

The Effect of Adherence to Practice Guidelines on Depression Outcomes

Kimberly A. Hepner, PhD; Melissa Rowe, PhD; Kathryn Rost, PhD; Scot C. Hickey, MA; Cathy D. Sherbourne, PhD; Daniel E. Ford, MD; Lisa S. Meredith, PhD; and Lisa V. Rubenstein, MD, MSPH

Background: Few studies have assessed clinician adherence to depression practice guidelines and the relationship between clinician adherence and depression outcomes.

Objective: To estimate how frequently specific guideline recommendations are followed and to assess whether following guideline recommendations is linked to improved depression outcomes.

Design: Observational analysis of data collected from 1996 to 1998 in 3 randomized clinical trials.

Setting: 45 primary care practices in 13 U.S. states.

Patients: 1131 primary care patients with depression.

Measurements: Expert panel methods were used to develop a patient survey–based index that measured adherence to clinical practice guidelines on depression. Rates of adherence to the 20 indicators that form the index were evaluated. Multivariable regression that controlled for case mix was used to assess how index scores predicted continuous and dichotomous depression measures at 12, 18, and 24 months.

Results: Quality of care was high (clinician adherence $\geq 79\%$) for 6 indicators, including primary care clinician detection of depression.

Quality of care was low (adherence, 20% to 38%) for 8 indicators, including management of suicide risk (3 indicators), alcohol abuse (2 indicators), and elderly patients; assessment of symptoms and history of depression; and treatment adjustment for patients who did not respond to initial treatment. Greater adherence to practice guidelines significantly predicted fewer depressive symptoms on continuous measures ($P < 0.001$ for 12 months, $P < 0.01$ for 18 months, and $P < 0.001$ for 24 months) and dichotomous measures ($P < 0.05$ for 18 and 24 months).

Limitations: Data are based on patient self-report. Possible changes in practice since 1998 may limit the generalizability of the findings.

Conclusions: Adherence to guidelines was high for one third of the recommendations that were measured but was very low for nearly half of the measures, pointing to specific needs for quality improvement. Guideline-concordant depression care appears to be linked to improved outcomes in primary care patients with depression.

Ann Intern Med. 2007;147:320-329.

For author affiliations, see end of text.

www.annals.org

Depression practice guidelines for primary care represent expert consensus, based on systematic review of several randomized clinical trials and other relevant literature, on how depression should be diagnosed and treated. Guidelines provide guidance for depression assessment, referral to mental health specialists, and follow-up in primary care (1–3). Numerous studies document low rates of adherence to guideline recommendations on antidepressant use or psychotherapy (4–6) and less-than-optimal outcomes in usual care (7). Far fewer studies have assessed adherence to the broader set of recommendations on additional aspects of care ranging from detection of depression symptoms to treatment discontinuation. National efforts to improve care (8) rely substantially on 1 or 2 guideline recommendations (for example, completion of a short-term care course of antidepressant medication). Knowing the strengths and gaps in care for primary care patients

with depression could provide a stronger basis for quality improvement.

We sought to determine whether adherence to the broader set of depression guideline recommendations improves depression outcomes and, if so, to identify which guideline recommendations are most and least likely to be followed. To achieve these aims, we used expert panel methods to develop 20 guideline-based quality indicators for depression care (9, 10). We applied the indicators to secondary data from surveys of 1131 representative depressed patients from 45 diverse and geographically dispersed primary care practices. We analyzed the link between performance on our measure and depression outcomes at 12, 18, and 24 months and assessed the level of performance on each of our indicators to identify which aspects of depression care were most and least in need of improvement.

See also:

Print

Editors' Notes 321
Editorial comment 342

Web-Only

Conversion of graphics into slides

METHODS

We performed secondary analysis of data collected through the Quality Improvement for Depression (QID) collaboration (11) between 1996 and 1998. The 4 studies in this collaboration used cluster randomized designs to evaluate encouragement strategies for implementing Agency for Healthcare Research and Quality guidelines for depression (1–3) in primary care. One QID study used self-

administered surveys (12), which had some questions that were worded differently from questions in the other 3 studies. We included only the 3 QID studies that used virtually identical telephone-administered surveys. Among these 3 studies, only 1 showed statistically significant improvement in depression-specific outcomes (13, 14). In the QID studies, experimental-group clinician leaders were given training, materials, and protocols to encourage them to provide collaborative care for depression-based quality improvement (Figure). Additional details on the encouragement approach used by each study are published elsewhere (11–14). For our observational analyses, we combined the usual care and experimental groups. The study was approved by the institutional review board at each study site.

Setting

Patients and clinicians were recruited from 45 primary care practices within 4 health care organizations in 13 states. Organizations included a multistate family practice network (11), 2 staff-model managed care organizations (Kaiser Permanente and Veterans Affairs [14]), and a network model managed care organization whose patients were cared for in private practices throughout Maryland (11, 15).

Sample

Researchers consecutively approached patients visiting participating practices to identify those with symptoms of depression on a brief self-report depression screening instrument derived from the World Health Organization Composite International Diagnostic Interview (CIDI) (16). Telephone interviewers used the full CIDI for Affective Disorders to assess consenting patients who tested positive on the CIDI screening instrument. We included 1019 patients who were positive for a major depressive disorder or dysthymia on the CIDI and 112 patients who were negative despite reporting at least 5 of 9 depression criteria (17) at recruitment (18). In the QID studies, follow-up surveys were completed by 85% of the enrolled patients at 12 months, 77% at 18 months, and 78% at 24 months.

Measures

We measured the process of care, case mix (based on prognosis), and outcomes for depression (Figure). Process-of-care measures included 139 self-report items (19). Case-mix measures (20) included 70 items used to create 5 scales (baseline severity of depression on the modified Center for Epidemiologic Studies–Depression [CES-D] scale, severity of physical symptoms common among people with depression, social support, number of depression symptoms lasting ≥ 2 years, and appropriate antidepressant use during ≥ 3 months in the past 6 months) that formed the basis for a validated depression prognostic index (DPI) (14, 21, 22). Higher DPI scores indicate worse prognosis.

Outcome measures included depression symptoms (continuous) and persistent depression (dichotomous) at 12, 18, and 24 months after baseline (the date of study enrollment). We measured depression symptoms by using

Context

Do clinicians who follow depression practice guidelines have better patient outcomes?

Contribution

This observational study from 45 primary care practices found that clinicians adhered well to only one third of the guideline recommendations for depression. Most clinicians initially recognized and treated depression; however, they often did not address suicide risk, assess alcohol use, adjust treatment appropriately, or follow through on long-term treatment plans. Better clinician adherence to recommendations was associated with a lower probability of persistent depressive symptoms among patients.

Caution

Data on depression-related care were based on patients' self-report.

Implication

Clinicians should try harder to deliver guideline-concordant depression care.

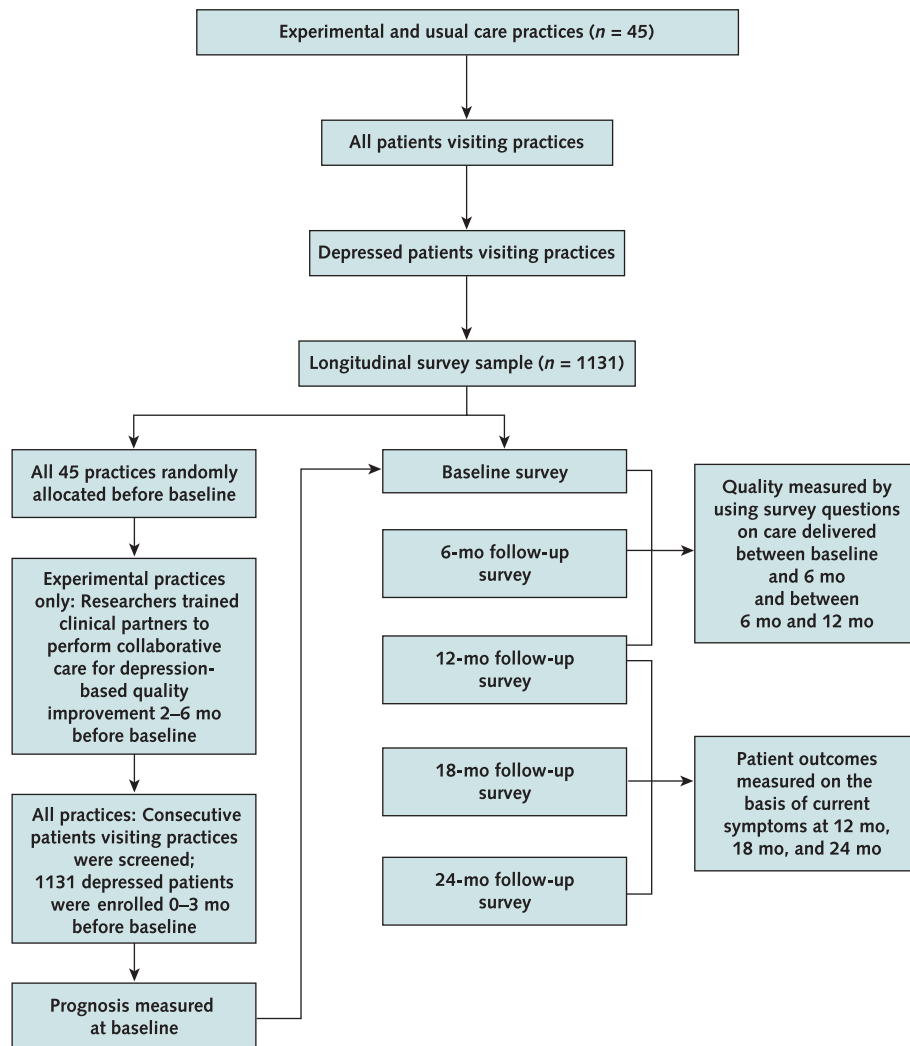
—The Editors

a modified version of the CES-D scale (23, 24) that more closely approximates criteria from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*; lower scores on the modified CES-D scale reflect fewer depression symptoms (range, 0 to 100). We measured the presence or absence of persistent depression by using a previously validated measure (25) that approximates current major depression. We defined “persistent depression” as scores that fell in the depressed range (according to previously established cutoff values) on each of 3 measures: the modified CES-D scale (score ≥ 20 , equivalent to ≥ 16 on the original CES-D), the mental health composite score of the Short Form-12 (26) (score ≤ 1 SD below the general population mean of 50), and stem items from the CIDI screening instrument that identifies current “probable depression” (16). Unlike the CIDI structured diagnosis instrument, the CIDI screening instrument requires the presence of depressed mood and anhedonia in the past month. Because patients must score in the depressed range on all 3 measures according to standard depression cutoffs (the measures map closely to diagnostic criteria from the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*) and major symptoms must be current, the patients identified by the composite measure for persistent depression are substantially and currently impaired.

Development of the Depression Quality Index

We developed preliminary quality review criteria based on Agency for Healthcare Research and Quality guidelines (1–3) and refined them through an expert panel process. We translated quality review criteria into scored, survey-

Figure. Quasi-experimental design of studies in the Quality Improvement for Depression collaboration.



based quality indicators, without testing any indicators against outcome until the index was complete. We validated the index by evaluating the relationship between quality scores and depression outcomes, adjusting for case mix.

We developed clinical algorithms for identifying standard, above-standard, or below-standard care relative to guideline statements. In standard care, a care process would be likely to ensure good to excellent outcomes for most patients, whereas below-standard care would lead to poor outcomes among some patients and above-standard care would have a high likelihood of leading to optimal outcomes.

We then submitted preliminary quality review criteria to an independent group of depression care experts (5 primary care practitioners and 2 mental health specialists) from the MacArthur Foundation Initiative on Depression and Primary Care, using a modified Delphi process (27).

Seven experts participated in 3 rounds of review and revision of the criteria and 1 final round of review to assess consensus on the final measure. The consensus process resulted in 20 quality review criteria (19). A strength of these criteria is that they account for the complexity of care decisions made over time on the basis of response to treatment.

We translated the final quality review criteria into measurable quality indicators based on survey variables. Each indicator identifies a patient population (for example, patients with persistent symptoms at 6 months) to which a care process or processes identified by the quality review criterion applies (for example, the patient should be seen by a mental health specialist). Scoring further divides the indicator into standard, above-standard, or below-standard care subsets on the basis of additional conditions specified in the quality review criterion (for example, per our panel's recommendations, care is below standard for suicide assess-

ment only when the patient came for a scheduled visit with no new medical problems). Each indicator generates a raw score of -9 to 9 , with standard care as zero.

Statistical Analysis

We used weighting and multiple imputation to account for difficult-to-survey subpopulations (such as ethnic minority groups). Weighting accounted for the probability of enrollment and attrition at each step, beginning with data on patients approached for screening. Multiple hot-deck imputation used for missing values for most items produced the 5 randomly imputed data sets that we used in the analyses (28). Item-level missing data were low ($<10\%$ for any item), and in previous analyses (29), the 5 data sets produced similar results. We adjusted SEs for uncertainty due to imputation (30). We used the sandwich estimator, also known as the *robust variance estimator* (31, 32), to adjust SEs for hierarchical sampling with clustering of patients in practices (31) in all regression analyses. We assessed the percent adherence to each indicator by identifying the population or subpopulation to which the overall indicator applied and the number of patients who received the care assessed by the indicator.

To create an overall depression quality index (DQI), we averaged scores across the 20 indicators. We first standardized indicator scores by creating a z score; indicators thus had equal weight when averaged. We used multivariable regression to assess whether higher DQI scores were statistically significantly associated with better depression outcomes at 12, 18, and 24 months. Regressions included only the DQI score and our case-mix adjustor (the DPI) as independent variables; the DPI reflects and replaces usual depression case-mix adjustors (14, 21, 29). We used linear regression to predict the continuous outcome measure and logistic regression to predict the dichotomous outcome measure. A P value less than 0.05 was considered significant.

We evaluated whether the relationship between the DQI and outcomes was stable relative to patient severity of illness by dividing our sample into baseline prognosis tertiles (DPI) and quality tertiles (DQI). We then calculated mean modified CES-D scale scores and the predicted probability of persistent depression (by using coefficients from the logistic regression analyses described above) at 12, 18, and 24 months for each set of prognostic and quality tertiles. We also estimated the relative risk for persistent depression at 12, 18, and 24 months among patients who received low-quality care compared with all other patients, adjusting for baseline depression prognosis by using the DPI.

Our analyses were observational but were based on data from randomized (intervention versus usual care) quality improvement experiments. We used hierarchical regression that first included the DPI and the DQI, followed by the patient's intervention group status, to test whether

the DQI accounted for changes in quality induced by the study interventions.

Role of the Funding Sources

This research was funded by the National Institute of Mental Health, Agency for Healthcare Research and Quality, the John D. and Catherine T. MacArthur Foundation, and Veterans Affairs Health Services Research & Development Service Center of Excellence for the Study of Healthcare Provider Behavior. The funding sources had no role in the design, analysis, or interpretation of the study or in the decision to submit the manuscript for publication.

RESULTS

Table 1 shows baseline characteristics of the study sample. About one third of the sample was male (29%), belonged to a minority group (28%), or was unemployed (36%). Few patients had less than high school education (13%) or were older than 65 years (7%).

Table 2 shows the characteristics of the clinicians and practices that provided care to these patients. Practices were diverse in size, financing, and ethnicity mix, ranging from small private or mixed-payer practices (55% of patients) to large managed care clinics (45% of patients).

Table 3 shows the proportion of patients whose care met the standard set by each of the 20 DQI indicators. Adherence ranged from 100% (for basic patient education) to 23% (for primary care clinician assessment of alcohol use or abuse). In general, primary care clinicians did well (guideline adherence $>70\%$) in recognizing and monitoring depression over the succeeding months, both in terms

Table 1. Patient Characteristics*

Characteristic	Data
Mean age (SD), y	45 (13.49)
Sex, n (%)	
Male	326 (29)
Female	805 (71)
Race, n (%)	
African American	125 (11)
White	809 (72)
Hispanic	97 (9)
Other	100 (9)
Age >65 y, n (%)	82 (7)
Married, n (%)	484 (43)
Employed, n (%)	670 (59)
Education less than high school, n (%)	144 (13)
Depression diagnoses, n (%)	
Dysthymia alone	6 (1)
Depression with dysthymia	172 (15)
Depression without dysthymia	841 (74)
Subthreshold depression	112 (10)
Received care at an enhanced-care site, n (%)	656 (58)
Mean modified CES-D scale score (SD)†	
At baseline	51 (21.37)
At 6 months	40 (25.53)

* Based on 1131 patients. Missing data were estimated by using multiple imputation. CES-D = Center for Epidemiologic Studies–Depression.

† Score ranges from 0 to 100, with scores >20 indicating a high likelihood of major depression (higher scores indicate more distress).

Table 2. Practice and Clinician Characteristics

Characteristic	Data
Health care organizations, <i>n</i>	4
Participating practices, <i>n</i>	45
Participating primary care providers, <i>n</i>	124
Clinician sex, <i>n</i> (%)	
Female	55 (44)
Male	66 (53)
Missing	3 (2)
Clinician race, <i>n</i> (%)	
Nonwhite	32 (26)
White	87 (70)
Missing	5 (4)
Clinician specialty, <i>n</i> (%)	
General/family practice	35 (28)
Internal medicine	89 (72)
Completed any continuing education for depression in the past 3 years, <i>n</i> (%)	87 (70)

of visits and of discussing depression with patients. Treatments, when initiated, usually matched patient needs and preferences. Clinicians (either psychiatrists or primary care clinicians) followed most depressed patients at least twice in 6 months and followed patients in whom antidepressant therapy was newly started at least 3 times in 6 months. Clinicians appropriately started antidepressant therapy or psychotherapy in patients with anxiety or panic, and most clinicians did not start long-term benzodiazepine therapy.

Primary care clinicians did less well (guideline adherence, 45% to 59%) in following through on treatment over time. Less than half of patients (46%) completed minimum treatment with antidepressants (≥ 2 months) or psychotherapy (≥ 4 visits); the rate was only slightly higher (51%) when we considered only patients who were amenable to receiving treatment. After patients completed treatment successfully, with resolution of symptoms, clinicians often failed to discuss depression during the succeeding 6 months (adherence, 59%). Patients who were not treated often were not monitored closely (adherence, 53%), and clinicians often did not try discontinuing antidepressant therapy in patients with no history of depression whose depression had resolved (adherence, 45%). Psychotherapy, when received, typically did not include even minimal elements of cognitive-behavioral therapy (adherence, 55%).

The lowest quality of care (adherence, 23% to 38%) occurred in adjustment of treatment in patients who did not respond over the first 6 months and management of suicidality, comorbid conditions, and depression in elderly persons. Most patients who remained depressed at 6 months did not have their treatment adjusted (adherence, 38%). Most elderly persons (age >64 years) did not complete a minimal course of appropriate treatment (adherence, 26%).

All 3 indicators for suicidality management fell into the lowest quality category. Suicidality usually was not assessed (adherence, 24%), was not treated with appropriate

medications or psychotherapy (adherence, 26%), and did not result in mental health specialist consultation (adherence, 36%). Of 144 patients with a previous suicide attempt or some indication of a suicide plan, only 51 (35%) saw any mental health specialist over the succeeding 6 months. Clinicians also usually did not assess alcohol use (23%), and patients with panic episodes or alcohol abuse who remained depressed at 6 months usually did not receive consultation with a mental health specialist over the succeeding 6 months (adherence, 30%). Rates were also low for assessment of mood disturbance, including assessment of depression history, depression symptoms, or manic symptoms (adherence, 34%).

Table 4 shows that after adjustment for patient prognosis, the DQI significantly predicted modified CES-D scale score at 12, 18, and 24 months and persistent depression at 18 and 24 months. Table 5 shows the mean modified CES-D scale scores at 12, 18, and 24 months, by baseline tertiles of prognosis (DPI) and quality of care (DQI). Patients with the middle or worst prognoses reported decreasing depressive symptoms when receiving higher quality of depression care at each time point.

Table 5 also shows the predicted probability of persistent depression. The effect of differences in quality is primarily seen for patients with the middle and worst prognoses at 18 and 24 months. For example, at 24 months, the predicted probability of persistent depression was 0.61 in patients in the worst prognosis group who received the lowest-quality care, compared with 0.51 in patients in the worst prognosis group who received the highest-quality care. Estimates of relative risk for persistent depression among patients who received low-quality care compared with all others was statistically significant at 18 months (relative risk, 1.22; $P < 0.05$), after adjustment for baseline depression prognosis. Relative risks at 12 months (1.04) and 24 months (1.09) were not significant.

Intervention group status (usual care versus intervention) was not statistically significant when added to regressions predicting depression symptoms or persistent depression for any time point (12, 18, or 24 months). Of the 20 indicators, only 3 differed significantly by intervention group status: appropriate treatment for patients with comorbid anxiety or panic, mental health specialist consultation for panic or alcohol use among patients still depressed at 6 months, and intensity of primary care treatment during the acute phase.

DISCUSSION

We found that primary care physicians provide high-quality care in terms of basic detection and initiation of treatment for depression but substantially lower-quality care in terms of completion of a minimal course of treatment for depression (especially among elderly persons) or assessment and treatment of psychiatric comorbid condi-

Table 3. Observed Adherence to Indicators of Quality of Depression Care*

Quality Indicator	Patients, <i>n</i>	Observed Adherence to Indicator, % [†]	Indicator Description
Low need for quality improvement[‡]			
Basic patient education	1003	100	Patients received depression education (discussed with a PCP or received depression information from another source).
PCP attention to depression during acute phase	743	88	The PCP discussed depression during follow-up with patients who had ≥ 2 primary care visits.
Treatment–patient match among treated patients	403	84	Patients who completed minimally appropriate treatment by 6 months had received treatment aligned with their symptom profiles (e.g., vegetative symptoms, life stress) and preferences.
Follow-up for patients with newly started antidepressant therapy	147	80	Patients with newly started appropriate antidepressant therapy received adequate follow-up (≥ 3 visits with a PCP or mental health specialist) over the succeeding 6 months.
Appropriate treatment for patients with comorbid anxiety or panic	176	80	Patients with anxiety or panic symptoms received appropriate antidepressant or psychotherapy and no more than 2 months of benzodiazepines (among patients not receiving benzodiazepine therapy before baseline).
PCP detection of depression	1131	79	The PCP explicitly detected the patient's depression over the 6 months after baseline by talking about depression, recommending psychotherapy, prescribing medication, or assessing suicidal ideation.
Primary care treatment intensity during acute phase	1131	73	Depressed patients followed in primary care visited the PCP at least twice in 6 months.
Medium need for quality improvement[§]			
PCP monitoring of vulnerable patients during continuation phase	802	59	Patients with symptom resolution but a history of depression episodes received PCP monitoring (at least 1 PCP visit in which depression was discussed) during the succeeding 6 months.
Quality of psychotherapy provided by a mental health specialist	295	55	Patients attending mental health services for psychotherapy received ≥ 1 element typical of cognitive-behavioral therapy (e.g., behavioral activation, reducing negative thoughts).
PCP monitoring of untreated patients who may be in watchful waiting	799	53	Patients not treated with antidepressants or psychotherapy received appropriate watchful waiting care (3 PCP visits; one-time discussion of depression).
Initial treatment completion	1003	46	Patients completed appropriate antidepressant therapy (≥ 2 months) or psychotherapy (4 visits).
Discontinuing antidepressant therapy in low-risk patients	109	45	Patients who completed appropriate antidepressant therapy (≥ 2 months), whose depression had resolved, and who were at low risk for recurrence had a trial discontinuation of antidepressant therapy.
High need for quality improvement			
Treatment adjustment among nonresponsive patients	640	38	Patients who remained depressed at 6 months while being treated with an appropriate antidepressant received a change in antidepressant, started psychotherapy, received additional psychotherapy visits, or were hospitalized.
Mental health specialist consultation for suicidal ideation	219	36	Patients with suicidal ideation received consultation with a mental health specialist over the next 6 months.
Depression history and symptom assessment by PCP	1131	34	The PCP assessed depression history, depression severity, or mania symptoms.
Consult with mental health specialist for panic or alcohol use among patients still depressed at 6 months	110	30	Patients with panic or alcohol abuse symptoms and ongoing depression received consultation with a mental health specialist (if they were not already being followed in mental health services).
Treatment for suicidal ideation among patients not already followed in mental health care	246	28	Patients with suicidal ideation received appropriate, low-toxicity antidepressant medication or appropriate psychotherapy.
Any appropriate treatment for depressed elderly persons	70	26	Depressed elderly persons completed appropriate treatment with antidepressants (≥ 2 months) or psychotherapy.
Suicide assessment by PCP	1131	24	The PCP assessed the patient for suicidal ideation.
Alcohol assessment by PCP	1131	23	The PCP assessed the patient for alcohol use or abuse.

* PCP = primary care physician.

[†] The number of patients eligible for the indicator who met the standard. In some cases, patient answers, such as “don't know,” further reduce the number of applicable patients.

[‡] Defined as observed adherence $\geq 73\%$.

[§] Defined as observed adherence of 45% to 59%.

^{||} Defined as observed adherence of 20% to 38%.

tions. In addition, we found little evidence of appropriate response to suicidal ideation, few changes in treatment in response to persistent or improved symptoms at 6 months,

and low rates of referral to mental health specialists for complex patients. When patients did access mental health specialists, nearly half did not report receiving any elements

Table 4. Regression Analyses to Predict Patient Outcomes at 12, 18, and 24 Months*

Predictor Variable	β Value (95% CI) for Predicting Depression Symptom Score	Odds Ratio (95% CI) for Predicting Persistent Depression
Outcome at 12 months (n = 876)		
Intercept	-1.47 (-5.7 to 2.7)	-3.52 (-4.2 to -2.8)
Depression prognostic index	1.00 (0.9 to 1.1)	1.08 (1.06 to 1.09)
Depression quality index	-7.86 (-11.6 to -4.1)	0.75 (0.54 to 1.03)
Outcome at 18 months (n = 731)		
Intercept	-0.42 (-5.5 to 4.6)	-3.46 (-3.9 to -3.0)
Depression prognostic index	0.96 (0.9 to 1.1)	1.07 (1.05 to 1.08)
Depression quality index	-7.00 (-11.0 to -3.0)	0.64 (0.41 to 0.98)
Outcome at 24 months (n = 745)		
Intercept	2.8 (-2.6 to 8.2)	-3.04 (-3.9 to -2.2)
Depression prognostic index	0.9 (0.8 to 1.0)	1.07 (1.05 to 1.08)
Depression quality index	-9.32 (-14.4 to -4.2)	0.59 (0.38 to 0.91)

* Adjusted for patient prognosis.

of evidence-based (cognitive-behavioral) therapy (33), such as being helped to reduce negative thoughts.

We found rates of treatment completion (46%) that were in the range reported by other investigators, such as Charbonneau and colleagues, who reported 45% completion of antidepressant treatment (4). Although most of the remaining quality measures that we report have not been systematically assessed across diverse practices, our results conceptually support the findings of previous depression studies. Our finding that primary care clinicians usually detect and initiate treatment of depression is consistent with findings that most primary care clinicians think that they should be treating depression, have basic knowledge

about treatment, and want to improve the depression care they provide (34). The high level of depression recognition that we found also helps to explain why interventions to improve depression care that focus on clinician education or reminders alone are not effective (35–37)—clinicians already know that their patients are depressed but still cannot achieve high-quality care.

The quality deficits that we identified also help explain results of studies showing that collaborative care interventions for depression improve outcomes (38). For example, the weaknesses in patient assessment for symptoms and comorbid conditions, completion of treatment, and treatment adjustment relative to symptoms over time highlight

Table 5. Mean Modified Center for Epidemiologic Studies–Depression Scale Scores and Predicted Probability of Persistent Depression at 12, 18, and 24 Months, by Prognosis and Quality Tertiles*

Baseline Prognostic Group†	Low-Quality Care			Middle-Quality Care			High-Quality Care		
	Patients, n	Mean (\pm SE) Modified CES-D Scale Score‡	Probability of Persistent Depression	Patients, n	Mean (\pm SE) Modified CES-D Scale Score‡	Probability of Persistent Depression	Patients, n	Mean (\pm SE) Modified CES-D Scale Score‡	Probability of Persistent Depression
12 months									
Best prognosis	80	26.1 \pm 2.8	0.18	121	23.5 \pm 1.9	0.16	86	23.8 \pm 2.3	0.15
Middle prognosis	90	39.2 \pm 2.7	0.38	90	35.1 \pm 2.6	0.35	90	33.7 \pm 1.8	0.33
Worst prognosis	84	55.5 \pm 2.0	0.63	97	51.6 \pm 2.0	0.59	104	46.9 \pm 2.1	0.57
18 months									
Best prognosis	70	24.1 \pm 2.3	0.18	96	25.0 \pm 1.9	0.15	72	23.8 \pm 2.0	0.13
Middle prognosis	74	37.1 \pm 2.8	0.35	79	35.9 \pm 2.5	0.31	102	32.7 \pm 2.4	0.28
Worst prognosis	71	56.2 \pm 2.3	0.58	77	49.7 \pm 3.1	0.53	90	46.2 \pm 2.0	0.49
24 months									
Best prognosis	72	28.4 \pm 3.0	0.23	103	26.4 \pm 2.2	0.19	74	22.9 \pm 2.3	0.16
Middle prognosis	75	37.2 \pm 2.3	0.41	72	33.7 \pm 2.4	0.35	111	32.5 \pm 1.7	0.31
Worst prognosis	74	55.2 \pm 2.3	0.61	74	49.9 \pm 2.8	0.56	90	49.4 \pm 2.9	0.51

* CES-D = Center for Epidemiologic Studies–Depression.

† Based on the depression prognostic index.

‡ Modified CES-D scores range from 0 to 100; higher scores indicate worse depression. Prognostic tertile cut-points are 33.0 and 44.3 (range, 8.2 to 73.8). Quality tertile cut-points are -0.11 and 0.19 (range, -2.1 to 1.0).

problems addressed by collaborative care intervention features, such as depression care manager–assisted comprehensive assessment, proactive symptom monitoring, and treatment adjustment. Our finding of low rates of referral to mental health specialists for complex patients is typically addressed in collaborative care interventions through stepped care (39) that prioritizes mental health specialist referrals on the basis of need. Finally, collaborative care interventions targeted to elderly persons have had particularly large effects (38, 40).

Consistent with published research (41, 42), the low rates of treatment completion that we found could not be attributed primarily to patients not wanting treatment. Most patients (85%) were amenable to treatment, and our indicator for treatment completion judged treatment completion only for this motivated group. We assessed care for the remaining patients under the indicator for watchful waiting.

After controlling for case mix by using a detailed index derived from baseline data, we found that depression quality of care was associated with outcomes at 12, 18, and 24 months after baseline for our depression symptom outcome (modified CES-D scale score) and at 18 and 24 months for persistent depression. A similar study found no process–outcome link (43). Previous studies showing process–outcome links focused on narrower sets of measures (35–40) or used employment as an outcome (44–46). Quality scores on our measure were not statistically significantly correlated with patient sickness, indicating that our algorithm-based method adequately accounted for differential patient needs such that patients with mild (good prognosis) depression scored similarly to those with severe (poor prognosis) depression. Outcomes also worsened relative to declining quality within each of 3 levels of sickness (prognosis), with the strongest outcome effects on patients with a poor prognosis—the group that we expect to be most vulnerable to persistent depression.

The study interventions focused on encouraging guideline-concordant care rather than introducing new, unproven treatments. Our quality scores also measured adherence to existing guidelines. We therefore expected, and indeed found, that the relationship between quality scores and outcomes would not be affected by experimental group assignment independent of whether the intervention improved performance. The degree of adherence to indicators that we found, however, may be somewhat higher than is typical for usual care, owing to modest improvements in quality in the intervention group (33).

The use of patient self-report to assess quality of care is both a limitation and a strength of our study. Self-reported quality measures have been validated (10, 47) and may have advantages for depression. There are no laboratory tests or physical findings for depression, and depression rarely leads to death. Patient self-report on depression diagnosis, severity, comorbid conditions, and outcomes is likely to be more accurate than medical record informa-

tion. For various medical conditions, care appears worse when measured on the basis of medical record review compared with reports from trained simulated patients (48, 49), owing to omitted documentation. Documentation is especially likely to be incomplete for stigmatized mental health conditions, such as depression. These advantages of patient self-report are accompanied by the disadvantages of administration and analysis of a long (139-item) survey, making the DQI impractical for routine clinical use in its current form. Patient recall may also be faulty. Ultimately, combinations of limited patient self-report and medical record review may provide the best compromise for use by such groups as the Health Plan Employer Data and Information Set (50).

Our study has other limitations. First, although the study involved diverse types of practices from a wide geographic area, some types of practices, such as publicly funded clinics other than Veterans Affairs clinics, were not represented. Second, care may have improved since our data were collected; the indicators should be reapplied in future research to assess change over time. Third, the study is observational and cannot establish causality—that is, we cannot prove that better process by our measure causes improved outcomes. Finally, the relationship between quality and outcomes, although significant, was modest.

In contrast, a strength of our approach is that our indicators were explicitly judged by an experienced expert panel as feasible for average primary care practices to achieve. To avoid bias, we did not analyze any process–outcome links until the entire index was built and scored, depending on guideline-based clinical logic, validation by independent experts, and evaluation of the index as a whole in relationship to outcomes. Future research should further explore the relationships between individual indicators and outcomes and among case mix, quality, and outcomes, as well as variations in quality relative to patient, provider, and practice characteristics.

In summary, we found notable strengths in how primary care physicians care for patients with depression, including various indications that primary care clinicians recognize and respond to depressive symptoms in their patients. However, we found important quality deficits, including pressing needs for clinicians to better detect and manage complex or poorly responding patients; encourage collaboration with mental health specialists; and increase treatment completion, especially among elderly patients.

From RAND Health Program, Santa Monica, California; Florida State University, Tallahassee, Florida; Johns Hopkins School of Medicine, Baltimore, Maryland; and Veterans Affairs Greater Los Angeles Healthcare System and David Geffen School of Medicine at the University of California, Los Angeles, Los Angeles, California.

Acknowledgments: The authors thank the MacArthur Foundation Initiative on Depression and Primary Care for assisting with the development of the quality review criteria; Bernadette Benjamin and Jose Arbelaez for expert programming; Jeff Smith, Maureen Carney, Chantal

Avila, and Carole Oken for data collection and data management; Kathryn Magruder and Paul Nutting for their work on the overall collaborations; and Bob Bell for statistical consultation. They also thank the organizations participating in the studies: Kaiser Permanente Medical Care Programs in the Northern California Region, Oakland, California; Veterans Affairs Greater Los Angeles, Los Angeles, California; Ambulatory Sentinel Practice Network, Denver, Colorado; and NYL Care Health Plans of the Mid-Atlantic, Greenbelt, Maryland. Finally, they thank the clinicians and patients who contributed their time and effort for these studies.

Grant Support: By the National Institute of Mental Health (grants R01MH64658, P50MH54623, R01 MH54444, and MH 63651), the Agency for Healthcare Research and Quality (grant R01-HS08349), the John D. and Catherine T. MacArthur Foundation (grant 96-42 901 A-HE), and the Veterans Affairs Health Services Research and Development Service Center of Excellence for the Study of Healthcare Provider Behavior (grant LIP 65-030).

Potential Financial Conflicts of Interest: *Consultancies:* K. Rost, D.E. Ford (Pfizer Inc.). *Honoraria:* K. Rost. *Grants pending:* K. Rost.

Requests for Single Reprints: Kimberly A. Hepner, PhD, RAND Health Program, 1776 Main Street, Santa Monica, CA 90407; e-mail, hepner@rand.org.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Depression Guideline Panel.** Clinical Practice Guideline, Depression in Primary Care: Volume 1. Detection and Diagnosis. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service Agency for Health Care Policy and Research; 1993. AHCPR publication no. 93-0550.
2. **Depression Guideline Panel.** Clinical Practice Guideline, Depression in Primary Care: Volume 2. Treatment of Major Depression. Rockville, MD: U.S. Department of Health and Human Services, Public Health Service Agency for Health Care Policy and Research; 1993. AHCPR publication no. 93-0551.
3. **Schulberg HC, Katon W, Simon GE, Rush AJ.** Treating major depression in primary care practice: an update of the Agency for Health Care Policy and Research Practice Guidelines. *Arch Gen Psychiatry.* 1998;55:1121-7. [PMID: 9862556]
4. **Charbonneau A, Rosen AK, Ash AS, Owen RR, Kader B, Spiro A 3rd, et al.** Measuring the quality of depression care in a large integrated health system. *Med Care.* 2003;41:669-80. [PMID: 12719691]
5. **Katon W, Schulberg H.** Epidemiology of depression in primary care. *Gen Hosp Psychiatry.* 1992;14:237-47. [PMID: 1505745]
6. **Kerr EA, McGlynn EA, Van Vorst KA, Wickstrom SL.** Measuring antidepressant prescribing practice in a health care system using administrative data: implications for quality measurement and improvement. *Jt Comm J Qual Improv.* 2000;26:203-16. [PMID: 10749005]
7. **Gilbody S, Whitty P, Grimshaw J, Thomas R.** Educational and organizational interventions to improve the management of depression in primary care: a systematic review. *JAMA.* 2003;289:3145-51. [PMID: 12813120]
8. **New Freedom Commission on Mental Health.** Achieving the Promise: Transforming Mental Health Care in America. Final Report. Rockville, MD; U.S. Department of Health and Human Services; 2003. Publication no. SMA-03-3832.
9. **Kahn KL, Rogers WH, Rubenstein LV, Sherwood MJ, Reinisch EJ, Keeler EB, et al.** Measuring quality of care with explicit process criteria before and after implementation of the DRG-based prospective payment system. *JAMA.* 1990;264:1969-73. [PMID: 2120476]
10. **Shekelle PG, MacLean CH, Morton SC, Wenger NS.** Assessing care of vulnerable elders: methods for developing quality indicators. *Ann Intern Med.* 2001;135:647-52. [PMID: 11601947]
11. **Rost KM, Duan N, Rubenstein LV, Ford DE, Sherbourne CD, Meredith LS, et al.** The Quality Improvement for Depression collaboration: general analytic strategies for a coordinated study of quality improvement in depression care. *Gen Hosp Psychiatry.* 2001;23:239-53. [PMID: 11600165]
12. **Wells KB, Sherbourne C, Schoenbaum M, Duan N, Meredith L, Unützer J, et al.** Impact of disseminating quality improvement programs for depression in managed primary care: a randomized controlled trial. *JAMA.* 2000;283:212-20. [PMID: 10634337]
13. **Rost K, Nutting P, Smith JL, Elliott CE, Dickinson M.** Managing depression as a chronic disease: a randomised trial of ongoing treatment in primary care. *BMJ.* 2002;325:934. [PMID: 12399343]
14. **Rubenstein LV, Meredith LS, Parker LE, Gordon NP, Hickey SC, Oken C, et al.** Impacts of evidence-based quality improvement on depression in primary care: a randomized experiment. *J Gen Intern Med.* 2006;21:1027-35. [PMID: 16836631]
15. **Meredith LS, Jackson-Triche M, Duan N, Rubenstein LV, Camp P, Wells KB.** Quality improvement for depression enhances long-term treatment knowledge for primary care clinicians. *J Gen Intern Med.* 2000;15:868-77. [PMID: 11119183]
16. **World Health Organization.** Composite International Diagnostic Interview (CIDI Core) Version 2.1. Interviewer's Manual. Geneva: World Health Organization; 1997.
17. **Zimmerman M, Coryell W.** The Inventory to Diagnose Depression (IDD): a self-report scale to diagnose major depressive disorder. *J Consult Clin Psychol.* 1987;55:55-9. [PMID: 3571659]
18. **Rost K, Nutting P, Smith J, Werner J, Duan N.** Improving depression outcomes in community primary care practice: a randomized trial of the quEST intervention. Quality Enhancement by Strategic Teaming. *J Gen Intern Med.* 2001;16:143-9. [PMID: 11318908]
19. **Hepner KA, Rowe M, Rost K, Hickey SC, Sherbourne CD, Ford DE, et al.** Assessing depression care in primary care settings using the Depression Quality Index. Santa Monica, CA: RAND. [Forthcoming]. Report no. TR-251-HLTH.
20. **Donabedian A, Bashur R.** An Introduction to Quality Assurance in Health Care. Oxford: Oxford Univ Pr; 2002.
21. **Hayward RA, Kent DM, Vijan S, Hofer TP.** Reporting clinical trial results to inform providers, payers, and consumers. *Health Aff (Millwood).* 2005;24:1571-81. [PMID: 16284031]
22. **Keeler EB, Kahn KL, Draper D, Sherwood MJ, Rubenstein LV, Reinisch EJ, et al.** Changes in sickness at admission following the introduction of the prospective payment system. *JAMA.* 1990;264:1962-8. [PMID: 2120475]
23. **Orlando M, Sherbourne CD, Thissen D.** Summed-score linking using item response theory: application to depression measurement. *Psychol Assess.* 2000;12:354-9. [PMID: 11021160]
24. **Radloff L.** The CES-D scale. A self-report depression scale for research in the general population. *Applied Psychiatric Measures.* 1977;1:385-401.
25. **Sherbourne CD, Wells KB, Duan N, Miranda J, Unützer J, Jaycox L, et al.** Long-term effectiveness of disseminating quality improvement for depression in primary care. *Arch Gen Psychiatry.* 2001;58:696-703. [PMID: 11448378]
26. **Ware J Jr, Kosinski M, Keller SD.** A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care.* 1996;34:220-33. [PMID: 8628042]
27. **Fink A, Siu AL, Brook RH, Park RE, Solomon DH.** Assuring the quality of health care for older persons. An expert panel's priorities. *JAMA.* 1987;258:1905-8. [PMID: 3656600]
28. **Little RJ.** *Statistical Analysis with Missing Data.* New York: Wiley; 1987.
29. **Rubenstein LV, Rayburn NR, Keeler EB, Ford DE, Rost KM, Sherbourne CD.** Predicting outcomes of primary care patients with major depression: development of a depression prognosis index. *Psychiatr Serv.* 2007; 58:1049-56.
30. **Schafer JL.** *Analysis of Incomplete Multivariate Data.* London: Chapman & Hall; 1997.
31. **Bell RM, McCaffrey DF.** Bias reduction in standard errors for linear regression with multi-stage samples. *Surv Methodol.* 2002;28:169-79.
32. **Williams RL.** A note on robust variance estimation for cluster-correlated data. *Biometrics.* 2000;56:645-6. [PMID: 10877330]
33. **Beck JS.** *Cognitive Therapy: Basics and Beyond.* New York: Guilford; 1995.
34. **Meredith LS, Rubenstein LV, Rost K, Ford DE, Gordon N, Nutting P, et al.** Treating depression in staff-model versus network-model managed care organizations. *J Gen Intern Med.* 1999;14:39-48. [PMID: 9893090]
35. **Rubenstein LV, McCoy JM, Cope DW, Barrett PA, Hirsch SH, Messer KS, et al.** Improving patient quality of life with feedback to physicians about

functional status. *J Gen Intern Med.* 1995;10:607-14. [PMID: 8583263]

36. Rollman BL, Hanusa BH, Lowe HJ, Gilbert T, Kapoor WN, Schulberg HC. A randomized trial using computerized decision support to improve treatment of major depression in primary care. *J Gen Intern Med.* 2002;17:493-503. [PMID: 12133139]
37. Thompson C, Kinmonth AL, Stevens L, Peveler RC, Stevens A, Ostler KJ, et al. Effects of a clinical-practice guideline and practice-based education on detection and outcome of depression in primary care: Hampshire Depression Project randomised controlled trial. *Lancet.* 2000;355:185-91. [PMID: 10675118]
38. Callahan CM, Boustani MA, Unverzagt FW, Austrom MG, Damush TM, Perkins AJ, et al. Effectiveness of collaborative care for older adults with Alzheimer disease in primary care: a randomized controlled trial. *JAMA.* 2006;295:2148-57. [PMID: 16684985]
39. Katon W, Von Korff M, Lin E, Simon G, Walker E, Unützer J, et al. Stepped collaborative care for primary care patients with persistent symptoms of depression: a randomized trial. *Arch Gen Psychiatry.* 1999;56:1109-15. [PMID: 10591288]
40. Unützer J, Katon W, Callahan CM, Williams JW, Hunkeler E, Harpole L, et al; IMPACT Investigators. Improving Mood-Promoting Access to Collaborative Treatment. Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. *JAMA.* 2002;288:2836-45. [PMID: 12472325]
41. Wells KB, Sherbourne CD. Functioning and utility for current health of patients with depression or chronic medical conditions in managed, primary care practices. *Arch Gen Psychiatry.* 1999;56:897-904. [PMID: 10530631]
42. Lenert LA, Sherbourne CD, Sugar C, Wells KB. Estimation of utilities for

the effects of depression from the SF-12. *Med Care.* 2000;38:763-70. [PMID: 10901359]

43. Dobscha SK, Gerrity MS, Corson K, Bahr A, Cuiwik NM. Measuring adherence to depression treatment guidelines in a VA primary care clinic. *Gen Hosp Psychiatry.* 2003;25:230-7. [PMID: 12850654]
44. Schoenbaum M, Unützer J, McCaffrey D, Duan N, Sherbourne C, Wells KB. The effects of primary care depression treatment on patients' clinical status and employment. *Health Serv Res.* 2002;37:1145-58. [PMID: 12479490]
45. Rost K, Fortney J, Coyne J. The relationship of depression treatment quality indicators to employee absenteeism. *Ment Health Serv Res.* 2005;7:161-9. [PMID: 16194001]
46. Rost K, Smith BS, Elliott CE, Dickinson LM. The effect of improving primary care depression management on employee absenteeism and productivity: a randomized trial. *Med Care.* [Forthcoming].
47. Hargraves JL, Hays RD, Cleary PD. Psychometric properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 adult core survey. *Health Serv Res.* 2003;38:1509-27. [PMID: 14727785]
48. Luck J, Peabody JW. Using standardised patients to measure physicians' practice: validation study using audio recordings. *BMJ.* 2002;325:679. [PMID: 12351358]
49. Luck J, Peabody JW, Dresselhaus TR, Lee M, Glassman P. How well does chart abstraction measure quality? A prospective comparison of standardized patients with the medical record. *Am J Med.* 2000;108:642-9. [PMID: 10856412]
50. Spettell CM, Wall TC, Allison J, Calhoun J, Kobylinski R, Fargason R, et al. Identifying physician-recognized depression from administrative data: consequences for quality measurement. *Health Serv Res.* 2003;38:1081-102. [PMID: 12968818]

EXPEDITED REVIEW

Annals invites authors of clinically important randomized, controlled trials to request expedited review and publication of their manuscripts. Send requests to Harold Sox (hsox@mail.acponline.org), Christine Laine (chrisl@mail.acponline.org), Michael Berkwits (mberkwits@acponline.org), or Cynthia Mulrow (cmulrow@acponline.org). We take extra efforts to provide thorough, high-quality, and timely critiques of trials that we expedite. We publish expedited trials that are accepted early online. We also provide readers ancillary material about selected trials, including registered protocols, lists of other ongoing and published relevant trials, lists of relevant published systematic reviews, and links to clinical sources that provide physicians and patients information about the topic of the trial.

Current Author Addresses: Drs. Hepner, Rowe, Sherbourne, and Meredith and Mr. Hickey: RAND Health Program, 1776 Main Street, Santa Monica, CA 90407.

Dr. Rost: Department of Medical Humanities and Social Sciences, Florida State University, 1115 West Call Street, Tallahassee, FL 32306.

Dr. Ford: Department of General Internal Medicine, Johns Hopkins School of Medicine, 733 North Broadway, Suite 115, Baltimore, MD 21205.

Dr. Rubenstein: Veterans Affairs Greater Los Angeles (152), 16111 Plummer Street, Sepulveda, CA 91343.

Author Contributions: Conception and design: K.A. Hepner, M. Rowe, K. Rost, C.D. Sherbourne, D.E. Ford, L.S. Meredith, L.V. Rubenstein. Analysis and interpretation of the data: K.A. Hepner, M. Rowe, K. Rost, S.C. Hickey, C.D. Sherbourne, D.E. Ford, L.V. Rubenstein.

Drafting of the article: K.A. Hepner, C.D. Sherbourne, L.V. Rubenstein. Critical revision of the article for important intellectual content: K.A. Hepner, M. Rowe, K. Rost, C.D. Sherbourne, D.E. Ford, L.S. Meredith, L.V. Rubenstein.

Final approval of the article: K.A. Hepner, M. Rowe, K. Rost, S.C. Hickey, C.D. Sherbourne, D.E. Ford, L.V. Rubenstein.

Provision of study materials or patients: L.S. Meredith, L.V. Rubenstein. Obtaining of funding: K. Rost, D.E. Ford, L.S. Meredith, L.V. Rubenstein.

Administrative, technical, or logistic support: M. Rowe, S.C. Hickey, L.V. Rubenstein.

Collection and assembly of data: M. Rowe, K. Rost, S.C. Hickey, D.E. Ford, L.S. Meredith, L.V. Rubenstein.