based on F -tests comparing differences across three intervention arms; the denominator degrees of freedom was adjusted for clustering; weights were used to adjust differential probability of enrollment ("Methods"). Data are presented as weighted percentage unless otherwise indicated.

QI-Meds, quality improvement programs supporting education and resources for medication management; QI-Therapy, quality improvement programs supporting evidence-based psychotherapy; UC, usual care.

shown). Among whites, there was a Level 1 significant intervention effect on use of any outpatient medical visit ($F(2,41) = 3.27, p = .05$); both QI-meds and QI-therapy reduced the likelihood of such a visit by about 6–7 percentage points relative to UC (lowest $t(42) = 2.21, p = .03$). The intervention effect was not statistically significant among minorities ($F(2,41) = .42, p = .66$).

Table 2: Intervention Effects on Primary Outcomes (N = 1,269)*

	UC (n = 421)	QI-Meds Group (n = 397)	QI-Therapy Group (n = 451)	Difference across Groups F(2,41) (p-Value)
Primary outcomes				
MHI5 score				
Mean	64.91	60.74	61.87	2.34 (.11)
95% CI	61.97–67.85	57.41–64.06	58.93–64.80	
t(42) [†] (p)	‡	– 2.02 (.05)	– 1.57 (0.12)	
Unmet need for treatment				
%	17.90	21.04	20.22	0.35 (.71)
95% CI	12.01–23.79	14.33–27.75	13.17–27.27	
t(42) [†] (p)	‡	0.75 (.46)	0.54 (.59)	
Intermediate outcome				
Effectiveness of coping with stressful event	2.33 (2.17–2.49)	2.67 (2.51–2.82)	2.52 (2.36–2.67)	3.91 (.03)
		2.99 (<.01)	1.68 (.10)	

*See “Methods” for model specifications (weighted for adjusting differential probability of enrollment; multiple imputation for item and wave nonresponse; actual completes at 9 years, 805; adjusted for covariates and group-level randomization).

[†] t-test for pairwise comparisons between QI-Meds group (or QI-Therapy group) and UC group.

[‡]Reference group.

QI-Meds, quality improvement programs supporting education and resources for medication management; QI-Therapy, quality improvement programs supporting evidence-based psychotherapy; UC, usual care.

The overall intervention effect on use of specialty visits was not statistically significant ($F(2,41) = .05, p = .95$). Among whites, there was a Level 1 significant intervention effect on use of any antidepressant medication ($F(2,41) = 3.29, p = .05$) with the highest likelihood for QI-meds and the lowest for QI-therapy. None of the interaction effects for these variables were statistically significant (largest $F(2,41) = 1.72, p = .19$).

Intermediate Outcomes (Tables 2 and 3)

Barriers to Care. For the likelihood of any delays/difficulties in care, there was a Level 1 significant interaction effect ($F(2,41) = 4.24, p = .02$). Among whites, there was a Level 1 significant overall intervention effect ($F(2,41) = 5.51, p = .01$), with a higher percentage of patients reporting any barrier in both QI-meds and QI-therapy compared with UC ($t(42) = 2.97$ and 2.63 , respectively, each $p \leq .01$). Among minorities, there was no significant overall intervention effect ($F(2,41) = 1.13, p = .33$).

Table 3: Intervention Effects on Services Use, Barriers, and Tangible Support Measures, by Ethnicity (N = 1,188)*

	UC Group (N = 398)	QI-Meds Group (N = 369)	QI-Therapy Group (N = 421)	Difference across Groups [‡] F(2,41) (p)
Outpatient medical visit				
Latino/African American combined	86.60 (78.95–94.25)	86.33 (77.15–95.51)	91.22 (83.50–98.94)	0.42 (.66)
% (95% CI)				
t(42) (p) [†]		-0.02 (.99)	0.83 (.41)	
White				
% (95% CI)	93.39 (89.84–96.94)	86.77 (81.39–92.15)	86.93 (82.30–91.57)	3.27 (.05)
t(42) (p) [†]		-2.30 (.03)	-2.21 (.03)	
QI-minority interaction				
Any mental health specialty visit				
Latino/African American combined	17.32 (7.20–27.44)	23.07 (10.95–35.19)	14.07 (6.91–21.23)	0.96 (.39)
% (95% CI)				
t(42) (p) [†]		0.74 (.46)	-0.56 (.58)	
% (95% CI)	23.67 (15.89–31.46)	21.15 (13.78–28.52)	25.75 (19.31–32.18)	0.46 (.63)
White				
t(42) (p) [†]		-0.45 (.65)	0.42 (.68)	
QI-minority interaction				
Any antidepressant				
Latino/African American combined	26.11 (13.65–38.57)	34.18 (21.87–46.49)	29.83 (20.15–39.51)	0.42 (.66)
% (95% CI)				
t(42) (p) [†]		0.87 (.39)	0.48 (.64)	
% (95% CI)	44.43 (36.71–52.15)	49.37 (40.17–58.57)	34.47 (26.78–42.17)	3.29 (.05)
White				
t(42) (p) [†]		0.77 (.45)	-1.79 (.08)	
QI-minority interaction				
t(42) (p) [†]				0.86 (.43)

Any barrier to care							
Latino/African American combined							1.13 (.33)
% (95% CI)						60.61 (51.25–69.97)	
<i>t</i> (42) (<i>p</i>) [‡]						– 1.53 (.13)	
White							5.51 (.01)
% (95% CI)						53.05 (44.48–61.63)	
<i>t</i> (42) [‡] (<i>p</i>)						2.63 (.01)	
QI-minority interaction							4.24 (.02)
Tangible support							
Latino/African American combined							0.18 (.83)
Mean (95% CI)						3.49 (3.24–3.75)	
<i>t</i> (42) (<i>p</i>) [‡]						– 0.57 (.57)	
White							3.86 (.03)
% (95% CI)						3.91 (3.74–4.09)	
<i>t</i> (42) (<i>p</i>) [‡]						– .39 (.70)	
QI-minority interaction							1.39 (.26)

*Sample of 1,188 who were enrolled to PIC study and still living at 9-year follow-up. Eighty-one from “other race” category were excluded from the analysis. Logistic regression for binary variables (outpatient medical visit, any mental health specialty visit, any antidepressant, any barrier to care) and linear regression for continuously scaled variable (tangible support). Models were adjusted for differential probability of enrollment, multiple imputation for item and wave nonresponse, baseline covariates, and group-level randomization. See “Data Analysis” for details.

[‡]Level 1 *F*-test for the overall intervention effects within a given specific ethnic group; or the overall interaction of intervention status with ethnicity.

[†]*t*-test for pairwise comparison between QI-Meds group (or QI-Therapy group) and UC group.
 PIC, partners-in-care; QI-Meds, quality improvement programs supporting education and resources for medication management; QI-Therapy, quality improvement programs supporting evidence-based psychotherapy; UC, usual care.

In analyses of specific barriers, there was a significant intervention effect among whites on barriers due to insurance not paying for treatment ($F(2,41) = 4.71, p = .01$), with a higher likelihood of reporting this as a barrier for QI-meds compared with UC ($t(42) = 2.90, p = .01$). Among whites there was a borderline significant effect on barriers due to difficulty finding providers ($F(2,41) = 3.06, p = .06$), with both intervention groups having greater difficulty than UC (lowest $t(42) = 2.09, p = .04$). These effects ranged from 7 to 13 percentage points, while for minorities, the range is from -5 to 2 (largest $t(42) = 0.77, p = .45$). The interaction effects were not statistically significant (largest $F(2,41) = 2.20, p = .12$).

In contrast, for barriers due to respondents thinking they could handle the problem on their own, there was a significant overall intervention effect among minorities ($F(2,41) = 6.06, p < .01$) and a significant interaction with ethnic status ($F(2,41) = 6.95, p < .01$). For minorities, the percentage with this barrier was 58.84 (95 percent confidence interval [CI] = 48.76–68.92) in UC, 31.24 (95 percent CI = 16.39–46.10) in QI-meds and 36.56 percent (95 percent CI = 24.88–48.25) in QI-therapy. Among whites, the percentage with this barrier did not differ significantly across intervention groups ($F(2,41) = 0.83, p = .44$). The measure of having a barrier due to racial discrimination was low across groups QI-meds, 0.37 percent (95 percent CI = 0.00–1.12); QI-therapy, 1.13 percent (95 percent CI = 0.00–2.42); UC, 4.05 percent (0.78–7.31). Because some subgroups reported no barrier of this type, we conducted exact logistic regression (Hirji 1992), using the permutation test to account for clinic level clustering (Manly 1997), and calculated p -values from a Monte Carlo approximated permutation test on 10,000 replicates for score statistics in exact logistic regression. The overall intervention effect is Level 1 significant among minorities ($p = .04$), with fewer barriers due to discrimination in both interventions than UC. The effect was not significant among whites ($p = .69$) and the interaction effect was not significant ($p = .68$).

Coping. There was a Level 1 significant intervention status effect on the effectiveness of coping with the most stressful event in the year ($F(2,41) = 3.91, p = .03$), with those in QI-meds having more difficulty compared with UC ($t(42) = 2.99, p < .01$). The overall intervention effect was Level 1 significant among minorities ($F(2,42) = 4.01, p = .03$), but not whites ($F(2,41) = 1.16, p = .321$), but the interaction effect was not statistically significant ($F(2,41) = 1.17, p = .32$).

Tangible Social Support. There was a Level 1 significant overall intervention effect among whites on level of tangible social support ($F(2,41) = 3.86$, $p = .03$); whites in QI-meds had less support compared with UC ($t(42) = -2.39$, $p = .02$) or QI-therapy ($t(42) = -2.30$, $p = .03$). The interaction effect was not statistically significant ($F(2,41) = 1.39$, $p = .26$).

DISCUSSION

Nine years is a long time to expect effects from QI for depression in primary care, but we have been surprised previously by observed direct and indirect long-term consequences of PIC interventions, and have documented the high magnitude of cumulative benefits for minorities across the follow-up years (Wells et al. 2007). At the 9-year time-point, it is likely that the intervention effects reflect both consequences of initial exposure, such as patient learning or changes in attitudes, and of terminating the interventions or clients losing access to facilitated care through changing providers. PIC was not designed to look at the consequences of program termination versus continuation, so we cannot tease out exposure and termination effects except by speculating based on the particular pattern of results.

For “main” outcomes (MHI5 and unmet need for appropriate care), we found little evidence for overall intervention status effects (across all three intervention conditions) at 9 years, confirming with more complete data partial findings on mental health outcomes in recent studies (Wells et al. 2007; Sherbourne et al. 2008). However, we found that QI-meds relative to UC reduced mental health outcome at 9 years, a finding consistent with the similar observed effect (overall and among minorities) on reduced coping with stress at 9 years. These results could have several explanations. One could be a shift away from psychological coping strategies due to the emphasis during the intervention on medications—at 9 years we also found continuing higher rates of antidepressant medication use among whites under QI-Meds than UC. Another explanation could be that the distress and coping difficulty reflects the consequences of changes in access to treatments previously found to be valuable. Among whites, for example, it appears that there was a perception for those in the intervention clinics, particularly QI-Meds, of difficulties later with insurance coverage and finding providers. Lower rates of medical visits among whites under QI-Meds compared with UC could similarly reflect an access problem.

Given that early on the interventions improved access to treatments, it is difficult to know what may have created the perceptions and experiences of barriers at very long-term follow-up. One possibility could be an upward shift in expectations for ease of getting appropriate care, from having had a prior experience with QI that was not met when facing new health care systems. Alternatively, subjects may actually have had access difficulties, or were concerned about them, as a result of having had prior care. For example, a documented history of depression treatment could lead to being excluded from insurance coverage for depression treatment at the time of a subsequent job change under some employment or insurance policies (i.e., preexisting condition exclusion). By encouraging patients to consider new treatments that they might not otherwise have accepted, some patients could have affected their future eligibility, or could fear this consequence, which might also reduce their coping alternatives and cause psychological distress for persons in high need of services. Again, this is speculation based on an unusual pattern of results. We note that we cannot tie up this possible explanation tightly based on the data for two reasons. First, we did not obtain information on the occurrence of these specific insurance limitations. Second, different findings making up this story, were observed across different specific intervention and cultural groups in the study, rather than consistently across specific groups. Further, owing to data limitations such as not having data on preexisting condition exclusions, we cannot model this specific explanation through a more formal structural analytic approach. We do hope to explore this and other possible explanations for the long-term outcome findings subsequently through an extensive qualitative data set developed as part of this study, which is not yet completed.

Among underserved minorities, we found that both interventions relative to UC reduced the likelihood of respondents thinking they could handle their problem on their own and reduced perceived barriers to care due to racial discrimination. From the perspective of origins of health care disparities (Smedley and Nelson 2003), this is an important outcome in its own right, that is, the interventions helped underserved minorities overcome culturally specific barriers, even though that did not result in reduced unmet need or greater access to treatments at 9 years, perhaps because of greater service availability problems for underserved minority groups. These are also issues for future studies including exploration of the PIC qualitative data.

There are important limitations to these findings, including use of particular health care systems in specific U.S. sites; moderate response rates; studying only certain minority groups; reliance on self-report measures

although interviewers were blinded to intervention status; and limited sample sizes and power/precision for some comparisons, requiring grouping of African Americans and Latinos.

Overall we can conclude that the main, intended intervention effects are over by 9 years, after yielding many years of benefit especially for minorities; yet there is a picture at 9 years suggesting new unintended consequences for particular interventions or cultural groups, ranging from perceived barriers, difficulty coping, increased distress, lower access—but also lowering of attitudinal barriers for minorities. These findings could offer clues for needed longer-term intervention supports or system or policy changes that we hope may emerge from new studies and in-depth examination of qualitative data.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This work was funded by grants from National Institute of Mental Health grants MH061570 and MH068639.

All authors on the paper contributed substantially to the work represented in this paper: Kenneth Wells (PI), Cathy Sherbourne (Co-PI), Jeanne Miranda (investigator), Naihua Duan (Statistician), Lingqi Tang (Statistician), and Bernadette Benjamin (Database Coordinator and Lead Programmer).

Maureen Carney provided project management and Barbara Levitan supervised data collection. We are grateful to project investigators Paul Koegele, Gery Ryan, and David Kennedy, who lead a separate qualitative follow-up study within the project.

There are no conflicts of interests. None of the authors nor the organizations with which the authors are currently affiliated have taken public stands (or a particular advocacy position) relevant to the manuscript. Preliminary results were presented at the Annual NIMH conference during the summer of 2007. NIMH funded this work and we plan to provide the project officer with an advance copy as a courtesy.

Disclosures: None.

REFERENCES

- Araya, R., G. Rojas, R. Fritsch, J. Gaete, M. Rojas, G. Simon, and T. J. Peters. 2003. "Treating Depression in Primary Care in Low-Income Women in Santiago, Chile: A Randomised Controlled Trial." *Lancet* 361 (9362): 995–1000.

- Arean, P. A., L. Ayalon, E. Hunkeler, E. H. Lin, L. Tang, L. Harpole, H. Hendrie, J. W. Williams Jr., and J. Unutzer. 2005. "Improving Depression Care for Older, Minority Patients in Primary Care." *Medical Care* 43 (4): 381-90.
- Asarnow, J. R., L. H. Jaycox, N. Duan, A. P. LaBorde, M. M. Rea, P. Murray, M. Anderson, C. Landon, L. Tang, and K. B. Wells. 2005. "Effectiveness of a Quality Improvement Intervention for Adolescent Depression in Primary Care Clinics: A Randomized Controlled Trial." *Journal of the American Medical Association* 293 (3): 311-9.
- Bell, R. 1999. *Depression PORT Methods Workshop (I)*. Santa Monica, CA: RAND.
- Bruce, M. L., T. R. Ten Have, C. F. Reynolds III, I. I. Katz, H. C. Schulberg, B. H. Mulsant, G. K. Brown, G. J. McAvay, J. L. Pearson, and G. S. Alexopoulos. 2004. "Reducing Suicidal Ideation and Depressive Symptoms in Depressed Older Primary Care Patients: A Randomized Controlled Trial." *Journal of the American Medical Association* 291 (9): 1081-91.
- Ciechanowski, P., E. Wagner, K. Schmalting, S. Schwartz, B. Williams, P. Diehr, J. Kulzer, S. Gray, C. Collier, and J. LoGerfo. 2004. "Community-Integrated Home-Based Depression Treatment in Older Adults: A Randomized Controlled Trial." *Journal of the American Medical Association* 291 (13): 1569-77.
- Depression Guidelines Panel. 1993a. *Depression in Primary Care II: Treatment of Major Depression*. Rockville, MD: U.S. Department of Health and Human Services, U.S. Public Health Service.
- . 1993b. *Depression in Primary Care, I: Detection and Diagnosis*. Rockville, MD: U.S. Department of Health and Human Services, U.S. Public Health Service.
- Dietrich, A. J., T. E. Oxman, J. W. Williams Jr., H. C. Schulberg, M. L. Bruce, P. W. Lee, S. Barry, P. J. Raue, J. J. Lefever, M. Heo, K. Rost, K. Kroenke, M. Gerrity, and P. A. Nutting. 2004. "Re-Engineering Systems for the Treatment of Depression in Primary Care: Cluster Randomised Controlled Trial." *British Medical Journal* 329 (7466): 602.
- Gilbody, S., P. Whitty, J. Grimshaw, and R. Thomas. 2003. "Educational and Organizational Interventions to Improve the Management of Depression in Primary Care: A Systematic Review." *Journal of the American Medical Association* 289 (23): 3145-51.
- Graubard, B. I., and E. L. Korn. 1999. "Predictive Margins with Survey Data." *Biometrics* 55 (2): 652-9.
- Hedrick, S. C., E. F. Chaney, B. Felker, C. F. Liu, N. Hasenberg, P. Heagerty, J. Buchanan, R. Bagala, D. Greenberg, G. Paden, S. D. Fihn, and W. Katon. 2003. "Effectiveness of Collaborative Care Depression Treatment in Veterans' Affairs Primary Care." *Journal of General Internal Medicine* 18 (1): 9-16.
- Hirji, K. F. 1992. "Computing Exact Distributions for Polytomous Response Data." *Journal of the American Statistical Association* 87 (418): 487.
- Hunkeler, E. M., W. Katon, L. Tang, J. W. Williams Jr., K. Kroenke, E. H. Lin, L. H. Harpole, P. Arean, S. Levine, L. M. Grypma, W. A. Hargreaves, and J. Unutzer. 2006. "Long Term Outcomes from the IMPACT Randomised Trial for Depressed Elderly Patients in Primary Care." *British Medical Journal* 332 (7536): 259-63.

- Hunkeler, E. M., J. Meresman, W. A. Hargreaves, B. Fireman, W. H. Berman, A. J. Kirsch, J. Groebe, S. W. Hurt, P. Braden, M. Getzell, P. A. Feigenbaum, T. Peng, and M. Salzer. 2000. "Efficacy of Nurse Telehealth Care and Peer Support in Augmenting Treatment of Depression in Primary Care." *Archives of Family Medicine* 9 (8): 700–8.
- Katon, W., P. Robinson, M. Von Korff, E. Lin, T. Bush, E. Ludman, G. Simon, and E. Walker. 1996. "A Multifaceted Intervention to Improve Treatment of Depression in Primary Care." *Archives of General Psychiatry* 53 (10): 924–32.
- Katon, W., C. Rutter, E. J. Ludman, M. Von Korff, E. Lin, G. Simon, T. Bush, E. Walker, and J. Unutzer. 2001. "A Randomized Trial of Relapse Prevention of Depression in Primary Care." *Archives of General Psychiatry* 58 (3): 241–7.
- Katon, W., and J. Unutzer. 2006. "Pebbles in a Pond: NIMH Grants Stimulate Improvements in Primary Care Treatment of Depression." *General Hospital of Psychiatry* 28 (3): 185–8.
- Katon, W., M. Von Korff, E. Lin, G. Simon, E. Walker, J. Unutzer, T. Bush, J. Russo, and E. Ludman. 1999. "Stepped Collaborative Care for Primary Care Patients with Symptoms of Depression: A Randomized Trial." *Archives of General Psychiatry* 56: 1009–115.
- Katon, W., M. Von Korff, E. H. Lin, G. Simon, E. Ludman, J. Russo, P. Ciechanowski, E. Walker, and T. Bush. 2004. "The Pathways Study: A Randomized Trial of Collaborative Care in Patients with Diabetes and Depression." *Archives of General Psychiatry* 61 (10): 1042–9.
- Katon, W., M. Von Korff, E. Lin, E. Walker, G. E. Simon, T. Bush, P. Robinson, and J. Russo. 1995. "Collaborative Management to Achieve Treatment Guidelines: Impact on Depression in Primary Care." *Journal of the American Medical Association* 273 (13): 1026–31.
- Katzelnick, D. J., M. Von Korff, H. Chung, L. P. Provost, and E. H. Wagner. 2005. "Applying Depression-Specific Change Concepts in a Collaborative Breakthrough Series." *Joint Commission Journal on Quality and Safety* 31 (7): 386–97.
- Lavori, P. W., R. Dawson, and D. Shera. 1995. "A Multiple Imputation Strategy for Clinical Trials with Truncation of Patient Data." *Statistics in Medicine* 14 (17): 1913–25.
- Little, R. 1988. "Missing-Data Adjustments in Large Surveys." *Journal of Business and Economic Statistics* 6 (3): 287–96.
- Manly, B. 1997. *Randomization, Bootstrap and Monte Carlo Methods in Biology*. London: Chapman & Hall.
- McCaffrey, D. F., and R. M. Bell. 2006. "Improved Hypothesis Testing for Coefficients in Generalized Estimating Equations with Small Samples of Clusters." *Statistics in Medicine* 25: 4081–98.
- Miranda, J., J. Y. Chung, B. L. Green, J. Krupnick, J. Siddique, D. A. Revicki, and T. Belin. 2003b. "Treating Depression in Predominantly Low-Income Young Minority Women: A Randomized Controlled Trial." *Journal of the American Medical Association* 290 (1): 57–65.
- Muñoz, R. J., S. Aguilar-Gaxiola, and J. Guzmán. 2000. *Manual de Terapia de Grupo para el Tratamiento Cognitivo-conductual de Depresión, Hospital General de San Francisco, Clinica de Depresión, 1986*. Santa Monica, CA: RAND.

- Muñoz, R. J., and J. Miranda. 2000. *Group Therapy for Cognitive Behavioral Treatment of Depression, San Francisco General Hospital Clinic 1986. Document MR01198/4*. Santa Monica, CA: RAND.
- Neumeier-Gromen, A., T. Lampert, K. Stark, and G. Kallischnigg. 2004. "Disease Management Programs for Depression: A Systematic Review and Meta-Analysis of Randomized Controlled Trials." *Medical Care* 42 (12): 1211–21.
- Orlando, M., and L.S. Meredith. 2002. "Understanding the Causal Relationship between Patient-Reported Interpersonal and Technical Quality of Care for Depression." *Medical Care* 40 (8): 696–704.
- Rost, K. M., N. Duan, L. V. Rubenstein, D. E. Ford, C. D. Sherbourne, L. S. Meredith, and K. B. Wells. 2001. "The Quality Improvement for Depression Collaboration: General Analytic Strategies for a Coordinated Study of Quality Improvement in Depression Care." *General Hospital of Psychiatry* 23 (5): 239–53.
- Rost, K. M., P. Nuttin, J. L. Smith, C. E. Elliott, and M. Dickinson. 2002. "Managing Depression as a Chronic Disease: A Randomized Trial of Ongoing Treatment in Primary Care." *British Medical Journal* 325: 934–37.
- Schafer, J. 1997. *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.
- Schoenbaum, M., J. Unutzer, D. McCaffrey, N. Duan, C. Sherbourne, and K. B. Wells. 2002. "The Effects of Primary Care Depression Treatment on Patients' Clinical Status and Employment." *Health Services Research* 37 (5): 1145–58.
- Schoenbaum, M., J. Unutzer, C. Sherbourne, N. Duan, L. V. Rubenstein, J. Miranda, L. S. Meredith, M. F. Carney, and K. Wells. 2001. "Cost-Effectiveness of Practice-Initiated Quality Improvement for Depression: Results of a Randomized Controlled Trial." *Journal of the American Medical Association* 286 (11): 1325–30.
- Shen, Z. 2000. "Nested Multiple Imputation." Ph.D. dissertation. Department of Statistics. Cambridge, MA: Harvard University.
- Sherbourne, C., M. Edelen, A. Zhou, C. Bird, N. Duan, and K. Wells. 2008. "How Therapy-Based Quality Improvement Intervention for Depression Affected Life Events and Psychological Well-Being over Time: A Nine-Year Longitudinal Analysis." *Medical Care* 46 (1): 78–84.
- Sherbourne, C. D., K. B. Wells, N. Duan, J. Miranda, J. Unutzer, L. Jaycox, M. Schoenbaum, L. S. Meredith, and L. V. Rubenstein. 2001. "Long-Term Effectiveness of Disseminating Quality Improvement for Depression in Primary Care." *Archives of General Psychiatry* 58 (7): 696–703.
- Simon, G. E., W. J. Katon, M. Von Korff, J. Unutzer, E. H. Lin, E. A. Walker, T. Bush, C. Rutter, and E. Ludman. 2001. "Cost-Effectiveness of a Collaborative Care Program for Primary Care Patients with Persistent Depression." *American Journal of Psychiatry* 158 (10): 1638–44.
- Simon, G. E., M. Von Korff, C. Rutter, and E. Wagner. 2001. "Randomised Trial of Monitoring, Feedback, and Management of Care by Telephone to Improve Treatment of Depression in Primary Care." *British Medical Journal* 320: 550–4.
- Smedley, B. D., A. Y. Stith, and A. R. Nelson (eds.) 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: Institute of Medicine, National Academy Press.

- Tang, L., N. Duan, R. Klap, T. Belin. 2007. "Contrasting Imputation Controlling for Intermediate Outcome Variables versus Weighting Using Baseline Covariates to Correct for Nonresponse Bias in a Longitudinal Study." In *2007 JSM Proceedings, Statistical Computing Section (CD-ROM)*, edited by A. S. Association. Alexandria, VA: American Statistical Association.
- Tang, L., J. Song, T. R. Belin, and J. Unutzer. 2005. "A Comparison of Imputation Methods in a Longitudinal Randomized Clinical Trial." *Statistics in Medicine* 24 (14): 2111–28.
- Unutzer, J., W. Katon, C. M. Callahan, J. W. Williams Jr., E. Hunkeler, L. Harpole, M. Hoffing, R. D. Della Penna, P. H. Noel, E. H. Lin, P. A. Arean, M. T. Hegel, L. Tang, T. R. Belin, S. Oishi, and C. Langston. 2002. "Collaborative Care Management of Late-Life Depression in the Primary Care Setting: A Randomized Controlled Trial." *Journal of the American Medical Association* 288 (22): 2836–45.
- Unützer, J., L. Rubenstein, W. J. Katon, L. Tang, N. Duan, I. T. Lagomasino, and K. B. Wells. 2001. "Two-Year Effects of Quality Improvement Programs on Medication Management for Depression." *Archives of General Psychiatry* 58 (10): 935–42.
- Wang, P. S., G. E. Simon, J. Avorn, F. Azocar, E. J. Ludman, J. McCulloch, M. Z. Petukhova, and R. C. Kessler. 2007. "Telephone Screening, Outreach, and Care Management for Depressed Workers and Impact On Clinical and Work Productivity Outcomes: A Randomized Controlled Trial." *Journal of the American Medical Association* 298 (12): 1401–11.
- Ware Jr., J. E., M. Kosinski, and S. Keller. 1995. *SF-12: How to Score the SF-12 Physical and Mental Health Summary Scales*. Boston: The Health Institute, New England Medical Center.
- Ware Jr., J. E., and C. D. Sherbourne. 1992. "The MOS 36-Item Short-Form Health Survey (SF-36). I. Conceptual Framework and Item Selection." *Medical Care* 30 (6): 473–83.
- Wells, K., C. Sherbourne, J. Miranda, L. Tang, B. Benjamin, and N. Duan. 2007. "The Cumulative Effects of Quality Improvement for Depression on Outcome Disparities over 9 Years: Results from a Randomized, Controlled Group-Level Trial." *Medical Care* 45 (11): 1052–59.
- Wells, K. B. 1999. "The Design of Partners in Care: Evaluating the Cost-Effectiveness of Improving Care for Depression in Primary Care." *Social Psychiatry and Psychiatric Epidemiology* 34 (1): 20–9.
- Wells, K. B., C. Sherbourne, M. Schoenbaum, N. Duan, L. Meredith, J. Unutzer, J. Miranda, M. F. Carney, and L. V. Rubenstein. 2000. "Impact of Disseminating Quality Improvement Programs for Depression in Managed Primary Care: A Randomized Controlled Trial." *Journal of the American Medical Association* 283 (2): 212–20.
- Wells, K. B., C. D. Sherbourne, M. Schoenbaum, S. Ettner, N. Duan, J. Miranda, J. Unutzer, and L. V. Rubenstein. 2004. "Five-Year Impact of Quality Improvement for Depression: Results of a Group-Level Randomized Controlled Trial." *Archives of General Psychiatry* 61 (4): 378–86.

World Health Organization. 1995. *Composite International Diagnostic Interview (CIDI), Version 2.1*. Geneva, Switzerland: World Health Organization.

SUPPLEMENTARY MATERIAL

The following supplementary material for this article is available online:
Appendix AS1. Author matrix.

This material is available as part of the online article from <http://www.blackwell-synergy.com/doi/abs/10.1111/j.1475-6773.2008.00871.x>
(this link will take you to the article abstract).

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.