



PROJECT MUSE[®]

Improving Care for Depression: Performance Measures, Outcomes and Insights from the Health Disparities Collaboratives

Steven Cole, MD
Kathy Reims, MD
Liz Kershner, MSW, LISW
Harriet G. McCombs, PhD
Kevin Little, PhD
Daniel E. Ford, MD, MPH

Abstract: This paper reports 10 measures, outcomes, and insights from HRSA Depression Health Disparities Collaboratives, representing attempts to accelerate evidence-based guidelines into practice. The authors analyze interviews with leadership of high-performing centers. Monthly data was submitted on 38,000 patients from 94 centers. Regression analyses were conducted to identify process measures predictive of better outcomes. Results indicated that these 10 measures of care were effective in guiding and quantifying improved outcomes. One measure, early and sustained response (ESR), proved particularly useful as it reflects long term outcomes. Regression analyses identified one process measure (Patient Health Questionnaire Reassessment) strongly associated with improved clinical outcomes ($n=37$, $R^2=44\%$). Interviews identified 18 process changes deemed pivotal for meaningful change. In sum, well-designed approaches utilizing proven improvement methodologies resulted in substantial enhancements in depression care. This approach and these measures, especially ESR and PHQ Reassessment, may improve depression care in other under-served settings.

Key words: Depression, depression in primary care, chronic care, primary care.

Major depressive disorder is the leading cause of disability in the U.S. for ages 15–44¹ years and affects approximately 14.8 million American adults (or about 6.7% of the U.S. population age 18 and older) in a given year.² Rates of depression are higher among low-income people than among those with higher incomes, and ethnic

STEVEN COLE is affiliated with the Stony Brook Medical Center and is a member of the faculty and Co-Chair of the Depression Health Disparities Collaborative (DHDC). *KATHY REIMS* is affiliated with CSI Solutions, LLC and the University of Colorado Health Sciences Center, and is a member of the faculty and a participant in the DHDC. *LIZ KERSHNER* is affiliated with CareSouth Carolina and is a member of the faculty and a participant in the DHDC. *HARRIET MCCOMBS* is affiliated with the Bureau of Primary Health Care and is a member of the faculty of the DHDC. *KEVIN LITTLE* is affiliated with Informing Ecological Design, LLC and is a member of the faculty and Improvement Advisor for DHDC. *DANIEL FORD* is affiliated with Johns Hopkins University School of Medicine and is a member of the faculty and Co-Chair of the DHDC. Inquiries may be sent to Steven Cole, MD at stevecolem@gmail.com.

minorities are disproportionately affected.³ Most patients with depression never see a mental health professional but many will seek primary care. Although physical health concerns are usually addressed, substantial numbers of primary care patients remain undiagnosed or under-treated for depression, reflecting the attitudes and practices of physicians, patients, families, and health care systems.⁴ A systematic review of 36 studies found that non-psychiatric physicians failed to make the diagnosis of depression in over one-half of patients seen nationally, and that less than one-fourth of people experiencing an episode of depression during a 12-month period received appropriate treatment.⁵ Because treatments for mental disorders are becoming more effective and more acceptable to patients, the case for screening and early recognition of these disorders is strengthened, particularly in the primary care setting.⁶

Based on the known incidence and prevalence of depression in medically underserved populations, many of the 20 million patients served by health centers nationwide have developed or will develop depression. The Health Resources and Services Administration (HRSA) recognized the importance of providing support, technical assistance and national infrastructure to improve care provided for patients of health centers and began a series of Health Disparities Collaboratives (HDCs) in 1999. Following a successful depression pilot sponsored by the Institute for Healthcare Improvement (IHI), HRSA sponsored three additional HDCs focused on depression care that began in 2002, 2003, and 2005.

The Depression Health Disparities Collaboratives (DHDC) were designed as a systematic approach to improve depression care; organizations and providers tested and measured practice innovations, then shared their experiences in an effort to accelerate learning and widespread implementation of successful change concepts that improved outcomes for patients with depression.

Katzelnick, Rubenstein, and Meredith described the initial DHDC in three earlier publications.^{7,8,9} This paper discusses the development of a set of ten depression measures leveraging the DHDC experience over the subsequent three years, the evolution of the DHDC quality improvement framework, and the development of best practices that correlate with improved care for depression in primary care settings.

Methods

The premise of our quality improvement work was that measures could be developed and used within health centers to improve the process and outcomes of care for patients with depression. Our goal was to develop measures that could be defined and applied consistently within and between health centers and sensitive enough to detect improvement over time, center by center. Through an iterative process, 10 measures were developed to measure processes and outcomes of depression care. These measures were reported monthly by health center teams. Teams from 94 health centers participated in the Depression Collaborative and reported on the results of core and optional measures. We used baseline data reported in 2006 from teams that represented all states except three (Delaware, Indiana, and Louisiana). Teams were nearly equally representative of urban (45%) and rural (55%) areas of the country. Also in 2006, we conducted semi-structured interviews with seven health centers identified as

high-performing teams based on a registry size of over 150 patients and 25% or more of patients showing a 50% reduction of symptoms. These interviews were conducted with health center medical and administrative leaders to determine the pivotal quality improvement processes related to achieving their high performance on the depression measures. In 2007, HRSA health centers were no longer required to report on the depression performance measures; however a mechanism was left in place for teams that wanted to continue submitting monthly reports. We used follow-up data reported in 2008 from 35 health centers including five high-performing teams that continued to report on the 10 performance measures. The goal of the analysis of the health center reports was to identify measures of positive responses to treatment and factors associated to improvement.

DHDC design. The DHDCs were designed to accelerate the translation of evidence-based care for depression into practice. Three important and well studied models were included: a learning system known to accelerate adult learning (Breakthrough Series methodology¹⁰), a rapid-cycle change methodology (Model for Improvement¹¹), and a well-respected framework for improving chronic illness care (Chronic Care Model¹²). Health centers from across the United States with experience in improving chronic illness care were invited to apply to participate in each DHDC. Each DHDC was conducted with the intent to continuously improve the care delivered to a heterogeneous group of patients with depression. Metrics were developed to monitor the changes in care as well as the impact of those changes on depression outcomes. Highlights regarding how the DHDC deployed these models are described in more detail below.

The DHDC framework. The DHDCs were created using the synergies of the three models mentioned above. The Breakthrough Series methodology¹⁰ provided the construct for collaborative learning. The fundamental principle behind this methodology is that there is a gap between what we know and what we do—in this case the gap was between evidence-based care for depression and “usual” depression care in ambulatory practice. Prior to the launch of the DHDC, experts in the field of depression gathered evidence regarding best practices for care of patients with depression in ambulatory settings. These best practices were then shared with multidisciplinary teams representing participating organizations utilizing various collaborative learning strategies over a thirteen month period. The team composition varied somewhat but typically included a physician who was committed to improving care for depressed patients (physician champion), a nurse care manager, a behavioral health specialist and a member of the senior leader team for the organization. For most of the teams the senior leader was the CEO.

Meetings called Learning Sessions allowed expert faculty to share best practices for depression care as well as change methodology to translate the evidence-based practice into local health center practice. Participants shared successful strategies to manage the changes in depression care within their local context. Faculty and participants worked together to identify challenges and craft solutions to improve outcomes. Action Periods, the time between Learning Sessions, allowed time for participating teams to make changes in the care for patients with depression and to monitor the impact of those changes in care outcomes. Teams were encouraged to include additional staff and senior leadership in adopting changes appropriate to their organization. Collaborative

participants and faculty maintained continual contact with each other through email, web portal, conference calls and site visits throughout each DHDC.

The translation of evidence-based care into practice is complicated and involves changes in how care is delivered, staff roles, how information is utilized and how patients are engaged. Even though health care professionals may understand evidence-based depression care, delivering that care reliably is often a challenge.

Utilizing the essential elements of the Care Model,¹¹ including the six domains of community relationships, organization of the health care system, self-management support, delivery system design, decision support and clinical information systems, the faculty organized a list of recommended changes for the depression care teams to test in their health care system. The changes recommended were considered likely to improve depression care or care outcomes based on strong evidence or experience.

The teams were encouraged to make changes in their care systems using the Model for Improvement, a simple yet powerful tool for accelerating improvement. The model has two parts:

- Three fundamental questions, which can be addressed in any order:
 - What are we trying to accomplish (aims)?
 - How will we know if a change is an improvement (measures)?
 - What changes can we make that will result in improvement (selecting changes to test)?
- *The Plan-Do-Study-Act (PDSA) cycle* to test and implement changes in real work settings. The PDSA cycle guides the test of a change to determine if the change is an improvement.

The Model for Improvement describes the PDSA cycle as a rapid method to plan and try a system change, observe the result and act on what was learned.

DHDC implementation. A total of 94 teams from health centers participated, distributed through the four 13-month DHDCs that began in 2000, 2002, 2003, and 2005. Teams came together for two-day learning sessions three times over the course of each DHDC to learn the three conceptual models, recommendations for implementing change from experts as well as to share practical strategies for improving depression care. Between learning sessions, teams met regularly in their real work settings and used the Plan-Do-Study-ACT (PDSA) cycle quality improvement methodology to test changes and to determine if their changes resulted in improved depression care. The teams used this method of action-oriented learning to guide their work. Depression Health Disparities Collaborative faculty and project staff supported teams as they worked to improve care for depression in their local health centers. One particularly effective support involved conducting teleconferences with teams to assess the impact of specific changes on recent performance.

Measurement of reliable and quantifiable performance indicators is an essential component of the Model for Improvement. At the time of the initial DHDC, measures of depression outcomes were limited to research settings and, with the exception of HEDIS pharmacy measures of continuity of antidepressant medications, no nationally recognized performance measures for depression care existed. Therefore, the faculty developed ten measures to support the work of the DHDC and refined them over

time, with input from teams based on their early experiences using the measures. The final versions of these measures are shown in Box 1. The DHDC teams were asked to submit performance data for these measures monthly. Faculty and teams monitored and analyzed progress toward improved depression care and depression care outcomes.

Challenges faced in developing performance measures for depression reflect the variability of the clinical presentation of depressed patients as well as the potentially negative impacts of the diagnosis. Depression represents a set of acute, sub-acute, and chronic conditions of varying levels of severity and stability, and patients are often in the process of transition into improvement or relapse. Discussions of how to improve care were complicated by these intricacies and the multiple depression subtypes included under the umbrella of a depression diagnosis (such as major depressive disorder, dysthymia, adjustment disorder with depressed mood, and others). Another barrier to overcome was the variation in readiness among clinicians to diagnose depression. Many primary care providers are reluctant to diagnose a condition called “major depression” and instead substitute a more socially acceptable term such as “adjustment” disorder despite the presence of symptoms clearly meeting diagnostic criteria. Common barriers to diagnosing major depression include social stigma, administrative difficulties, or insurance barriers.¹³ Thus, the depression faculty were spurred to adopt a standardized approach to depression evaluation and to create a new term, *clinically significant depression* (CSD) (defined as having a PHQ equal to 10 or greater AND any diagnosis of depression) that clearly warranted clinical attention.

The Patient Health Questionnaire 9 (PHQ-9)¹⁴ was selected to standardize the approach to depression evaluation for the DHDC. The PHQ-9 is a brief, nine-item depression assessment scale based directly on the diagnostic criteria for Major Depressive Disorder in the Diagnostic and Statistical Manual Fourth Edition (DSM-IV).¹⁵ Scores range from zero to 27, with higher scores reflecting more patient-reported depressive symptoms. The PHQ-9 has been established as a valid and reliable tool for both the diagnosis of depression as well as monitoring of changes during treatment.¹⁶ Although additional clinical assessment is required for diagnosis, this tool was used to screen for depression using a standardized instrument as well as to monitor the process of treatment outcome over time.

Patients were considered to have CSD if they received any depression diagnosis and scored 10 or higher on the PHQ-9. A score of 10 or greater was chosen as the cut-off for CSD because among primary care patients, this score is strongly associated with a diagnosis of major depression (88% sensitivity and 88% specificity).¹⁷ Clinicians were receptive to the use of this new term, and accepted the idea that patients with CSD should receive care to alleviate their symptoms. Screening, diagnosis, treatment, and appropriate follow-up are all components of depression care regardless of the specific type of depression. Defining CSD allowed teams to focus on improving care processes, regardless of the specific depression diagnosis.

All teams participating in the DHDCs reported on a set of four core measures and were able to choose additional, optional measures. Table 1 details the refined definitions of the core and optional measures used by DHDCs.

One measure, in particular, represents an entirely new concept in the measurement of chronic depression care. *Early and sustained response* (or ESR) is the first and only

Box 1.
ALL MEASURES^a

Core Measures			
#	Brief Name Denominator	Numerator	Clinical Relevance
1	Early and Sustained Response (ESR)	All patients with a diagnosis of Clinically Significant Depression (CSD) who have a 50% or greater reduction in Patient Health Questionnaire (PHQ) All CSD Patients	Documented early and sustained response to treatment; established metric of clinical improvement in randomized clinical trials
2	6 Month Repeat PHQ	All patients with a diagnosis of depression who have a PHQ score within the last 6 months (180 days). All patients with a diagnosis of depression	Clinical reassessment of depressive symptoms
3	Self-Management Goal Setting	All patients with a self-management goal within the last 12 months All patients with a diagnosis of depression	Importance of specific self-management support
4	5 point Drop in PHQ	All CSD patients who have a 5 point or greater reduction in PHQ within six months (180 days) of their New Episode PHQ ^b , and New Episode PHQ is within the last 12 calendar months. (If there is more than one Current PHQ within 6 months (180 days) of New Episode PHQ, take the most recent Current PHQ) All CSD patients with at least one Current PHQ within 6 months (180 days) of their most recent New Episode PHQ ^b and most recent New Episode PHQ is within the last 12 calendar months.	Documentation of clinically significant improvement in 6 months

(Continued on p. 160)

Box 1. (continued)

Optional Measures			
#	Brief Name Denominator	Numerator	Clinical Relevance
5	Early Follow-Up	All CSD patients with follow-up 1 to 3 weeks (7–21 days) after their last New Episode PHQ ^b and last New Episode PHQ is within the last 12 calendar months.	Importance of early clinical follow-up as per treatment guidelines
6	PHQ Reassessment	All patients with CSD All CSD patients with documented PHQ reassessment between 4–8 weeks (28–56 days) after their last New Episode PHQ ^b and last New Episode PHQ is within the last 12 calendar months. All CSD patients with a New Episode PHQ ^b within the last 12 calendar months.	Importance of reassessing depression severity after treatment initiation
7	Patients with diagnosis of major depression or dysthymia remaining on antidepressant for at least 6 months	All patients with a diagnosis of major depression or dysthymia (Why not just CSD?) who have been taking an antidepressant for at least 6 months (180 days). All patients with a diagnosis of major depression or dysthymia for at least 6 months (180 days).	Importance of duration of therapy as per treatment guidelines
8	Response to treatment in 12 months	All patients with a diagnosis of CSD who, 4 months (120 days) or longer after the last New Episode PHQ ^b , (occurring within the last 12 calendar months), have a 50% or greater reduction in PHQ. All patients with a diagnosis of CSD 4 months (120 days) or longer after the last New Episode PHQ ^b , and last New Episode PHQ is within the last 12 calendar months.	This measure allows health centers to measure outcomes on a yearly basis and compare performance improvement over time.

(Continued on p. 161)

Box 1. (continued)

Optional Measures			
#	Brief Name Denominator	Numerator	Clinical Relevance
9	Remission in 12 months	<p>All patients with a diagnosis of CSD who, 4 months (120 days) or longer after their last New Episode PHQ^b (occurring within the last 12 calendar months), have a most recent PHQ score less than 5.</p> <p>All patients with a diagnosis of CSD 4 months (120 days) or longer after their last New Episode PHQ^b, and last New Episode PHQ is within the last 12 calendar months.</p>	<p>This measure allows health centers to measure outcomes on a yearly basis and compare performance improvement over time.</p>
10	Depressed patients who improve in function	<p>All patients with a diagnosis of depression who, on their last New Episode Function,^c had a score >0 and whose last Current Function had a reduced score. (Date of Current Function must be later than the date of the last New Episode Function). This could be anywhere from a 1–3 point drop on the function question (#10). (Note that only patients with a New Episode Function within the last 12 months will be included in this measure).</p> <p>All patients with a diagnosis depression who have a score of >0 on their last New Episode Function, which must be within the last 12 calendar months.</p>	<p>Documentation of improvement in function with treatment</p>

^aThe brief names shown in this table have been updated and clarified from those used in Depression Health Disparities Collaboratives (DHDC)s; however, the numerators, denominators and calculations remain unchanged.

^bNew Episode PHQ: a PHQ score associated with the clinician’s decision that a new clinical episode needs treatment.

^cNew Episode Function: Patient’s response on question 10 of the PHQ9 corresponding to the new episode of depression. The score rates the patient’s ability to function on a scale from zero to three.

depression care measure to our knowledge that assesses a population of depressed patients' long term outcome over time. Specifically, the measure allows monitoring response to care: through the waxing and waning of the aggregate of each depressed individual's most current symptom scores, through early response (if it occurs) through relapses over time and including further treatment responses if any should occur in the future.

Early and sustained response is calculated utilizing standardized definitions of a numerator and denominator and then multiplying by 100 to convert to percentage of the population of patients with CSD that have at least a 50% reduction in their PHQ score. The denominator of ESR is the total number of patients with a PHQ score of 10 or greater, requiring any depression diagnosis (e.g., major depressive disorder, adjustment disorder with depressed mood) as well as the clinician's assessment that the patient's depressive condition should become the focus of a new treatment plan. Once a patient has been diagnosed with CSD, he/she remains in the denominator of this measure unless the clinician determines that the depression has resolved. Depression faculty encouraged clinicians to be cautious in deciding that depression was resolved considering that recurrence of depression is common.

The numerator of ESR is the number of CSD patients whose most recent PHQ-9 score reflects a 50% or greater reduction in PHQ-9 score compared to the PHQ-9 score that qualified the patient for the CSD diagnosis. This measure is calculated without any time constraints for most recent PHQ-9 or baseline PHQ-9. The absence of time constraints for this measure is what makes this measure especially unique and especially important. From the point of view of health of the population, the question of how long it took for a patient or group of patients to turn from depressed to "not depressed" is much less important than whether or not they made the change at all and whether they remained in that state of new health. It is the absence of time constraints which gives us this very simple window into health ("depressed" vs. "not depressed") in the numerator, independently of time.

We can illustrate how ESR operates for both short and long-term outcomes by considering an imaginary typical health center case. Mr. Rogers is a 45-year-old carpenter who came to the clinic in June 2002 for diabetes care and was also diagnosed with clinically significant depression (CSD), with a PHQ-9 score of 14. As a CSD patient, he was added to the denominator for the ESR measure in June 2002, but was not in the numerator because he had not had a 50% reduction in his PHQ-9 score. In August, he took the PHQ-9 a second time and scored 7. This score reflected a 50% reduction from baseline, so he was added to the numerator of the ESR measure, as a "responder." (Clinically one might consider him to be an "early responder," since only eight weeks had passed since treatment initiation). In November, he was seen in the clinic again and scored 9 on the PHQ. His depressive symptoms were slightly increased from his score in August and no longer reflected a 50% or greater drop when compared to baseline scores in June 2002. Therefore, Mr. Rogers was no longer considered a "responder" and was no longer counted in the numerator of this measure. As a result, the overall measure performance decreased because Mr. Rogers was taken out. The clinician adjusted Mr. Roger's treatment in November 2002, and his PHQ-9 scores in January, March, June

and December of 2003 were 7, 6, 7, and 3 respectively. All of these scores were 50% or less than his baseline score of 14 in June 2002, so he was included in the numerator of the ESR measure as a patient with a response which was sustained throughout all of 2003. It is perhaps important to note that we do not offer any specific definition of either “early” or “sustained” response; these are just rough approximations to observed clinical realities. We say Mr. Rogers’ response was “sustained” throughout 2003 simply because he did not relapse in 2003 and lose his response. (In June 2004, we can imagine that Mr. Rogers loses his job and catches his wife having an affair with his best friend, which leads to a divorce. He gets very depressed again, scoring 20 on the PHQ-9 on June 25, 2011. The clinician initiates new treatment, which could be a new medication, a new course of psychotherapy of both, for a new episode of depression and this PHQ is considered a “New Episode PHQ” for the “New Episode.”)

The case of Mr. Rogers illustrates how ESR provides a dynamic, long-term response rate of the entirety of a health center’s depressed patient population. Conversely, short-term, time limited measures may convey inflated response rates because they will not reflect the illness trajectory of patients with chronic depression which is characterized by fluctuating symptoms, relapse and recurrence. As a dynamic, chronic care measure, ESR provides a snapshot of the health of the population with respect to chronic illness care for depression.

Another novel measure created for this collaborative, one of process, also merits attention. Previous studies and our clinical reasoning suggested that early clinician reassessment (e.g., 1–2 months) after initiation of treatment, particularly antidepressant medication treatment, could have a significant impact on outcome.¹⁸ For example, once a patient is started on a medication, if he/she is seen and carefully reassessed with the PHQ in four to eight weeks, the clinician has an objective measure by which he/she can make a reasonable clinical determination whether to increase medication dose, continue the dose, switch to another medication, or make other changes. Without this objective measurement, there is much more subjectivity to the assessment. Moreover, without any observation at all, there will not be any opportunity to adjust treatment to meet the patient’s need at all. Expressed as an objective measurement then, this view would suggest that PHQ reassessment within four to eight weeks would be strongly associated with high ESR. To help understand this possible relationship, the association between ESR and four-to-eight-week PHQ reassessment, as well as the associations between all other measures, were tested using Pearson correlation coefficients. Regression analyses were also completed (see Figure 1 below).

Monthly reports. Teams submitted monthly performance data for core measures to the Health Disparities National Reporting (HDNR) website in the form of a standardized Excel® spreadsheet. Teams could also voluntarily choose to monitor and report optional measures. Each team from all four DHDCs continued to submit data until the end of the reporting period in 2006, though some teams found value in the reporting process and continued to report well beyond the projects’ initial timelines.

Through the HDNR website, teams could review their own data and DHDC faculty could analyze both aggregate and team-specific reports for each measure. Faculty used these analyses to coach the teams at least monthly. Though an initial screening was

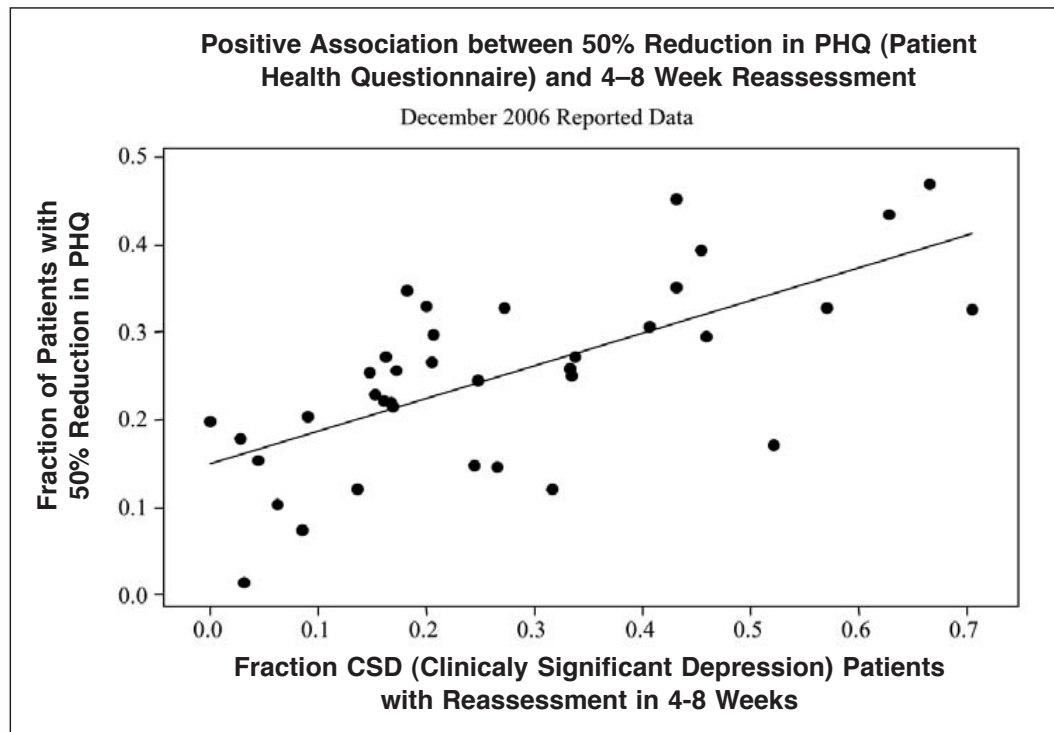


Figure 1. Association between ESR (Early and Sustained Response) and PHQ reassessment measures.

performed on data submissions for gross errors, the ultimate responsibility for the integrity of these data rested with the individual health centers, as they could revise previously submitted reports to correct data entry errors or other problems.

Analysis of data reported. Analyses were designed to give overview of performance over time for all measures at two periods of time. Data reported until 2006 shows improvement throughout the DHDC timeline. We wanted to demonstrate that participation in the DHDCs result in improved care, but also that the improvement could be sustained over time. To this end, we took another snapshot in December 2008 and focused our analysis on the sustainable improvement of DHDC teams. The performance of several teams was well above other teams. We looked specifically at this subset of teams to understand why they were so successful.

Analysis of High-Performance Teams. In 2006 the HDC depression faculty developed the criteria for selecting the highest-performing DHDC teams, hereafter known as High-Performance Teams (HPTs) to learn more about their common pathways to success. The criteria for selection included the performance on the first measure (ESR) of greater than the threshold/benchmark goal (40% with a minimum registry size of 150 patients during December 2006).

The authors conducted semi-structured telephone interviews with the leadership of the seven HPTs in 2006 to better understand organizational factors correlated with successful improvement in care for depressed patients. Teams were provided a copy of the interview questions prior to the interview.

The seven HPTs centers are located in various regions of the country: New Hampshire, South Carolina, Texas, Alaska, Missouri, Nebraska, and Colorado. All but two centers are located in medically underserved communities with designations of health professional shortage areas. These health centers have a medical provider staff size ranging from four to 39. Several centers provided care to the homeless and most were Ryan White grantees. All of the centers provide pharmacy services and dental services either on-site or through community contracts. All HPTs provide on-site integrated behavioral health care, although several did not begin this service until their participation in the DHDCs.

Results

Core measures. Available data for all DHDC measures were analyzed as a snapshot of aggregate team performance in December 2006 and again in December 2008. We also reported performance data of our HPTs at those times. In 2006, the total depression collaborative registry size for all 94 centers was 38,124 patients, including 23,928 with CSD.

Table 1 provides snapshots of team performance on core measures in 2006 and 2008. Data are separated into three groups: the All Teams column details data from all teams reporting in December 2006, the Continued Reporters column summarizes data from

Table 1.

PERFORMANCE DATA FOR CORE MEASURES

#	Measure	All Teams	Continued Reporters		High Performance Teams (HPTs)	
		2006 Mean Patients (#Centers)	2006 Mean Patients (Centers)	2008 Mean Patients (Centers)	2006 Mean Patients (Centers)	2008 Mean Patients (Centers)
1	Early and Sustained Response (ESR)	24.6% 23,928 (94)	28.5% 12,571 (35)	29.7% 14,764 (35)	40.7% 3,076 (5)	42.2% 3,566 (5)
2	6 Month Repeat Patient Health Questionnaire (PHQ)	44.3% 38,124 (94)	49.2% 17,430 (35)	37.2% 21,545 (35)	62.2% 4,332 (5)	51.6% 5,280 (5)
3	Self-Management Goal-Setting	42.6% 38,124 (94)	51.4% 17,430 (35)	40.3% 21,545 (35)	63.6% 4,332 (5)	50.5% 5,280 (5)
4	5 Point Drop in PHQ	54.6% 3,479 (77)	57.7% 2,003 (31)	58.7% 1,520 (32)	61.9% 612 (5)	61.4% 420 (5)

those teams who continued to report through 2008, and the High Performance Teams column displays data from those Continued Reporters who displayed exemplary results. For each group, the mean performance on the Core Measures, number of patients impacted, and the number of health centers reporting is shown.

All health centers reported data for each of the first three Core Measures. Measure four, a five-point reduction in PHQ score for CSD patients within six months of initial treatment, was introduced in 2005; therefore, health centers completing a DHDC prior to 2005 did not report on it as a Core Measure. All subsequent participants (77 health centers or 82% of total participants) reported on measure four.

Thirty-five of the 94 health centers (37%), which served 17,430 patients with depression including 12,571 with CSD, continued to report data through 2008, at which time these health centers served 21,545 patients with depression including 14,764 with CSD. This increase reflects additional clinicians and their patients that participated in enhanced care processes for depression as well as more effective approaches to screening and diagnosis.

Consistent with our interest in and commitment to sustainability, we reported data only for those high-performing teams who continued to report through 2008: five of the seven HPTs (71%) continued to report outcomes in 2008. The five HPTs reported results on core measures in both 2006 and 2008 reflecting 5,280 patients with depression, including 3,566 with CSD. As shown in Table 2, the performance of HPTs exceeded that of other teams for each core measure in 2006 and again in 2008.

Health centers reported 24.6% of their CSD patient's PHQ scores improved 50% or more compared to their initial PHQ scores, and 54.6% of all patients' PHQ scores improved by at least five points within six months of treatment. Performance on two core measures (measures two and three) that reflect processes of care indicated that 44.3% of all depressed patients had a documented PHQ score within the previous six months and 42.6% had self self-management goals within the last 12 months.

Optional measures. A smaller number of centers reported on the optional measures. Table 3 provides snapshots of team performance on optional measures in 2006 and 2008. Data are separated into three groups as presented in Table 2.

In 2006 the number of health centers reporting on each measure ranged from 4–37, including 411–4,246 patients with all categories of depression or CSD. In 2008 the number of community health centers reporting ranged from 3–17, including 256–2,605 patients with all categories of depressions or CSD.

In 2006 and 2008, the most frequently reported optional measure by health centers was the reassessment of CSD patients within four to eight weeks after initial treatment. The four-to-eight-week reassessment measure was reported in 2006 by 37 health centers for 4,246 CSD patients in their registries and in 2008, 17 centers reported the measure for 2,605 CSD patients.

The measure of the proportion of patients remaining on an antidepressant for at least six months after initiation of treatment was reported by 23 CHCs for 4,380 patients in 2006 and by nine CHCs for 2,471 patients in 2008. In 2006, 16 CHCs reported that 51.3% of all depressed patients in their registries (n=1,110) had shown improvement in function from the date of entry into the registry and in 2008 seven CHCs reported 48% of all depressed patients (n=684) had shown improvement.

Table 2.**PERFORMANCE DATA FOR OPTIONAL MEASURES**

#	Measure	All Teams	Continued Reporters		High Performance Teams	
		2006 Mean Patients (Centers)	2006 Mean Patients (Centers)	2008 Mean Patients (Centers)	2006 Mean Patients (Centers)	2008 Mean Patients (Centers)
5	Early Follow-Up	30.7% 1,386 (9)	34.6% 885 (6)	31.9% 761 (6)	31.5% 813 (3)	35.5% 679 (4)
6	PHQ (Patient Health Questionnaire) Reassessment	27.4% 4,246 (37)	34.9% 1,967 (15)	27.4% 2,605 (17)	51.5% 648 (3)	37.7% 806 (4)
7	Patients with diagnosis of major depression or dysthymia remaining on antidepressant for at least 6 months	42.3% 496 (7)	42.7% 405 (5)	42.5% 325 (3)	43.8% 329 (2)	46.6% 277 (3)
8	Response to treatment in 12 months	21.7% 411 (4)	22.3% 391 (3)	28.1% 256 (3)	23.7% 329 (2)	29.4% 177 (2)
9	Remission in 12 months	79.5% 4,380 (23)	80.1% 2,570 (9)	72.4% 2,471 (9)	72.5% 570 (1)	69.2% 1,150 (1)
10	Depressed patients who improve in function	51.3% 1,110 (16)	54.1% 735 (7)	48.0% 684 (7)	56.0% 373 (2)	54.1% 266 (2)

Correlation and regression analyses: Relationship between measures. Data values for December 2006 were used to explore inter-correlations between all measures. The only two measures highly inter-correlated in 2006 were ESR and PHQ Reassessment. Figure 1 is the scatter plot showing that an increase in PHQ Reassessment scores was associated with an increase in ESR scores for all 37 health centers that reported both ESR and PHQ Reassessment in December 2006. (Pearson correlation coefficient = .67, $p < .000$, $R^2 = 43.6\%$.)

This observed relationship makes clinical sense and indeed, was initially predicted as likely. In fact, the measure itself was created to help guide and encourage teams to perform this second severity assessment 1–2 months after initiating treatment, in order to have the opportunity to re-evaluate the impact of initial choice of treatment and have the opportunity to make adjustments based on initial response.

Interview results with HPTs. In 2006, telephone semi-structured interviews were conducted with the leadership of seven of the highest performing community health center teams to identify the common pathways to their success in implementing the depression measures.

When asked to describe how they managed depression care in their clinics prior to their participation in the depression collaborative, most centers reported having no standardization of care for depression. Each provider treated the depressed patient “as they thought best.” None of the centers were using the PHQ-9 or other tools for the routine screening of depression. All but two centers had only limited access to behavioral health provider (BHP) expertise. The teams were asked to identify their key implementations of the care model that they deemed to be of pivotal importance in achieving their successful outcomes.

Analysis of the interviews identified 18 common drivers of improvement for depression care, embedded within all six components of the chronic care model. These drivers are outlined in Box 2. Interviewers were impressed with the level of detail provided for each of these drivers and the extent to which these drivers were integrated into the fabric of each organization.

Discussion

Providing care for patients with depression has proven to be challenging in the primary care setting, often resulting in suboptimal patient outcomes.⁵ This paper describes an overall quality improvement framework that was used successfully in a heterogeneous group of health centers to improve care for depression¹⁹.

Practice teams participating in the DHDCs were able to improve both care delivery and care outcomes for depression even for populations challenged with a high incidence of poverty, low insurance coverage, and limited access to behavioral health supports. These improvements were supported by a well established framework that encouraged translation of evidence into practice, standardized^{20–24} performance measurement over time and support in using data to improve performance. Certain characteristics correlating with implementation of the Care Model were demonstrated by a small group of High-Performance Teams who were able to achieve impressive sustained long-term patient care response rates comparable to benchmarks reported in acute short-term experimental studies.

It is worth noting that the first measure, early and sustained response (ESR) is a measure that captures overall response rate of a population over time, including both early as well as sustained responses. It has no specific start, stop, or cut-off dates and includes all patients begun on treatment, even those dropping out.

With respect to the issue of drop outs, it is also worth noting that high-performing centers, in fact, typically distinguished themselves^{25–27} with vigorous proactive efforts

Box 2.

DRIVERS OF IMPROVEMENT FOR DEPRESSION CARE

<p>Organization of Health Care System</p>	<ul style="list-style-type: none"> • Visible financial and moral support from the CEO and Medical Director • A “day to day” leader responsible for coordination of sustaining and spreading change concepts • A physician champion to promote acceptance among peers
<p>Delivery System Design</p>	<ul style="list-style-type: none"> • Train multidisciplinary care teams to deliver depression care • Weekly team meetings during implementation phase to “drill down” the data on process and outcome measures • Use PDSA (plan/do/study/act) cycles of improvement • Develop written evidenced-based clinical processes for the care of depression and standardize system wide • Emphasize culturally competent depression care management
<p>Decision Support</p>	<ul style="list-style-type: none"> • Use of the Patient Health Questionnaire-9 for routine depression screening and monitoring of severity • Same day access to on-site integrated behavioral health provider expertise • Behavioral health providers use a mix of brief behavioral change interventions and traditional Cognitive-Behavioral Therapy, Interpersonal Psychotherapy, Problem-Solving Therapy for individuals and families • Psychotropic medication management
<p>Clinical Information System</p>	<ul style="list-style-type: none"> • Electronic management of depression process and outcome measures • Use of data for routine quality improvement reviews and required reporting
<p>Patient Self Management Support</p>	<ul style="list-style-type: none"> • Train care team on use of “Ultra-Brief Personal Action Planning (UB-PAP™)” • Document self-management goals
<p>Community Relationships</p>	<ul style="list-style-type: none"> • Present depression measure population data to funding sources and payers • Develop Memoranda of Agreement/contractual arrangements with local resources including mental health providers

to keep patients in treatment who might, in this typically under-served population, otherwise have missed follow-up appointments. These vigorous outreach efforts not only helped keep drop-out rates low, contributing to ESR response rates by tracking improvement when present, but also helped increase performance on the process measure four-to-eight-week PHQ reassessment, which itself seemed to help predict outcome on ESR. Patients who received the benchmark four-to-eight-week PHQ re-assessment

were able to receive the clinical benefit of appropriate adjustment of treatment based on the results of this re-assessment.

These two measures, the four-to-eight-week PHQ reassessment and the ESR, were effective in guiding treatment in day-to-day clinical care, driving quality improvement efforts, and achieving benchmark outcomes. We believe that HPTs achieved and sustained response rates of 40%, a response rate equivalent to those achieved in short-term experimental clinical trials, provides a useful benchmark of excellence for quality improvement efforts in other settings.

We should point out a possible limitation of the interviews with the high-performing centers. These interviews certainly identified 18 elements related to the care model that leaders thought contributed to the success of their efforts. These change efforts were convincing to the interviewers because of the level of detail that was provided and the clear degree to which these changes were integrated into the organizations. However, we cannot be sure that these were the changes that made the difference and were causally related to the improvements demonstrated. It is certainly possible that other centers may have made the same changes without noting any improvements in outcomes. Further study is certainly needed.

Another limitation of this study is that participation was limited to health centers. Fortunately, similar infrastructure support has been shown to be helpful to support practice improvements in private practices in other programs such as Improving Performance in Practice (IPIP)²⁸ and some of the patient-centered medical home (PCMH)²⁹ demonstration projects. An additional limitation is that the data were self-reported and not validated by a third party. However, since data were primarily utilized by the teams to refine procedures and improve outcomes, it was advantageous to the health centers to ensure that data were valid. Finally, the data were collected as performance measure sets, not at the individual patient level.

Conclusion. Well-designed measures and systematic approaches utilizing proven improvement methodologies resulted in enhancements in care delivery and outcomes for patients with depression in health centers. In particular, one process measure, four-to-eight-week PHQ re-assessment, and one outcome measure, early and sustained response (ESR), both proved useful as measures for chronic depression care quality improvement efforts. We suggest that both these measures may be useful more generally to the field. Centers achieved notable outcomes on ESR that were comparable in magnitude to improvements in other published reports of depression outcomes in general, and, in particular, of performance improvement efforts for depression in primary care. The authors suggest that this approach to improvement guided by performance measures may benefit care for depression in other primary care settings.

Acknowledgments

The authors wish to thank HRSA (contract number HSH250200866002C), which provided partial supportive funding for the writing of this paper. The authors would also like to acknowledge the co-Chairs of the 2000 Collaborative, David Katzelnick and Michael Von Korff, and faculty of the 2000 Collaborative including Henry Chung, as well as the care teams of the 17 health centers who laid important groundwork for

subsequent work. Finally, the faculty (including Henry Chung, David Pollack, and Tom Oxman) of the 2002, 2003, and 2005 Collaboratives also merit special thanks for their considerable efforts and participation at all levels of development and implementation of the ideas promoted through the Collaborative. The authors would also like to acknowledge the seven “High-Performing Teams” of Ammonoosuc Community Health Center (CHC), CareSouth Carolina, Inc. CHC, City of Austin CHC, Interior CHC, Missouri Highlands CHC, One World CHC, and Pueblo CHC for their participation in the interviews and willingness to share their success strategies. Finally, we appreciate and acknowledge the editing assistance of Colin Reims.

Appendix 1—Interview Questions

1. First of all, tell me about the structure of your center.
 - a. Integrated behavioral health vs. not integrated
 - b. Received mental health expansion grant
 - c. Large (many sites) vs. small
 - d. Urban/rural
 - e. Homeless?
 - f. Other?
2. We would like to get your own thoughts on your pathways to success. That is, what did you do that seemed to be of pivotal importance in obtaining your outstanding results?
3. Please tell me about the organizational support that was devoted to this effort. What resources were used?
4. What was the composition of the change team? What were the factors contributing the way the individuals on the team worked together? Were any roles changed to help facilitate efficient operation of the team?
5. What communications were utilized within your center to facilitate the implementation of changes and to obtain “buy-in”?
6. Tell me about your early successes. What were they and how did they help you gain momentum?
7. How did you institutionalize change? How did you change your clinical culture? Did job descriptions change?
8. How has this change process been similar to or different from other organizational changes in your center?
9. Did your center participate in any other collaboratives? Which ones? Was that before or after your participation in the depression collaborative?
10. If you did participate in other collaboratives, please describe how previous experience in one collaborative influenced process or outcomes in the next collaborative.

Notes

1. World Health Organization. The world health report 2004: changing history, Annex Table 3: burden of disease in DALYs by cause, sex, and mortality stratum in WHO regions, estimates for 2002. Geneva, Switzerland: World Health Organization, 2004.
2. Kessler RC, Chiu WT, Demler O, et al. Prevalence, severity, and co morbidity of twelve-month DSM-IV disorders in the National Co-morbidity Survey Replication (NCS-R). *Archives of General Psychiatry*. 2005 Jun;62(6):617–27.
3. Pratt LA, Brody D. Depression in the United States household population, 2005–2006. Hyattsville, MD: U.S. Public Health Service, 2008.
4. Cepoiu M, McCusker J, Cole M, et al. Recognition of depression by non-psychiatric physicians—a systematic literature review and meta-analysis. *J Gen Intern Med*. 2008 Jan;23(1):25–36.
5. Mental Health Quality Enhancement Research Initiative. Strategic plan. Washington, DC: Mental Health Quality Enhancement Research Initiative Center, 2006.
6. Owen R, Magruder K, Jonson R, et al. Screening for mental health and substance use disorders. Washington, DC: U.S. Department of Justice, 2006.
7. Katelnick DJ, Von Korff M, Chung H, et al. Applying depression-specific change concepts in a collaborative breakthrough series. *Jt Comm J Qual Patient Saf*. 2005 Jul;31(7):386–97.
8. Rubenstein LV. Review: collaborative care was effective for depression in primary care in the short and longer term. *Evid Based Med*. 2007 Aug;12(4):109.
9. Meredith LS, Mendel P, Pearson M, et al. Implementation and maintenance of quality improvement for treating depression in primary care. *Psychiatric Serv*. 2006 Jan; 57(1):48–55.
10. Institute for Healthcare Improvement. The breakthrough series: IHI's collaborative model for achieving breakthrough improvement. Boston, MA: Institute for Healthcare Improvement, 2003.
11. Wagner EH, Austin BT, Davis C, et al. Improving chronic illness care: translating evidence into action. *Health Aff (Millwood)*. 2001 Nov–Dec;20(6):64–78.
12. Langley GL, Nolan KM, Nolan TW, et al. The improvement guide: a practical approach to enhancing organizational performance (1st ed.). San Francisco, CA: Jossey-Bass, 1996.
13. Rost K, Smith R, Matthews DB, et al. The deliberate misdiagnosis of major depression in primary care. *Arch Fam Med*. 1994 Apr;3(4):333–7.
14. Löwe B, Unützer J, Callahan CM, et al. Monitoring depression treatment outcomes with the patient health questionnaire–9. *Med Care*. 2004 Dec;42(12):1194–201.
15. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (4th ed.). Washington, DC: American Psychiatric Association, 2000.
16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001 Sep;16(9):606–13.
17. Lin EH, Von Korff M, Russo J, et al. Can depression treatment in primary care reduce disability? A stepped care approach. *Arch Fam Med*. 2000 Nov;9(10):1052–8.
18. Katon W, Rutter C, Ludman EJ, et al. A randomized trial of relapse prevention of depression in primary care. *Arch Gen Psychiatry*. 2001 Mar;58(3):241–7.
19. Katon W, Von Korff M, Lin E, et al. Collaborative management to achieve treatment guidelines. Impact on depression in primary care. *JAMA*. 1995 Apr 5;273(13):1026–31.
20. Olivarius NF, Beck-Nielsen H, Andreasen AH, et al. Randomised controlled trial of structured personal care of type 2 diabetes mellitus. *BMJ*. 2001 Oct 27;323(7319):970–5.

21. Renders CM, Valk GD, Franse LV, et al. Long-term effectiveness of a quality improvement program for patients with type 2 diabetes in general practice. *Diabetes Care*. 2001 Aug;24(8):1365–70.
22. Norris SL, Engelgau MM, Narayan KM. Effectiveness of self-management training in type 2 diabetes: a systematic review of randomized controlled trials. *Diabetes Care*. 2001 Mar;24(3):561–87.
23. Maly RC, Leake B, Frank JC, et al. Implementation of consultative geriatric recommendations: the role of patient-primary care concordance. *J Amer Geriatr Soc*. 2002 Aug;50(8):1372–80.
24. Heisler M, Bouknight RR, Hayward RA, et al. The relative importance of physician communication, participatory decision-making and patient understanding in diabetes self-management. *J Gen Intern Med*. 2002 Apr;17(4):243–52.
25. Lorig KR, Sobel DS, Stewart AL, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing utilization and costs: a randomized trial. *Med Care*. 1999 Jan;37(1):5–14.
26. Löwe B, Kroenke K, Herzog W, et al. Measuring depression outcome with a short self-report instrument: sensitivity to change of the Patient Health Questionnaire (PHQ-9). *J Affective Disorders*. 2004 Jul;78(8):131–40.
27. Rost K, Nutting P, Smith JL, et al. Managing depression as a chronic disease: a randomized trial of treatment in primary care. *BMJ*. 2002 Oct 26;325:934.
28. Margolis PA, Dewalt DA, Simon JE, et al. Designing a large-scale multilevel improvement initiative: the improving performance in practice program. *J Continuing Education in the Health Professions*. 2010 Summer;30(3):187–96.
29. Bitton A, Martin C, Landon BE. A nationwide survey of patient-centered medical home demonstration projects. *J Gen Intern Med*. 2010 Jun;25(6):584–92.