Quality Improvement With Pay-for-Performance Incentives in Integrated Behavioral Health Care

| Jürgen Unützer, MD, MPH, MA, Ya-Fen Chan, PhD, Erin Hafer, MPH, Jessica Knaster, MPH, Anne Shields, RN, MPH, Diane Powers, MA, Richard C. Veith, MD

Behavioral health problems are among the most common and disabling health conditions worldwide.[1] They often co-occur with chronic medical diseases and can substantially worsen associated health outcomes.[1] When these problems are not effectively treated, they can impair self-care and adherence to medical and mental health treatments and are associated with increased mortality and increased overall health care costs.[2]

National surveys have consistently demonstrated that more Americans receive mental health care from primary care providers than from mental health specialists.[3] Most patients prefer an integrated approach in which primary care and mental health providers work together to address medical and behavioral health needs. In reality, however, medical, mental health, and substance abuse services are fragmented and delivered in separate "silos" with little to no effective collaboration. In a recent survey, two thirds of primary care providers (PCPs) reported that they could not access effective mental health services for their patients.[4]

Currently, the most robust research evidence for improving mental health outcomes in primary care comes from studies of collaborative care programs for common mental disorders, such as depression.[5] In such programs, PCPs are part of a collaborative care team that may include nurses, clinical social workers, psychologists, and psychiatrists who can support medication management prescribed by PCPs and provide evidence-based mental health treatments in primary care. Core components of successful programs include the concepts of measurement-based care and stepped care in which treatments are systematically changed or intensified if patients do not show substantial improvement in target clinical outcomes.[6] In the largest trial of collaborative care to date, participants were more than twice as likely as those in usual care to

Objectives. We evaluated a quality improvement program with a pay-for-performance (P4P) incentive in a population-focused, integrated care program for safety-net patients in 29 community health clinics.

Methods. We used a quasi-experimental design with 1673 depressed adults before and 6304 adults after the implementation of the P4P program. Survival analyses examined the time to improvement in depression before and after implementation of the P4P program, with adjustments for patient characteristics and clustering by health care organization.

Results. Program participants had high levels of depression, other psychiatric and substance abuse problems, and social adversity. After implementation of the P4P incentive program, participants were more likely to experience timely follow-up, and the time to depression improvement was significantly reduced. The hazard ratio for achieving treatment response was 1.73 (95% confidence interval = 1.39, 2.14) after the P4P program implementation compared with pre-program implementation.

Conclusions. Although this quasi-experiment cannot prove that the P4P initiative directly caused improved patient outcomes, our analyses strongly suggest that when key quality indicators are tracked and a substantial portion of payment is tied to such quality indicators, the effectiveness of care for safety-net populations can be substantially improved. (*Am J Public Health.* 2012;102: e41–e45. doi:10.2105/AJPH.2011.300555)

experience a substantial improvement in their depression over 12 months.[7] They also had less physical pain,[8] better social and physical functioning, and better overall quality of life than did patients in usual care settings.[9]

Although there was compelling research evidence supporting collaborative care for depression in primary care by the year 2000,[5] large-scale implementations of this approach have only started to emerge over the past few years, and there remain few reports of the effectiveness of such programs when implemented outside of research trials. Examples include the Depression Improvement Across Minnesota, Offering a New Direction (DIAMOND) program in Minnesota[10] and the Washington State Mental Health Integration Program (MHIP; http://integratedcare-nw.org) in which over 100 community health clinics and 30 community mental health centers partner to provide integrated care for safety-net patients with medical and behavioral health needs.

Funded by the State of Washington and administered by the Community Health Plan of Washington (CHPW), a nonprofit managed care plan, in collaboration with Public Health—Seattle & King County, MHIP provides medical and mental health services for low-income adults who are temporarily disabled due to a physical or mental health condition and expected to be unemployed for at least 90 days (adults covered in the State of Washington's Disability Lifeline Program), veterans and family members of veterans, the uninsured, low-income mothers and their children, and low-income older adults. Behavioral health care is provided in the primary care clinic through a collaborative approach including a PCP and a care coordinator, a consulting psychiatrist assigned to each of the primary care–based teams, and other behavioral health providers, if available. Each care coordinator receives weekly caseload consultation with a consulting psychiatrist to review cases and develop a treatment plan, which might include

medication recommendations, psychosocial support and brief psychotherapeutic interventions by the care coordinator, and referrals to other services that are clinically indicated (e.g., substance abuse counseling). Patients who are too challenging to be cared for in primary care are referred to a partnering community mental health center for additional treatment.

MHIP was initiated in 29 community health clinics in the 2 most populous counties in Washington State representing the metropolitan Seattle–Tacoma area in late 2007. In 2010, the program was expanded to over 100 community health clinics and 30 community mental health centers statewide. Expert faculty from the AIMS (Advancing Integrated Mental Health Solutions) Center at the University of Washington provided training, technical assistance, and a web-based tracking system[11] to help support systematic outcome tracking and quality improvement. All program participants are tracked in this registry, which captures clinical diagnoses assigned by clinicians working with patients and clinical outcomes using validated clinical rating scales such as the PHQ-9 (Patient Health Questionnaire) for depression.[12] This information is gathered for all participants at an initial assessment and at each subsequent contact with a care coordinator.

Initial experience with MHIP suggested substantial variation in the quality and outcomes of care provided across the participating community health clinics. To reduce this variation and improve the overall effectiveness of the program, the program sponsors instituted a quality improvement program with a pay-for-performance (P4P) incentive. Before 2009, participating clinics received full payment for the cost of the care coordinators deployed in the participating primary care clinics. Outcomes were monitored by MHIP staff, and technical assistance was provided to support struggling sites, but no financial incentives were tied to performance. After the P4P incentive program went into effect on January 1, 2009, 25% of the annual program funding to participating clinics was contingent on meeting several quality indicators, including timely follow-up of patients in the program (2 or more contacts per month for at least half of the active caseload), psychiatric consultation for patients who do not show clinical improvement, and regular tracking of psychotropic medications

used. Participating clinics and providers received regular feedback on their quality indicators through the web-based clinical tracking system and training and technical assistance to help improvement on these indicators through an all-day in-person training workshop for care coordinators (http://chpw.org/gau) and monthly webinars provided by the University of Washington AIMS Center.

There is very limited experience with P4P incentives in behavioral health care,[13] and we know of no published studies of such incentives in the context of population-focused, primary care–based collaborative care programs. In this article, we take advantage of this real-world experiment and examine changes in quality of care and patient outcomes observed among MHIP participants before and after implementation of the P4P incentive program.

METHODS

Our analytical sample includes all 7941 MHIP participants with a PHQ-9 depression score of 10 or greater, indicating clinically significant depression, who were served in one of the 29 clinics participating in the metropolitan Seattle–Tacoma area between January 2008 and December 2010. All participants were assigned to 1 of 2 groups: individuals who entered the program prior to January 1, 2009, the date that the P4P incentive program was initiated (n = 1673), and those who enrolled after the initiation of the incentive program (n = 6304).

Data on demographic and clinical characteristics of participants, processes of care, and clinical outcomes were obtained from the web-based registry that is used by all MHIP clinicians to track and coordinate care. This web-based tracking tool was originally developed for a large randomized-controlled trial of collaborative care for depression[11] and has subsequently been used to track patient outcomes in several other large treatment trials[14] and quality improvement initiatives.[10]

Quality of care was assessed by whether follow-up contact was initiated within 2 and 4 weeks after the participant's initial assessment for the program and whether participants had a psychiatric consultation if they were not improving. Improvement in depression severity was defined as achieving a 50% reduction

from the baseline score or a score of less than 10 on the PHQ-9.[12] Our analyses included the following covariates: age, gender, unstable housing, elevated risk for suicide, and comorbid psychiatric diagnoses listed by treating providers (anxiety, bipolar disorder, posttraumatic stress disorder, substance abuse, or cognitive disorder). Participants with a reported history of suicide attempt or current suicidal ideation on PHQ item 9 were considered as having an elevated risk of suicide. Unstable housing was identified at baseline assessment if participants reported needing assistance with housing, being homeless, or being referred to housing support services.

Sample characteristics, clinical presentation, quality and outcomes of care in the pre- and post-P4P groups were examined using the χ^2 test for dichotomous variables and t-test for continuous variables. We used the Kaplan–Meier method to estimate the cumulative probability of achieving a 50% or greater reduction or a score less than 10 in the PHQ-9 before and after initiation of the P4P incentive program. Participants were censored on the date of last observation or disenrollment if they did not achieve the aforementioned depression improvement. We further conducted a Cox proportional hazard model to estimate hazard ratio for depression improvement with adjustment for covariate effects. The Cox model was conducted using clustered robust variance to take into account the nesting of patients within participating health care organizations. The proportional hazards assumption was examined with specifying variables that vary with respect to time. Any violation of the assumption was handled by incorporating an interaction of covariate and log time to the model. Data were analyzed using Stata Version 11 (College Station, TX).

All analyses were conducted on de-identified data collected for quality improvement purposes and were not considered research requiring individual patient consent by the University of Washington's institutional review board. Only aggregate data are presented.

RESULTS

All 7941 MHIP participants with clinically significant depression symptoms (PHQ-9 score of 10 or greater) were included in our analyses

(Table 1). The mean age of participants was 41.3 years (SD = 11.9 years) and approximately half of the program participants were men. The mean depression severity on the PHQ-9 was 18.1 (SD = 4.8), indicating moderate to severe depression. More than half of the participants (57.7%) reported thoughts of death or suicide at baseline and more than half (56.7%) reported problems with stable housing. In addition to depression, additional clinical diagnoses reported by treating providers included anxiety (63.5%), bipolar disorder (17.1%), posttraumatic stress disorder (21.4%), and substance abuse disorders (20.5%). Participants with comorbid psychotic (2.9%) or cognitive disorders (2.2%) were less common.

A total of 1673 MHIP participants with depression enrolled in the program before 2009, and 6304 participants enrolled after the P4P initiative went into effect. Participants in both groups were similar in demographic and clinical characteristics (Table 1). Depressed MHIP participants enrolled post–P4P implementation were slightly younger (mean [SD] = 41.1 years [12.2 years] versus 41.9 years [11.0 years]; $P = .014$) and slightly more likely to be women (52% versus 48%; $P = .004$) than were those enrolled before the P4P incentive was implemented. Post–P4P implementation participants were somewhat more likely to have comorbid psychiatric disorders, such as anxiety (67% versus 51%; $P < .001$), posttraumatic stress disorder (23% versus 16%; $P < .001$), or a cognitive disorder (2.4% versus 1.4%; $P = .012$). We adjusted for these demographic and clinical differences in subsequent analyses.

Pre–P4P implementation participants were significantly less likely to have early follow-up (within 2 or 4 weeks of the initial assessment) than were those treated after the P4P program was instituted in 2009 (Table 2). For example, 72% of those seen post–P4P implementation had follow-up with a care coordinator within 4 weeks compared with only 53% of those seen before implementation ($P < .001$). On the other hand, the total number of follow-up contacts with a care coordinator over the course of the program actually decreased somewhat from 6.2 (SD = 8.6) pre–P4P implementation to 5.5 (SD = 6.8) post–P4P implementation ($P = .002$). The proportion of program participants who had a psychiatrist review their case and make recommendations for treatment to the patient's PCP increased from 49% pre–P4P implementation to 60% post–P4P implementation ($P < .001$).

Figure 1 illustrates Kaplan–Meier survival curves examining the time elapsed until participants achieve the desired clinical improvement before and after the P4P-based quality improvement instituted in 2009. The findings show that the rate of achieving a 50% or greater reduction or a score of less than 10 on the PHQ-9 was significantly higher after the introduction of the P4P program (log-rank test; $P < .001$). Additionally, analyses show that the median time elapsed for reaching this improvement benchmark in depression was reduced from approximately 64 weeks pre–P4P implementation to 25 weeks postimplementation. After adjusting for demographic and clinical differences, we found that participants enrolled post–P4P implementation had a 1.73-fold increased likelihood of achieving either a 50% or greater reduction from the baseline or a PHQ-9 score less than 10 than did participants who enrolled before the P4P program was instituted (95% confidence interval = 1.39, 2.14; $P < .001$).

DISCUSSION

To our knowledge, this is the first data-based report of a P4P initiative in the context of a large-scale collaborative care program for primary care patients with depression and other common mental health disorders. Patients served in this program had moderate to severe depression and a high degree of clinical complexity such as high rates of suicidal ideation and psychiatric comorbidities. They also faced substantial social challenges. The majority of program participants were unemployed

TABLE 1—Baseline Sample Demographic and Clinical Characteristics: Washington State Mental Health Integration Program, January 2008–December 2010

		Year of Enrollment		
Variables	Total (n = 7941), Mean ± SD or No. (%)	Before 2009 (n = 1637), Mean ± SD or No. (%)	After 2009 (n = 6304), Mean ± SD or No. (%)	P
Age, y	41.3 ± 11.9	41.9 ± 11.0	41.1 ± 12.2	.014
Gender[a]				.004
Male	3853 (48.5)	846 (51.7)	3007 (47.7)	
Female	4087 (51.5)	791 (48.3)	3296 (52.3)	
Suicidality[b]				.623
No	3309 (42.3)	689 (42.9)	2620 (42.2)	
Yes	4512 (57.7)	919 (57.1)	3593 (57.8)	
Unstable housing[c]				.967
No	2365 (43.3)	444 (43.3)	1921 (43.3)	
Yes	3102 (56.7)	581 (56.7)	2521 (56.7)	
PHQ-9	18.1 ± 4.8	18.0 ± 4.7	18.1 ± 4.8	.571
Comorbidities				
Anxiety	5044 (63.5)	829 (50.6)	4215 (66.9)	< .001
Bipolar	1357 (17.1)	298 (18.2)	1059 (16.8)	.178
Psychotic disorder	229 (2.9)	51 (3.1)	178 (2.8)	.53
PTSD	1698 (21.4)	255 (15.6)	1443 (22.9)	< .001
Substance abuse	1624 (20.5)	334 (20.4)	1290 (20.5)	.957
Cognitive disorder	176 (2.2)	23 (1.4)	153 (2.4)	.012
Chronic pain	507 (6.4)	116 (7.1)	391 (6.2)	.193

Note. PHQ = Patient Health Questionnaire; PTSD = posttraumatic stress disorder.
[a] 1 patient did not provide information on gender.
[b] 120 (1.5%) patients did not provide information on suicidality.
[c] 2474 (31%) patients did not provide information on housing.

TABLE 2—Quality of Care Outcomes: Washington State Mental Health Integration Program, January 2008–December 2010
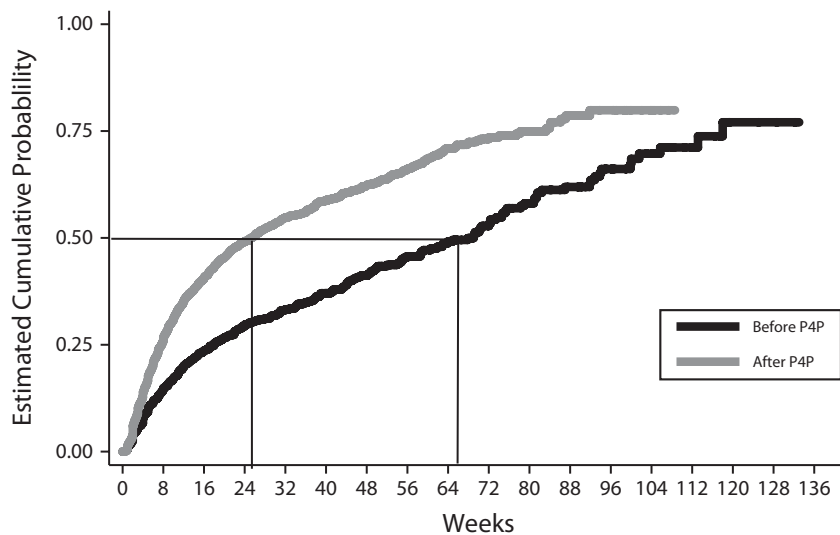
Variables	Total (n = 7941), % or Mean ±SD	Year of Enrollment		P
		Before 2009 (n = 1637), % or Mean ±SD	After 2009 (n = 6304), % or Mean ±SD	
Any follow-up contacts/attempts within 2 wk after initial assessment	55.9	42.4	59.3	< .001
Any follow-up contacts/attempts within 4 wk after initial assessment	68.0	52.6	71.8	< .001
No. of follow-up contacts/attempts in first 4 wk after initial assessment	1.33 ±1.29	0.97 ±1.19	1.42 ±1.30	< .001
Total number of follow-up contacts during treatment[a]	5.70 ±7.27	6.17 ±8.63	5.54 ±6.76	.002
Any psychiatric consultation during the treatment	57.6	49.4	59.8	< .001

[a]Subset of patients who were discharged from the program (n = 6609).

due to a medical or a mental health–related disability, and more than half of program participants had problems with stable housing. Despite these formidable challenges, we found that the majority of program participants achieved improvements in their depression during their participation in the program.

Our analyses also suggest that the institution of a quality improvement program with a P4P incentive substantially improved the quality and outcomes of care provided by the program. After the institution of the P4P incentive program, participants were substantially more likely to experience a significant improvement in depression severity, and the time to improvement was dramatically reduced compared with before the P4P incentive was implemented. These improvements in clinical outcomes were consistent with improvements observed in the quality of care that were the intended aims of the P4P initiative, such as early follow-up and psychiatric consultation for patients who were not improving.

Earlier research on use of P4P incentives in behavioral health care concluded that such programs are not "magic bullets" and may require substantial investments in and commitment to quality infrastructure, in particular the ability to track systematically the quality and outcomes of care provided.[13] Our experience is entirely consistent with this conclusion. We believe that systematic collection of key quality parameters (e.g., timely follow-up) and clinical outcomes (e.g., PHQ-9 scores for depression) during the course of treatment is required to identify patients who are not improving as expected and to change treatments as needed. This approach, also called measurement-based stepped care,[6] has been identified as a core component of effective quality improvement programs not only for depression but also for other common medical conditions.[15] At a program level, the timely availability of data on quality and outcomes of care enables the implementation of meaningful and effective P4P incentive programs.

Our analyses have several important strengths. We used data from a large, real-world sample of safety-net patients served in 29 community health centers. We included all program participants with clinically significant depression, resulting in a diverse and representative sample of patients, clinics, and providers. Because MHIP routinely tracks key process indicators and clinical outcomes using a web-based tracking system, we were able to overcome one of the key limitations of earlier work on P4P in behavioral health care: the lack of systematic data on quality and outcomes of care.[13]

Our study also has several important limitations. Our data are from a natural experiment without a contemporaneous or randomly assigned control group, and we cannot prove that the quality improvement program and the associated P4P incentive that focused on improving specific quality indicators was causally related to the improved outcomes observed. We are, however, not aware of other systematic changes in the program or the participating clinics that might have accounted for the improvements observed after 2009. Program participants after 2009 had

Note. P4P = pay-for-performance.

FIGURE 1—Kaplan–Meier survival curve for time to the first improvement in depression before and after P4P-based quality improvement: Washington State Mental Health Integration Program, January 2008–December 2010.

somewhat higher rates of psychiatric comorbidity, but we controlled for such patient characteristics in our analyses and do not believe that the differences in outcomes can be explained by observed characteristics of the patients served. The P4P incentive program was implemented in the context of a larger quality improvement effort in which participating clinics were offered ongoing training of care management staff, and it is not possible to attribute program improvements to any 1 particular component of the overall quality improvement effort. Given these limitations, our findings about improvements in quality and outcomes of care after initiation of the P4P program should be tested using a more rigorous design such as a randomized-controlled experiment. Other limitations include the fact that we are using clinical rating scales completed by patients and working diagnoses assigned by providers in the context of their routine clinical work and not independent research diagnostic assessments or assessments of depression severity. Data from the largest primary care–based treatment trial for depression to date suggest, however, that PHQ-9 scores collected by care managers during routine clinical work are closely correlated with other depression severity measures administered by researchers.[16]

Health care reform, mental health parity, and the emergence of the patient-centered health care home provide important opportunities for the integration of mental health and primary care. The proposed expansion of Medicaid will create a need for mental health services for millions of adults who are primarily served in the nation's primary health care systems. After 20 years of building a robust research evidence base for integrated mental health care,[5] the time has now come for payers to provide the right incentives and tools for organizations to implement evidence-based programs that can serve large populations of patients with common behavioral health problems. Our experience with MHIP is consistent with earlier randomized-controlled trials of collaborative care for low income populations with depression[17–19] and suggests that evidence-based integrated care programs in community health centers can be effective in treating low-income populations with complex mixtures of depression, chronic medical problems, and social stressors. When clinical outcomes and key quality indicators are routinely tracked and a substantial portion of the payment for care is tied to quality indicators such as adequate follow-up and consultation for patients who are not improving, the quality and effectiveness of such programs can be substantially improved. ∎

About the Authors

Jürgen Unützer, Ya-Fen Chan, Diane Powers, and Richard C. Veith are with the Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle. Erin Hafer is with the Community Health Plan of Washington, Seattle. Jessica Knaster is with Public Health of Seattle and King County, Seattle, WA. Anne Shields is with the Washington State Department of Health, Olympia, WA.

Correspondence should be sent to Jürgen Unützer, MD, MPH, MA, Professor and Vice Chair, Psychiatry and Behavioral Sciences, University of Washington, 1959 NE Pacific Street, Box 356560, Seattle, WA 98195-6560 (e-mail: unutzer@uw.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This article was accepted October 27, 2011.

Contributors

J. Unützer and Y.-F. Chen contributed to the study design, data collection, and analysis. E. Hafer contributed to the study design and data collection. J. Knaster, A. Shields, and D. Powers contributed to data collection. R. C. Veith contributed to study design. All authors contributed to writing and editing the article.

Human Participant Protection

All analyses were conducted on de-identified data collected for quality improvement purposes and were not considered research requiring individual patient consent by the University of Washington's institutional review board. Only aggregate data are presented.

References

1. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. *Lancet.* 2007;370(9590):851–858.

2. Katon W, Ciechanowski P. Impact of major depression on chronic medical illness. *J Psychosom Res.* 2002;53(4):859–863.

3. Wang PS, Demler O, Olfson M, Pincus HA, Wells KB, Kessler RC. Changing profiles of service sectors used for mental health care in the United States. *Am J Psychiatry.* 2006;163(7):1187–1198.

4. Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. *Health Aff.* 2009;28(3):w490–501.

5. Gilbody S, Bower P, Fletcher J, Richards D, Sutton AJ. Collaborative care for depression: A cumulative meta-analysis and review of longer-term outcomes. *Arch Intern Med.* 2006;166:2314–2321.

6. Katon W, Unutzer J, Wells K, Jones L. Collaborative depression care: history, evolution and ways to enhance dissemination and sustainability. *Gen Hosp Psychiatry.* 2010;32(5):456–464.

7. Unutzer J, Katon W, Callahan CM, et al. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA.* 2002;288(22):2836–2845.

8. Lin EH, Katon W, Von Korff M, et al. Effect of improving depression care on pain and functional outcomes among older adults with arthritis: a randomized controlled trial. *JAMA.* 2003;290(18):2428–2434.

9. Hunkeler EM, Katon W, Tang L, et al. Long term outcomes from the IMPACT randomised trial for depressed elderly patients in primary care. *BMJ.* 2006;332 (7536):259–263.

10. Korsen N, Pietruszewski P. Translating evidence to practice: two stories from the field. *J Clin Psychol Med Settings.* 2009;16(1):47–57.

11. Unützer J, Choi Y, Cook IA, Oishi S. A web-based data management system to improve care for depression in a multicenter clinical trial. *Psychiatr Serv.* 2002;53 (6):671–673, 678.

12. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* 2001;16(9):606–613.

13. Bremer RW, Scholle SH, Keyser D, Knox Houtsinger JV, Pincus HA. Pay for performance in behavioral health. *Psychiatr Serv.* 2008;59(12):1419–1429.

14. Roy-Byrne P, Craske MG, Sullivan G, et al. Delivery of evidence-based treatment for multiple anxiety disorders in primary care: a randomized controlled trial. *JAMA.* 2010;303(19):1921–1928.

15. Katon WJ, Lin EHB, Von Korff M, et al. Collaborative care for patients with depression and chronic illnesses. *N Engl J Med.* 2010;363(27):2611–2620.

16. Lowe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care.* 2004;42(12):1194–1201.

17. Miranda J, Duan N, Sherbourne C, et al. Improving care for minorities: can quality improvement interventions improve care and outcomes for depressed minorities? Results of a randomized, controlled trial. *Health Serv Res.* 2003;38(2):613–630.

18. Ell K, Xie B, Quon B, Quinn DI, Dwight-Johnson M, Lee PJ. Randomized controlled trial of collaborative care management of depression among low-income patients with cancer. *J Clin Oncol.* 2008;26(27):4488–4496.

19. Ell K, Katon W, Xie B, et al. Collaborative care management of major depression among low-income, predominantly Hispanic subjects with diabetes: a randomized controlled trial. *Diabetes Care.* 2010;33(4):706–713.