Identifying Physician-Recognized Depression from Administrative Data: Consequences for Quality Measurement

Claire M. Spettell, Terry C. Wall, Jeroan Allison, Jaimee Calhoun, Richard Kobylinski, Rachel Fargason, and Catarina I. Kiefe

Background. Multiple factors limit identification of patients with depression from administrative data. However, administrative data drives many quality measurement systems, including the Health Plan Employer Data and Information Set (HEDIS®).
Methods. We investigated two algorithms for identification of physician-recognized depression. The study sample was drawn from primary care physician member panels of a large managed care organization. All members were continuously enrolled between January 1 and December 31, 1997. *Algorithm 1* required at least two criteria in any combination: (1) an outpatient diagnosis of depression or (2) a pharmacy claim for an antidepressant. *Algorithm 2* included the same criteria as algorithm 1, but required a diagnosis of depression for all patients. With algorithm 1, we identified the medical records of a stratified, random subset of patients with and without depression ($n = 465$). We also identified patients of primary care physicians with a minimum of 10 depressed members by algorithm 1 ($n = 32{,}819$) and algorithm 2 ($n = 6{,}837$).
Results. The sensitivity, specificity, and positive predictive values were: Algorithm 1: 95 percent, 65 percent, 49 percent; Algorithm 2: 52 percent, 88 percent, 60 percent. Compared to algorithm 1, profiles from algorithm 2 revealed higher rates of follow-up visits (43 percent, 55 percent) and appropriate antidepressant dosage acutely (82 percent, 90 percent) and chronically (83 percent, 91 percent) ($p < 0.05$ for all).
Conclusions. Both algorithms had high false positive rates. Denominator construction (algorithm 1 versus 2) contributed significantly to variability in measured quality. Our findings raise concern about interpreting depression quality reports based upon administrative data.

Key Words. Depression, HEDIS, quality of care, predictive value

Administrative medical, pharmacy, and membership files of managed care organizations offer relatively low-cost, convenient data sources for examining patterns of care at the population level. Cohorts defined from administrative data often drive quality measurement and reporting systems, such as the Health Plan Employer Data and Information Set (HEDIS®). HEDIS is a set of standardized performance measures for comparisons among managed care organizations (Hanchak et al. 1996). Administratively defined cohorts are also

used from quality improvement (Weiner et al. 1990, 1995; Romano, Roos, and Jollis 1993; Garnick, Hendricks, and Comstock 1994; Leatherman et al. 1991) and disease management programs. For example, at-risk individuals may be identified from administrative files to receive reminders for annual mammography or influenza vaccination. The Centers for Medicare and Medicaid Services often use administrative data to identify patients for national Medicare quality improvement projects (Jencks and Wilensky 1992).

Regardless of the condition, several factors limit the accuracy of administrative algorithms for disease identification. These factors include incompleteness of data submitted by providers for capitated visits and procedures (encounters) and fee-for-service procedures (claims) to payors; limited clinical detail in the International Classification of Disease (ICD), Clinical Procedure Terminology (CPT), and Diagnostic Related Group (DRG) systems; and inaccuracy of demographic information in administrative files. For example, administrative pharmacy databases will not contain evidence of treatment if the physician only gives the patient samples from the office and does not write a formal prescription. The patient could also have the prescription filled at a nonparticipating pharmacy without using a pharmacy ID card. Therefore, to use administrative databases effectively for quality improvement and profiling, careful attention must be given to disease identification algorithms (Benesch et al. 1997; Weintraub et al. 1999).

Depression, in particular, presents many additional challenges in the use of health plan administrative data. These problems include failure of the physician to recognize depression (Wells et al. 1989; Kessler, Cleary, and

---

Supported in part by the Academic Medicine and Managed Care Forum, grant nos. HS09446 and HS1112403

Address correspondence to Claire M. Spettell, Ph.D., Research Director, Health Informatics/ USQA, 980 Jolly Road, Mailstop U38A, Blue Bell, PA 19422. Terri C. Wall, M.D., M.P.H., is Assistant Professor of General Pediatrics, University of Alabama at Birmingham, and is with the UAB Center for Outcomes and Effectiveness Research and Education. Jeroan J. Allison, M.D., M.S., Epi., is Associate Professor of Medicine, University of Alabama at Birmingham, and is with the UAB Center for Outcomes and Effectiveness Research and Education, the UAB Division of General Internal Medicine, and the UAB Division of Preventive Medicine. Jaimee Calhoun, M.A., Ed., is Program Coordinator, University of Alabama at Birmingham, UAB Center for Outcomes and Effectiveness Research and Education. Richard Kobylinski is Senior Research Analyst, USQA/Aetna, Inc., Blue Bell, PA. Rachel Fargason, M.D., is with the Department of Psychiatry, University of Alabama at Birmingham. Catarina Kiefe, Ph.D., M.D., is Professor of Medicine and Biostatistics, University of Alabama at Birmingham, and is with the UAB Center for Outcomes and Effectiveness Research and Education, the UAB Division of Preventive Medicine, and the Birmingham Veterans Affairs Medical Center.

Burke 1985; Borus et al. 1988), failure of the physician and patient to report depression because of the stigma of mental illness (Hoyt et al. 1997; Hirschfeld et al. 1997; Rost et al. 1994), and confounding of diagnosis by medical comorbidity (Tylee, Freeling, and Kerry 1993; Epstein et al. 1996; Koenig et al. 1993; Cohen-Cole and Stoudemire 1987; Coulehan et al. 1990). Because of these difficulties, one cannot assume that successful approaches to identifying patients with other diagnoses from claims data may be necessarily extrapolated to depression (Hanchak et al. 1996; Kiefe et al. 2001; Ellerbeck et al. 1995; Marciniak et al. 1998). However, the high prevalence of depression (Simon, Von Korff, and Barlow 1995; Simon and Von Korff 1995; Henderson and Pollard 1992; Hall and Wise 1995) and the well-documented deficiencies in the diagnosis and treatment of depression (Rost et al. 1994; Wells 1994; Rogers et al. 1993; Norquist et al. 1995; Katon et al. 1992; Lemelin et al. 1994; Bouhassira et al. 1998) make this an important area of quality assessment and improvement.

Existing research has not yet successfully addressed the impact of these difficulties. The purpose of this paper is to describe the process we used to evaluate and refine an algorithm for identifying physician-recognized depression using multistate data from a large managed care organization. We also explored the effect of changes in the algorithm on apparent changes in quality of care.

METHODS

Overview

We examined two algorithms that used administrative data to identify patients diagnosed with depression by their primary care physicians. The study sample was drawn from primary care physician member panels of a large managed care organization (MCO). For all comparisons, we used a physician diagnosis of depression in the medical record as the standard. Next, we used administrative data to explore the variability in physician performance on identical quality measures for two contemporaneous patient samples, each constructed from different algorithms. *Algorithm 1* was designed to maximize sensitivity by allowing patients to be identified from either administrative diagnostic codes or pharmacy data. *Algorithm 2* was designed to decrease the false positive identifications that result from the use of pharmacy data alone.

Patient Data and Eligibility Criteria

Administrative data from primary care encounters, specialist encounters, claims, and pharmacy databases were linked with the MCO's membership

and provider files. Although each patient was assigned to a primary care physician, we included treatment and follow-up events from any physician in the plan. We obtained our study samples from the pool of all members aged 12 years and older with medical and pharmacy benefits who were continuously enrolled as members in the MCO's health plans in the Mid-Atlantic or Northeast regions of the United States between January 1, 1997, and December 31, 1997 ($n = 892{,}786$).

Disease-Identification Algorithms

Algorithm 1 relied on a combination of diagnostic and pharmacy codes from administrative databases. The ICD-9 codes identified depressive disorders (296.2–296.36; 300.4; 311). Bipolar affective disorder (i.e., manic disorders) and depression with psychosis were specifically excluded. Pharmacy codes included: (1) monoamine oxidase inhibitors (MAOIs), (2) tricyclic antidepressants, (3) tetracyclic antidepressants, (4) selective serotonin reuptake inhibitors (SSRIs), (5) serotonin 2-receptor antagonists, (6) alpha-2-receptor antagonists, and (7) other miscellaneous antidepressants (e.g., modified cyclics). Patients less than 19 years of age with prescriptions for imipramine were excluded because imipramine may be used to treat enuresis or attention deficit hyperactivity disorder. Patients with prescriptions for lithium were also excluded.

Algorithm 1 required that members have at least two events in the administrative data with each event satisfying one of the following criteria: (1) an outpatient encounter with a primary diagnosis of depression or (2) a pharmacy claim for an antidepressant medication. Algorithm 1 could be satisfied by two events from the same category. Because a diagnostic code was not required, some members may have been identified because they filled two or more prescriptions for antidepressants. One event (either encounter diagnosis or pharmacy claim) for depression was not considered sufficient for identification because of the potential for miscoding or the use of another family member's pharmacy card when filling prescriptions.

Algorithm 2 required that one of the two events from Algorithm 1 actually be an encounter with a diagnosis of depression. The other event could either be another encounter with a diagnosis of depression or a pharmacy claim for an antidepressant medication.

Protocol for Medical Record Identification and Review

Our protocol for medical record abstraction, which has been published elsewhere, included a computerized abstraction module with extensive

instructions and synonym documentation, detailed abstractor training with an instruction manual, careful attention to data security and tracking of the records, and quality assurance (Allison, Wall et al. 2000). The study protocol was approved separately by the Institutional Review Boards of the University of Alabama at Birmingham and the MCO. Abstractors were trained to protect patient confidentiality. Medical record review was performed by the MCO, and no member-identifying information was released to the collaborating academic institution.

The original algorithm (Algorithm 1) was used to identify patients with and without depression for inclusion in the medical record sample. In defining the sample for medical record review, we focused on patients with a new diagnosis of depression. Therefore, we only included members that met the criteria of Algorithm 1 during our 12-month study window and for whom the MCO had no record of depression-related treatment in the 12 months preceding January 1, 1997. We used a stratified random sampling methodology that matched depressed and nondepressed patients within brackets of age, gender, and number of comorbid medical conditions.

From administrative data, 892,786 members were eligible for the study in the Mid-Atlantic and Northeast regions. Of these, 53,170 patients met criteria for depression by Algorithm 1 using administrative data, and the remaining 839,616 patients did not meet criteria. Based upon the stratified randomized methodology described above, we abstracted the charts of 234 patients with depression and 231 patients without depression. The time frame of abstraction was from July 1, 1996, to December 31, 1997. Approximately 10 percent of all medical records were dually abstracted with an overall interrater agreement of at least 95 percent for all main variables. In particular, interrater agreement for physician-recorded diagnosis of depression was 98 percent.

Diagnosis of Depression

The clinical standard for the diagnosis of depression is based on the *Diagnostic and Statistical Manual of Mental Disorders, Version IV* (American Psychiatric Association 1994). The DSM-IV criteria are traditionally ascertained through a structured medical interview (Spitzer et al. 1992; Robins et al. 1981). A diagnosis of major depression requires the presence of one of the primary symptoms of depression (depressed mood or anhedonia) plus four additional symptoms for more than two weeks (American Psychiatric Association 1994). Additional symptoms include impaired concentration or cognitive dulling, thoughts of suicide, loss of energy/fatigue, altered appetite or weight change,

feelings of worthlessness and guilt, disturbances of sleep, and psychomotor retardation or agitation.

Because quality measurements are not currently based on DSM-diagnosed depression, we used the standard of physician-recognized depression determined by documentation in the medical record.

Quality Measures

One purpose of quality measures is to accurately capture the essence of evidence-based clinical guidelines in a quantitative fashion, allowing large amounts of data to be processed for improving delivery of medical care (Weissman et al. 1999; Hofer et al. 1997; Turpin et al. 1996; Harr, Balas, and Mitchell 1996). Important foundations for quality measures include: (1) strength of supportive evidence (evidence obtained from multiple randomized controlled trials given greatest emphasis); (2) consensus of professional societies about targeted intervention; (3) existence of a performance gap with documented need to improve care; (4) ability to improve care based upon quality measure, after consideration of practical resource constraints; (5) availability of adequate and economically feasible data sources; and (6) the severity and consequences of the underlying condition.

We applied the above principles to the 1993 guidelines on treatment of depression issued by the Agency for Health Care Policy and Research (AHCPR), currently the Agency for Healthcare Research and Quality (AHRQ) (1993). The guidelines divide treatment of depression into the acute, continuation, and maintenance phases. The goal of the acute treatment phase is to achieve remission, that is, to remove all signs and symptoms of the current episode of depression and to restore psychosocial and occupational functioning. Continuation treatment is intended to prevent relapse. Recovery is achieved when the patient has been asymptomatic for at least four to nine months, at which time the clinician may consider tapering or stopping antidepressant medication under certain circumstances. Maintenance treatment prevents subsequent episodes in those at risk for recurrence. We developed three measures specifically for the acute phase (adequate follow-up, medication adherence, minimum medication dosage), two measures specifically for the continuation phase (medication adherence and minimum medication dosage), and one measure for both the acute and continuation phase (adequate trial before switching medications). Table 1 provides a more detailed definition and rationale for each measure.

Table 1: Quality Measures Based on the 1993 Guidelines for Treatment of Major Depression by the Agency for Health Care Policy and Research

Measure	*Data Source*	*Criteria*	*Guideline-Based Rationale*
1. Follow-up (acute phase)	Administrative (claims/ encounter)	Proportion of patients with at least 1 visit within 6 weeks of initial antidepressant medication	After initial diagnosis, patients should be seen frequently until symptoms resolve, then every 4 to 12 weeks.
2. Medication adherence (acute phase)	Pharmacy	Proportion of patients filling at least 90 days of therapy during the 118 days from the first antidepressant fill	Lack of adherence to medication is associated with worse outcomes. Physicians should promote patient adherence.
3. Medication adherence (chronic phase)	Pharmacy	Proportion of patients filling at least 120 days of therapy during the 155 days from the first antidepressant fill	Medication should be continued for 4 to 9 months after onset of remission.
4. Minimum medication dose (acute phase)	Pharmacy	Proportion of patients started on at least the minimum therapeutic dose approved by Federal Drug Administration	When prescribed, medication should be administered in dosages shown to alleviate symptoms.
5. Minimum medication dose (chronic phase)	Pharmacy	Proportion of patients continued on at least the minimum therapeutic dose approved by Federal Drug Administration	For continuation treatment, medication should be prescribed at the same dosage necessary to control symptoms in the acute phase.
6. Adequate trial before switching (either phase)	Pharmacy	Proportion of patients receiving a minimum 25-day trial of initial antidepressant before receiving new antidepressant exclusive of low dose tricyclic or trazadone added for insomnia	When prescribed, medication should be continued for a sufficient length of time to permit a reasonable assessment of response, generally 4 to 6 weeks.

Analyses

Bivariate comparisons were made with the chi-square statistic for dichotomous variables and the t-test for continuous variables (Rosner 1995).

We first compared demographic characteristics and comorbidities for abstracted cases by the presence of algorithm-defined depression. Next, we examined the DSM depression symptoms for patients with and without physician-recognized depression. From the sample of 465 abstracted medical records, we compared the operating characteristics (true/false positives, true/false negatives, sensitivity, specificity, and predictive values) of each algorithm, taking physician-diagnosed depression as the standard. We first derived predictive values based on the prevalence of depression in our sample. We then examined change in predictive values over a broad prevalence range of physician-recognized depression.

Finally, we identified patients of all physicians who had a minimum of 10 members with depression by Algorithm 1 ($n = 32{,}819$ patients) and Algorithm 2 ($n = 6{,}837$ patients). Using administrative data, we then compared aggregate physician performance on each of the six quality measures for both patient samples.

RESULTS

Although the ages of the patients with and without depression were similar, a higher proportion of the patients with depression were female. In addition, patients with depression tended to have more comorbidities (Table 2).

DSM-IV Diagnosis of Depression

The diagnosis of depression was recorded in 121 of the 465 abstracted medical records. Among those 121 medical records, only 9 percent contained

Table 2: Patient Characteristics by Presence of Algorithm-Defined Depression[a]

	Algorithm Positive for Depression	*Algorithm Negative for Depression*
N	234	231
Female	70%	49%
Median age (years)	44	41
Number of comorbid conditions	1.3	0.79

[a]Members were identified from 1997 Managed Care Organization administrative data by Algorithm 1; reported data ($n = 465$) were obtained from chart review of stratified random subset.

Table 3: Physician Documentation of Depressive Symptoms[a]

	Physician-Recognized Depression from Medical Record	
	Yes	*No*
N	121	344
Depressed mood	52.9%	3.5%
Anhedonia	12.4%	0.3%
Worthlessness/guilt	7.4%	0%
Loss of energy/fatigue	27.3%	9.9%
Psychomotor retardation/agitation	5.8%	1.7%
Cognitive dulling	11.6%	2.0%
Appetite/weight change	28.1%	11.3%
Suicidal thoughts	4.1%	0.3%
Insomnia/hypersomnia	36.4%	5.5%
Formal diagnosis by symptom criteria[b]	9.1%	0%

[a]Members were identified from 1997 Managed Care Organization administrative data by Algorithm 1; reported data ($n = 465$) were obtained from chart review of stratified random subset.
[b]Diagnosis of depression from Diagnostic and Statistical Manual of Mental Disorders, Version IV (DSM-IV).

documentation that met the American Psychiatric Association's rigorous DSM-IV criteria for diagnosis of depression (Table 3).

Comparison of Disease-Identification Algorithms

Table 4 gives the operating characteristics of both algorithms. Algorithm 1, which required any two depression-related events (diagnosis or medication claim) in the administrative data, had a sensitivity of 95 percent and a specificity of 65 percent. The sensitivity was quite high because there were few cases where the algorithm identified a member as not depressed but the medical record indicated a diagnosis of depression (false negative cases). Of the 234 cases identified with depression by Algorithm 1, 54 percent had no diagnosis of depression in administrative data and were identified from antidepressant use only. As expected, there was a high rate of antidepressant use among the false positive cases (84 percent). Compared to Algorithm 1, Algorithm 2, which required a diagnosis of depression, yielded a higher specificity (88 percent) and lower sensitivity (52 percent). The rate of antidepressant use among false positive cases was less at 53 percent.

Unlike sensitivity and specificity, predictive values depend upon the prevalence of the index condition in the population. The method we used to

Table 4: Depression Identification Algorithm Operating Characteristics[a]

Patient Subset/Characteristic	*Algorithm 1*	*Algorithm 2*
Algorithm "depressed," *n* (%)		
• True positive	115 (25)	63 (14)
• False positive	119 (26)	40 (9)
Algorithm "non-depressed," *n* (%)		
• True negative	225 (48)	304 (65)
• False negative	6 (1)	58 (12)
All patients, %		
• Sensitivity	95.0	52.1
• Specificity	65.4	88.4
• Positive predictive value[b]	49.1	60.6
• Negative predictive value[b]	97.4	84.0

[a]Members were identified from 1997 Managed Care Organization administrative data by Algorithm 1; reported data ($n = 465$) were obtained from chart review of stratified random subset; comparison standard: physician diagnosis of depression in medical record.

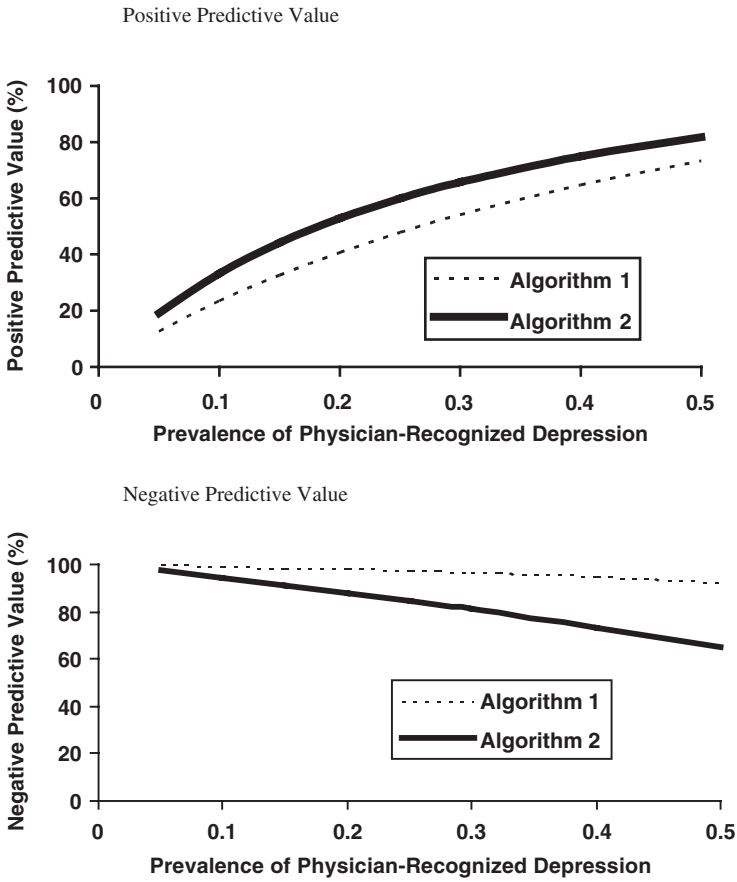
[b]Based upon 26% prevalence of physician-recognized depression as determined by sampling plan; because the high prevalence of physician-recognized depression is related to the sampling strategy of this study, predictive values over a range of lower prevalence rates are presented in Figures 1A and 1B.

construct our sample yielded a prevalence of 26 percent (121/465) for physician-recognized depression. Based upon this prevalence, the positive predictive values for Algorithms 1 and 2 were 49 percent and 61 percent, respectively, and the negative predictive values were 97 percent and 84 percent, respectively (Table 4). Figures 1A and 1B depict how the predictive values of both algorithms vary as the prevalence of physician-recognized depression varies. In the prevalence range of 5–10 percent, the positive predictive values of both algorithms remain below 33 percent, and negative predictive values remain above 94 percent. With a prevalence of 20 percent, the positive predictive values remain below 53 percent and the negative predictive values above 88 percent.

Variation in Physician Performance Profiles by Disease-Identification Algorithm

Using administrative medical and pharmacy data, we examined variation in office-level performance on six quality measures for Algorithms 1 and 2. Algorithm 1 identified 32,819 members as depressed, compared to 6,837 members identified by Algorithm 2. The number of primary care provider offices with a minimum of 10 depressed patients decreased from 1,756 with Algorithm 1 to 414 with Algorithm 2.

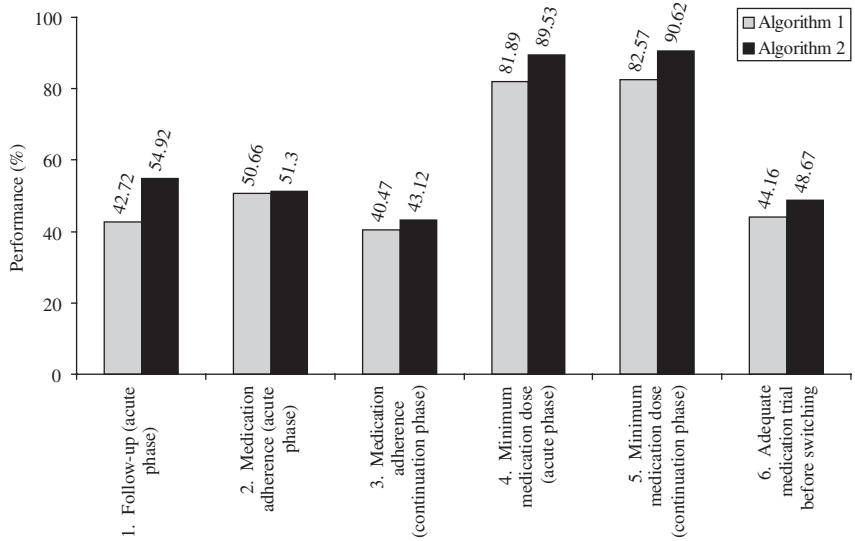
Figure 1: Predictive Values of Two Algorithms for Identifying Physician-Recognized Depression by Prevalence[a]



[a]Members were identified from 1997 Managed Care Organization administrative data by Algorithm 1; reported data (n = 465) were obtained from chart review of stratified random subset.

There were significant differences in quality performance reflected by the two algorithms (Figure 2). More specifically, for Algorithm 2, there were significantly higher rates of follow-up visits after initiation of antidepressant medication, appropriate dosage and medication adherence, and appropriate medication trial before switching to another antidepressant.

Figure 2: Mean Primary Care Physician Quality Performance by Disease-Identification Algorithm[a]



[a]Members were identified from 1997 Managed Care Organization administrative data separately with Algorithm 1 (n = 32,819) and Algorithm 2 (n = 6,837); quality measures are defined in Table 1 and based on administrative data; all comparisions are significant at $p < 0.05$.

DISCUSSION

Our work demonstrates the difficulty in identifying patients with depression from administrative data. Recognizing that depression is underreported in administrative data, we specifically designed Algorithm 1 to explore the effects of using pharmacy codes as primary identifiers of depression. The pharmacy inclusion criteria for Algorithm 1 were broad, thus capturing more members (increased sensitivity) at the expense of generating more false positive diagnoses (decreased specificity and positive predictive value). The more stringent Algorithm 2, which required a diagnosis of depression, had a better specificity, but much lower sensitivity.

Both algorithms suffered from low positive predictive value, and thus frequently falsely classify patients as having depression. Predictive value depends upon the prevalence of the underlying condition. Only in highly selected primary care patient populations does the prevalence of depression approach 20 percent (Pearson et al. 1999), corresponding to a positive

predictive value of less than 53 percent for both algorithms. One study found the prevalence to range between 5 to 10 percent for unselected elderly patients (Barry et al. 1998), corresponding to a positive predictive value of less than 33 percent for both algorithms. This means that if administrative data were used to derive quality measures for depression, only 33 percent of those patients in the denominator would actually have a physician diagnosis of depression by chart review.

These findings are especially important given the close relationship of our quality measures to the HEDIS measure for Antidepressant Medication Management. Linking Algorithm 2 with quality measures 1, 2, and 3 approximates the current HEDIS technical specifications (Allison, Wall et al. 2000). Algorithm 2 uses the same pharmacy and ICD-9 codes as HEDIS for denominator construction. Both approaches require each member to have at least 12 months of continuous enrollment in a managed care plan with pharmacy benefits.

However, there are some differences between our quality measures and the HEDIS measure. Our quality measures were intended to reflect the 1993 AHCPR guidelines and not to duplicate the HEDIS Antidepressant Medication Management measure, which was in draft format at the time of data collection for this study. Similar to the HEDIS measure, we required a new diagnosis of depression because many of our quality measures apply to the acute phase of depression. In contrast to the HEDIS measure, which allows either a primary or secondary diagnosis of depression, we required a primary diagnosis.

Quality measure 1 differs from the corresponding HEDIS measure by requiring one visit to a primary care provider within 6 weeks of diagnosis, whereas the HEDIS measure requires three visits within 12 weeks. We made our criteria for follow-up more lenient because, in certain cases, telephone contact without an office visit is appropriate and would not be reflected in the administrative data. We included two measures of antidepressant medication adherence. Our measure 2, Adherence during the Acute Phase of Treatment, is similar to the HEDIS measure, as both reflect adherence during the first three months of treatment, allowing for gaps in medication supply. Our measure 3, Adherence during the Maintenance Phase of Treatment, examined adherence within the first four months of treatment, while the HEDIS measure examines adherence during the first six months. We constructed measure 3 to reflect adherence to the minimum recommended by the 1993 AHCPR guideline (i.e., four months). Our work also reveals important differences in quality measurement according to which algorithm was used to define the denominator. Algorithm 1, associated with a positive

predictive value and corresponding higher false positive rate, led to lower rates for each quality measure. In fact, three of the measures differed by 7–8 percent. Such underreporting of quality performance is important and may lead to loss of credibility in provider feedback with crippling of improvement efforts (Allison, Calhoun et al. 2000). Therefore, when planning a quality improvement project, positive predictive value is probably the most important operating characteristic of a disease-identification algorithm.

The number of identified members, which increases with the sensitivity of the algorithm, has implications when generating performance profiles. Creating valid physician profiles requires sufficient patient numbers. In this study, the impact of requiring a diagnostic code to identify members with depression reduced the eligible member population by two-thirds. Consequently, fewer practices meet minimum volume criteria for individualized performance profiles.

Even beyond the difficulties imposed by administrative data, several factors contribute to the diagnostic challenges of depression. Although depression affects up to 10 percent of the U.S. population at an estimated annual cost of $44 billion (Hall and Wise 1995) and produces impairment in quality of life similar to that of other serious chronic diseases (Wells, Stewart et al. 1989), the diagnosis is often missed by physicians. Primary care physicians recognize only about one-half of all depressed patients in the outpatient setting (Wells, Hays et al. 1989; Kessler, Cleary, and Burke 1985; Borus et al. 1988). The detection rate by primary care physicians falls to 30 percent for patients with significant medical comorbidity (Tylee, Freeling, and Kerry 1993). This may result from physicians attributing signs and symptoms of depression to other medical illness. Somatic symptoms used in making the diagnosis of depression (e.g., fatigue, sleep disturbance, weight loss) are also presenting features of many other medical illnesses. Subjective symptoms such as depressed mood and anhedonia may also be inappropriately regarded as an understandable reaction to illness.

Concern about patient confidentiality and the potential for jeopardizing reimbursement and other benefits may also lead physicians to deliberately substitute alternative diagnoses on claims and encounters. In a survey of 440 primary care physicians randomly selected from the membership lists of professional organizations, 50 percent of respondents reported substituting another diagnostic code in the prior two weeks for one or more patients who met the criteria for major depression (Rost et al. 1994). Physicians may underreport depression to protect patients from social stigma and possible occupational and legal consequences (Hoyt et al. 1997;

Hirschfeld et al. 1997). For example, medical records are often subpoenaed during custody hearings. In addition, many physicians may be uncertain about making such diagnoses because of limited training with behavioral health disorders. As a result, valid cases of depression are not identified by physicians, let alone by algorithms based on administrative data. Furthermore, patients identified from administrative data may represent more severe cases (Valenstein et al. 2000).

Some patients reluctantly express psychological symptoms and may deny mood changes. These patients may present instead with a variety of nonspecific somatic complaints such as headache, abdominal pain, insomnia, weight loss, or low energy. This symptom overlap leads to a complex interaction between depression and medical comorbidity. Therefore, medical illness frequently presents as depression and depression as other medical illness. This problem is especially troubling in the elderly, who suffer from a higher burden of comorbidity (Coulehan et al. 1990).

The overlap in treatment of depression and other medical conditions makes identification of depressed patients from pharmacy data difficult. Antidepressants are now used for a wide variety of diseases other than depression such as chronic pain, neuropathic pain, fibromyalgia, chronic fatigue syndrome, migraine and tension headaches, irritable bowel disease, premenstrual dysphoric disorder, insomnia, eating disorders, premature ejaculation, panic disorder, post-traumatic stress disorder, social phobia, and anger attacks (Barkin et al. 1996a, 1996b; Compas et al. 1998; Davies et al. 1996; Fishbain et al. 1998; Keck and McElroy 1997; McQuay et al. 1996; Merskey 1997; Metz and Pryor 2000; Moreland and St. Clair 1999; Pappagallo 1999; Simon and Von Korff 1997). These syndromes have variable overlap with clinical depression, and it is often difficult to discern which problem is primary (Keck and McElroy 1997; Clarke 1998). Often a corresponding ICD-9 diagnosis of depression cannot be found when an antidepressant prescription is written at an office visit. Given the tolerability and safety of the newer antidepressants, clinicians may tend to prescribe these agents for nonspecific psychiatric symptoms or behavioral problems (e.g., stress) without a making a clear diagnosis. As a result, there is concern about inappropriate use of antidepressants (Bouhassira et al. 1998).

The use of antidepressants for disorders other than depression and the treatment of depression without a diagnosis are both reflected in the rate of antidepressant use among members with a false positive diagnosis. Algorithm 2 produced fewer cases of false positive identification and, even among the

false positive cases, the rate of antidepressant use was lower for Algorithm 2 (53 percent versus 84 percent).

The limited observation period available through administrative databases of health plans is both a strength and limitation. Administrative data permits longitudinal observation at the member level, unlike certain other data types. However, annual member disenrollment averaged 29 percent in 1999 for HMOs reporting to NCQA's *Quality Compass 2000* (2000). This limits identification of new cases. Although recently enrolled members may appear to be newly identified with a disease, it is possible that the disease is long-standing. Variation in quality performance for new cases of major depression may partially result from undetected variation in disease chronicity. Katon did not find important differences in antidepressant treatment patterns in a staff-model HMO after adjusting for multiple factors, including prior history of depressive episodes (Katon et al. 2000).

Our paper raises the need for caution in interpreting quality measures based on administrative data. For example, we found that rates of appropriate antidepressant treatment (e.g., follow-up visits, appropriate dosage) were substantially higher when the specifications for the member population required a diagnosis of depression. Likewise, Kerr found that variations in the specifications of quality-of-care measures for depression treatment influenced conclusions about the adequacy of antidepressant prescribing patterns in two managed care practices (Kerr et al. 2000). Kerr varied the definition of a new episode of depression (four-month versus nine-month clean period) and the minimum number of visits with a diagnosis of depression. Patients with two or more visits coded for depression were more likely to receive antidepressants at the appropriate dosage than those with only one visit coded for depression. This may, in part, result because increased algorithm specificity from the requirement of a diagnosis code may lead to the identification of patients with more severe depression as opposed to a temporary crisis or generalized stress.

Our study also has specific implications for the training of primary care physicians. We found that primary care physicians who documented a diagnosis of depression in the medical record rarely documented the symptoms required to make that diagnosis using the DSM-IV criteria for major depression. This finding may reflect poor documentation rather than poor interviewing skills. Other studies suggest that interviewing style of the primary care physician is related to the recognition of depression (Badger et al. 1994; Robbins et al. 1994). Training and feedback based upon medical record review has been shown to increase both recognition of depression and documentation of symptoms (Linn and Yager 1980). In addition, automated

office screening tools have been shown to increase recognition of depression without placing excessive demands on physicians (Zung et al. 1983; Moore, Silimperi, and Bobula 1978; Hoeper et al. 1984; Magruder-Habib, Zung, and Feussner 1990).

CONCLUSIONS

We found that accurate identification of patients with physician-recognized depression from administrative data poses significant difficulty. In addition, observed quality varied significantly with algorithm operating characteristics, with lower observed quality being associated with lower algorithm specificity and a greater number of members being falsely identified as having depression. This suggests that low-quality performance may, in part, be attributed to the specific algorithms used to identify the study population. Low specificity, and the associated false classification of patients as having depression, may inappropriately lower quality performance scores and decrease confidence in performance feedback.

REFERENCES

Agency for Health Care Policy and Research. 1993. *Depression in Primary Care: Treatment of Major Depression.* Agency for Health Care Policy and Research, Rockville, MD.

Allison, J. J., J. W. Calhoun, T. C. Wall, C. M. Spettell, C. A. Fargason, N. W. Weissman, and C. I. Kiefe. 2000. "Optimal Reporting of Health Care Process Measures: Inferential Statistics as Help or Hindrance?" *Managed Care Quarterly* 8 (4): 1–10.

Allison, J. J., T. C. Wall, C. M. Spettell, J. Calhoun, C. A. Fargason, R. Kobylinski, R. Farmer, and C. I. Kiefe. 2000. "The Art and Science of Chart Review." *Joint Commission Journal on Quality Improvement* 26 (3): 115–36.

American Psychiatric Association. 1994. *Diagnostic and Statistical Manual of Mental Disorders, 4th Edition.* American Psychiatric Association, Washington, DC.

Badger, L. W., F. V. deGruy, J. Hartman, M. A. Plant, J. Leeper, R. Ficken, A. Maxwell, E. Rand, R. Anderson, and B. Templeton. 1994. "Psychosocial Interest, Medical Interviews, and the Recognition of Depression." *Archives of Family Medicine* 3 (10): 899–907.

Barkin, R. L., T. R. Lubenow, S. Bruehl, B. Husfeldt, O. Ivankovich, and S. J. Barkin. 1996a. "Management of Chronic Pain. Part I." *Disease-A-Month* 42 (7): 389–454.

———. 1996b. "Management of Chronic Pain. Part II." *Disease-A-Month* 42 (8): 457–507.

Barry, K. L., M. F. Fleming, L. B. Manwell, L. A. Copeland, and S. Appel. 1998. "Prevalence of and Factors Associated with Current and Lifetime Depression in Older Adult Primary Care Patients." *Family Medicine* 30 (5): 366–71.

Benesch, C., D. M. Witter Jr., A. L. Wilder, P. W. Duncan, G. P. Samsa, and D. B. Matchar. 1997. "Inaccuracy of the International Classification of Diseases (ICD-9-CM) in Identifying the Diagnosis of Ischemic Cerebrovascular Disease." *Neurology* 49 (3): 660–4.

Borus, J. F., M. J. Howes, N. P. Devins, R. Rosenberg, and W. W. Livingston. 1988. "Primary Health Care Providers' Recognition and Diagnosis of Mental Disorders in Their Patients." *General Hospital Psychiatry* 10 (5): 317–21.

Bouhassira, M., M. P. Allicar, C. Blachier, A. Nouveau, and F. Rouillon. 1998. "Which Patients Receive Antidepressants? A 'Real World' Telephone Study." *Journal of Affective Disorders* 49 (1): 19–26.

Clarke, D. M. 1998. "Psychological Factors in Illness and Recovery." *New Zealand Medical Journal* 111 (1076): 410–2.

Cohen-Cole, S. A., and A. Stoudemire. 1987. "Major Depression and Physical Illness: Special Considerations in Diagnosis and Biologic Treatment." *Psychiatric Clinics of North America* 10 (1): 1–17.

Compas, B. E., D. A. Haaga, F. J. Keefe, H. Leitenberg, and D. A. Williams. 1998. "Sampling of Empirically Supported Psychological Treatments from Health Psychology: Smoking, Chronic Pain, Cancer, and Bulimia Nervosa." *Journal of Consulting and Clinical Psychology* 66 (1): 89–112.

Coulehan, J. L., H. C. Schulberg, M. R. Block, J. E. Janosky, and V. C. Arena. 1990. "Medical Comorbidity of Major Depressive Disorder in a Primary Medical Practice." *Archives of Internal Medicine* 150 (11): 2363–7.

Davies, H. T., I. K. Crombie, W. A. Macrae, K. M. Rogers, and J. E. Charlton. 1996. "Audit in Pain Clinics: Changing the Management of Low-Back and Nerve-Damage Pain." *Anaesthesia* 51 (7): 641–6.

Ellerbeck, E. F., S. F. Jencks, M. J. Radford, T. F. Kresowik, A. S. Craig, J. A. Gold, H. M. Krumholz, and R. A. Vogel. 1995. "Quality of Care for Medicare Patients with Acute Myocardial Infarction. A Four-State Pilot Study from the Cooperative Cardiovascular Project." *Journal of the American Medical Association* 273 (19): 1509–14.

Epstein, S. A., J. J. Gonzales, J. S. Onge, J. Carter-Campbell, K. Weinfurt, M. Leibole, and R. L. Goldberg. 1996. "Practice Patterns in the Diagnosis and Treatment of Anxiety and Depression in the Medically Ill: A Survey of Psychiatrists." *Psychosomatics* 37 (4): 356–67.

Fishbain, D. A., R. B. Cutler, H. L. Rosomoff, and R. S. Rosomoff. 1998. "Do Antidepressants Have an Analgesic Effect in Psychogenic Pain and Somatoform Pain Disorder? A Metaanalysis." *Psychosomatic Medicine* 60 (4): 503–9.

Garnick, D. W., A. M. Hendricks, and C. B. Comstock. 1994. "Measuring Quality of Care: Fundamental Information from Administrative Datasets." *International Journal for Quality in Health Care* 6 (2): 163–77.

Hall, R. C., and M. G. Wise. 1995. "The Clinical and Financial Burden of Mood Disorders: Cost and Outcome." *Psychosomatics* 36 (2): S11–8.

Hanchak, N. A., S. R. Harmon-Weiss, P. D. McDermott, A. Hirsch, and N. Schlackman. 1996. "Medicare Managed Care and the Need for Quality Measurement." *Managed Care Quarterly* 4 (1): 1–12.

Harr, D. S., E. A. Balas, and J. Mitchell. 1996. "Developing Quality Indicators as Educational Tools to Measure the Implementation of Clinical Practice Guidelines." *American Journal of Medical Quality* 11 (4): 179–85.

Henderson, J. G. Jr., and C. A. Pollard. 1992. "Prevalence of Various Depressive Symptoms in a Sample of the General Population." *Psychological Reports* 71 (1): 208–10.

Hirschfeld, R. M., M. B. Keller, S. Panico, B. S. Arons, D. Barlow, F. Davidoff, J. Endicott, J. Froom, M. Goldstein, J. M. Gorman, R. G. Marek, T. A. Maurer, R. Meyer, K. Phillips, J. Ross, T. L. Schwenk, S. S. Sharfstein, M. E. Thase, and R. J. Wyatt. 1997. "The National Depressive and Manic-Depressive Association Consensus Statement on the Undertreatment of Depression." *Journal of the American Medical Association* 277 (4): 333–40.

Hoeper, E. W., G. R. Nycz, L. G. Kessler, J. D. Burke Jr., and W. E. Pierce. 1984. "The Usefulness of Screening for Mental Illness." *Lancet* 1 (8367): 33–5.

Hofer, T. P., S. J. Bernstein, P. A. Hayward, and S. DeMonner. 1997. "Validating Quality Indicators for Hospital Care." *Joint Commission Journal on Quality Improvement* 23 (9): 455–67.

Hoyt, D. R., R. D. Conger, J. G. Valde, and K. Weihs. 1997. "Psychological Distress and Help Seeking in Rural America." *American Journal of Community Psychology* 25 (4): 449–70.

Jencks, S. F., and G. R. Wilensky. 1992. "The Health Care Quality Improvement Initiative: A New Approach to Quality Assurance in Medicare." *Journal of the American Medical Association* 268 (7): 900–3.

Katon, W., M. Von Korff, E. Lin, T. Bush, and J. Ormel. 1992. "Adequacy and Duration of Antidepressant Treatment in Primary Care." *Medical Care* 30 (1): 67–76.

Katon, W., C. M. Rutter, E. Lin, G. Simon, M. Von Korff, T. Bush, E. Walker, and E. Ludman. 2000. "Are There Detectable Differences in Quality of Care or Outcome of Depression across Primary Care Providers?" *Medical Care* 38 (6): 552–61.

Keck, P. E. Jr., and S. L. McElroy. 1997. "New Uses for Antidepressants: Social Phobia." *Journal of Clinical Psychiatry* 58 (14, supplement): 32–6; discussion 37–8.

Kerr, E. A., E. A. McGlynn, K. A. Van Vorst, and S. L. Wickstrom. 2000. "Measuring Antidepressant Prescribing Practice in a Health Care System Using Administrative Data: Implications for Quality Measurement and Improvement." *Joint Commission Journal on Quality Improvement* 26 (4): 203–16.

Kessler, L. G., P. D. Cleary, and J. D. Burke Jr. 1985. "Psychiatric Disorders in Primary Care: Results of a Follow-up Study." *Archives of General Psychiatry* 42 (6): 583–7.

Kiefe, C. I., J. J. Allison, O. D. Williams, S. D. Person, M. T. Weaver, and N. W. Weissman. 2001. "Improving Quality Improvement Using Achievable Benchmarks for Physician Feedback." *Journal of the American Medical Association* 258 (22): 2871–9.

Koenig, H. G., H. J. Cohen, D. G. Blazer, K. R. Krishnan, and T. E. Sibert. 1993. "Profile of Depressive Symptoms in Younger and Older Medical Inpatients with Major Depression." *Journal of the American Geriatrics Society* 41 (11): 1169–76.

Leatherman, S., E. Peterson, L. Heinen, and L. Quam. 1991. "Quality Screening and Management Using Claims Data in a Managed Care Setting." *Quality Review Bulletin* 17 (11): 349–59.

Lemelin, J., S. Hotz, R. Swensen, and T. Elmslie. 1994. "Depression in Primary Care. Why Do We Miss the Diagnosis?" *Canadian Family Physician* 40: 104–8.

Linn, L. S., and J. Yager. 1980. "The Effect of Screening, Sensitization, and Feedback on Notation of Depression." *Journal of Medical Education* 55 (11): 942–9.

McQuay, H. J., M. Tramer, B. A. Nye, D. Carroll, P. J. Wiffen, and R. A. Moore. 1996. "A Systematic Review of Antidepressants in Neuropathic Pain." *Pain* 68 (2–3): 217–27.

Magruder-Habib, K., W. W. Zung, and J. R. Feussner. 1990. "Improving Physicians' Recognition and Treatment of Depression in General Medical Care: Results from a Randomized Clinical Trial." *Medical Care* 28 (3): 239–50.

Marciniak, T. A., E. F. Ellerbeck, M. J. Radford, T. F. Kresowik, J. A. Gold, H. M. Krumholz, C. I. Kiefe, R. M. Allman, R. A. Vogel, and S. Jencks. 1998. "Improving the Quality of Care for Medicare Patients with Acute Myocardial Infarction: Results from the Cooperative Cardiovascular Project." *Journal of the American Medical Association* 279 (17): 1351–7.

Merskey, H. 1997. "Pharmacological Approaches Other Than Opioids in Chronic Non-Cancer Pain Management." *Acta Anaesthesiologica Scandinavica* 41 (1, part 2): 187–90.

Metz, M. E., and J. L. Pryor. 2000. "Premature Ejaculation: A Psychophysiological Approach for Assessment and Management." *Journal of Sex and Marital Therapy* 26 (4): 293–320.

Moore, J. T., D. R. Silimperi, and J. A. Bobula. 1978. "Recognition of Depression by Family Medicine Residents: The Impact Of Screening." *Journal of Family Practice* 7 (3): 509–13.

Moreland, L. W., and E. W. St. Clair. 1999. "The Use of Analgesics in the Management of Pain in Rheumatic Diseases." *Rheumatic Diseases Clinics of North America* 25 (1): 153–91, vii.

Norquist, G., K. B. Wells, W. H. Rogers, L. M. Davis, K. Kahn, and R. Brook. 1995. "Quality of Care for Depressed Elderly Patients Hospitalized in the Specialty Psychiatric Units or General Medical Wards." *Archives of General Psychiatry* 52 (8): 695–701.

Pappagallo, M. 1999. "Aggressive Pharmacologic Treatment of Pain." *Rheumatic Diseases Clinics of North America* 25 (1): 193–213, vii.

Pearson, S. D., D. J. Katzelnick, G. E. Simon, W. G. Manning, C. P. Helstad, and H. J. Henk. 1999. "Depression among High Utilizers of Medical Care." *Journal of General Internal Medicine* 14 (8): 461–8.

"Quality Compass 2000 Shows Steady Improvement in HEDIS Performance Data." 2000. *Capitation Rates and Data* 5 (10): 112–5.

Robbins, J. M., L. J. Kirmayer, P. Cathebras, M. J. Yaffe, and M. Dworkind. 1994. "Physician Characteristics and the Recognition of Depression and Anxiety in Primary Care." *Medical Care* 32 (8): 795–812.

Robins, L. N., J. E. Helzer, J. Croughan, and K. S. Ratcliff. 1981. "National Institute of Mental Health Diagnostic Interview Schedule: Its History, Characteristics, and Validity." *Archives of General Psychiatry* 38 (4): 381–9.

Rogers, W. H., K. B. Wells, L. S. Meredith, R. Sturm, and M. A. Burnam. 1993. "Outcomes for Adult Outpatients with Depression under Prepaid or Fee-for-Service. Financing." *Archives of General Psychiatry* 50 (7): 517–25.

Romano, P. S., L. L. Roos, and J. G. Jollis. 1993. "Adapting a Clinical Comorbidity Index for Use with ICD-9-CM Administrative Data: Differing Perspectives." *Journal of Clinical Epidemiology* 46 (10): 1075–9; discussion 1081–90.

Rosner, B. 1995. *Fundamentals of Biostatistics*, 4th ed. Belmont, CA: Wadsworth Publishing.

Rost, K., R. Smith, D. B. Matthews, and B. Guise. 1994. "The Deliberate Misdiagnosis of Major Depression in Primary Care." *Archives of Family Medicine* 3 (4): 333–7.

Simon, G. E., and M. Von Korff. 1997. "Prevalence, Burden, and Treatment of Insomnia in Primary Care." *American Journal of Psychiatry* 154 (10): 1417–23.

———. 1995. "Recognition, Management, and Outcomes of Depression in Primary Care." *Archives of Family Medicine* 4 (2): 99–105.

Simon, G. E., M. Von Korff, and W. Barlow. 1995. "Health Care Costs of Primary Care Patients with Recognized Depression." *Archives of General Psychiatry* 52 (10): 850–6.

Spitzer, R. L., J. B. Williams, M. Gibbon, and M. B. First. 1992. "The Structured Clinical Interview for DSM-III-R (SCID) I: History, Rationale, and Description." *Archives of General Psychiatry* 49 (8): 624–9.

Turpin, R. S., L. A. Darcy, R. Koss, C. McMahill, K. Meyne, D. Morton, J. Rodriguez, S. Schmaltz, P. Schyve, and P. Smith. 1996. "A Model to Assess the Usefulness of Performance Indicators." *International Journal for Quality in Health Care* 8 (4): 321–9.

Tylee, A. T., P. Freeling, and S. Kerry. 1993. "Why Do General Practitioners Recognize Major Depression in One Woman Patient Yet Miss It in Another?" *British Journal of General Practice* 43 (373): 327–30.

Valenstein, M., T. Ritsema, L. Green, F. C. Blow, A. Mitchinson, J. F. McCarthy, K. L. Barry, and E. Hill. 2000. "Targeting Quality Improvement Activities for Depression: Implications of Using Administrative Data." *Journal of Family Practice* 49 (8): 721–8.

Weiner, J. P., S. T. Parente, D. W. Garnick, J. Fowles, A. G. Lawthers, and R. H. Palmer. 1995. "Variation in Office-Based Quality. A Claims-Based Profile of Care Provided to Medicare Patients with Diabetes." *Journal of the American Medical Association* 273 (19): 1503–8.

Weiner, J. P., N. R. Powe, D. M. Steinwachs, and G. Dent. 1990. "Applying Insurance Claims Data to Assess Quality Of Care: A Compilation of Potential Indicators." *Qrb. Quality Review Bulletin* 16 (12): 424–38.

Weintraub, W. S., C. Deaton, L. Shaw, E. Mahoney, and D. C. Morris. 1999. "Can Cardiovascular Clinical Characteristics Be Identified and Outcome Models Be Developed from an In-patient Claims Database?" *American Journal of Cardiology* 84 (2): 166–9.

Weissman, N. W., J. J. Allison, C. I. Kiefe, R. M. Farmer, M. T. Weaver, O. D. Williams, I. G. Child, J. H. Pemberton, K. C. Brown, and C. Baker. 1999. "Achievable Benchmarks of Care: The ABCs of Benchmarking." *Journal of Evaluation in Clinical Practice* 5 (3): 269–81.

Wells, K. B. 1994. "Depression in General Medical Settings: Implications of Three Health Policy Studies for Consultation-Liaison Psychiatry." *Psychosomatics* 35 (3): 279–96.

Wells, K. B., R. D. Hays, M. A. Burnam, W. Rogers, S. Greenfield, and J. E. Ware Jr. 1989. "Detection of Depressive Disorder for Patients Receiving Prepaid or Fee-for-Service Care: Results from the Medical Outcomes Study." *Journal of the American Medical Association* 262 (23): 3298–302.

Wells, K. B., A. Stewart, R. D. Hays, M. A. Burnam, W. Rogers, M. Daniels, S. Berry, S. Greenfield, and J. Ware. 1989. "The Functioning and Well-being of Depressed Patients. Results from the Medical Outcomes Study." *Journal of the American Medical Association* 262 (7): 914–9.

Zung, W. W., M. Magill, J. T. Moore, and D. T. George. 1983. "Recognition and Treatment of Depression in a Family Medicine Practice." *Journal of Clinical Psychiatry* 44 (1): 3–6.