

Enhancing Electronic Health Record Measurement of Depression Severity and Suicide Ideation: A Distributed Ambulatory Research in Therapeutics Network (DARTNet) Study

Robert J. Valuck, PhD, RPh, Heather O. Anderson, PhD, Anne M. Libby, PhD, Elias Brandt, BS, Cathy Bryan, MHA, Richard R. Allen, MS, Elizabeth W. Staton, MSTC, David R. West, PhD, and Wilson D. Pace, MD

Background: Depression is a leading cause of morbidity worldwide. The majority of treatment for depression occurs in primary care, but effective care remains elusive. Clinical decision making and comparative studies of real-world antidepressant effectiveness are limited by the absence of clinical measures of severity of illness and suicidality.

Methods: The Distributed Ambulatory Research in Therapeutics Network (DARTNet) was engaged to systematically collect data using the 9-item Patient Health Questionnaire (PHQ-9) at the point of care. We used electronic health records (EHRs) and the PHQ-9 to capture, describe, and compare data on both baseline severity of illness and suicidality and response and suicidality after diagnosis for depressed patients in participating DARTNet practices.

Results: EHR data were obtained for 81,028 episodes of depression (61,464 patients) from 14 clinical organizations. Over 9 months, data for 4900 PHQ-9s were collected from 2969 patients in DARTNet practices (this included 1892 PHQ-9s for 1019 adults and adolescents who had at least one depression diagnosis). Only 8.3% of episodes identified in our depression cohort had severity of illness information available in the EHR. For these episodes, considerable variation existed in both severity of illness (32.05% with no depression, 26.89% with minimal, 19.54% with mild, 12.04% with moderate, and 9.47% with severe depression) and suicidality (69.43% with a score of 0, 22.58% with a score of 1, 4.97% with a score of 2, and 3.02% with a score of 3 on item 9 of the PHQ-9). Patients with an EHR diagnosis of depression and a PHQ-9 (n = 1019) had similar severity but slightly higher suicidality levels compared with all patients for which PHQ-9 data were available. The PHQ-9 showed higher sensitivity for identifying depression response and emergent (after diagnosis) severity and suicidality; 25% to 30% of subjects had some degree of suicidal thought at some point in time according to the PHQ-9.

Conclusions: This study demonstrated the value of adding PHQ-9 data and prescription fulfillment data to EHRs to improve diagnosis and management of depression in primary care and to enable more robust comparative effectiveness research on antidepressants. (J Am Board Fam Med 2012;25:582–593.)

Keywords: Clinical Epidemiology, Depression, Practice-based Research, Practice-based Research Networks, Suicide

Depression is a common, chronic but episodic, and costly condition for which primary care physicians

provide the majority of care.¹ Depression is severe or very severe in approximately 60% of patients and is the leading cause of disability worldwide as

This article was externally peer reviewed.

Submitted 15 February 2011; revised 2 March 2012; accepted 19 March 2012.

From the Department of Clinical Pharmacy, Skaggs School of Pharmacy and Pharmaceutical Sciences (RJV, HOA, AML) and the Department of Family Medicine, School of Medicine (RJV, HOA, AML, EWS, PRW, WDP), University of Colorado, Anschutz Medical Campus, the American Academy of Family Physicians, Denver, CO

(EB); QED Clinical, Inc., dba CINA, Dallas, TX (CB); and Peak Statistical Services, Evergreen, CO (RRA).

Funding: This project was funded under contract no. HHSA290200500371, task order no. 4 from the Agency for Healthcare Research and Quality, US Department of Health and Human Services, as part of the Developing Evidence to Inform Decisions about Effectiveness (DEcIDE) program.

measured by the number of years lived with a disabling condition.² In primary care practices, more than 50% of depressed patients go unrecognized, and among the half who do receive treatment, it is adequate in approximately 42%, resulting in only 22% of all patients being adequately treated as evaluated by medication use and frequency of follow-up.¹

Assessing depression severity at the time of diagnosis is important for determining prognosis and treatment. Increased severity is associated with higher health care service utilization, worse outcomes, and less likelihood of remission.^{3,4} Patients with severe depression also may take longer to respond to treatment.⁵ Perhaps most importantly, baseline assessment of severity provides a starting point for evaluating the effectiveness of treatment, making adjustments to treatment, and recognizing patients who have treatment-resistant depression.⁶

Assessing suicidality is an important component of assessing depression severity. Each year 31,000 people in the United States, 60% with histories of major depression, die by suicide; 650,000 are treated emergently after a suicide attempt.⁷ Depression increases the risk for suicide by 10- to 20-fold.⁸ The lifetime rate of suicide attempts among those with major depression is 8%. Several direct, patient-based measures including the 9-item Patient Health Questionnaire (PHQ-9)⁹ and Quick Inventory of Depressive Symptomatology¹⁰ provide scores of severity and functional impairment and have been validated for repeated use to measure change over time. Such measures are variably used in routine primary care practice.¹¹

Comparative effectiveness studies of antidepressants leave several important knowledge gaps, including the need for “real-world” studies that can

assess differences in antidepressant effectiveness on a large scale, with the ability to increase our understanding of baseline severity of illness and effects on rare events such as suicidality.¹² Furthermore, little published evidence exists that describes baseline severity of depression and suicide ideation in real-world, primary care practice settings or provides guidance for researchers and clinicians regarding the best ways to measure these phenomena on a large scale. To address these gaps, we examined depression diagnosis, severity, and treatment patterns in primary care using the Distributed Ambulatory Research in Therapeutics Network (DARTNet).^{13,14} DARTNet is research network that links data from electronic health records (EHRs) from 25 organizations representing more than 1700 clinicians and more than 3 million patients. Using data from EHRs and direct measurement methods, this article reports the baseline severity of illness and suicidality, as well as subsequent response and suicidality (after diagnosis), for depressed patients in participating DARTNet practices.

Methods

Patients and Data Collection

We obtained data from 14 DARTNet practices that agreed to participate in this particular study. Among DARTNet practices, these practices were diverse in terms of both geography (located in 9 states from the East coast to the South to West coasts) and practice size (solo practitioner to 12-clinician practice).

We used 3 different approaches to collect data for this study: existing data within each participating DARTNet organization’s EHRs were supplemented by prescription fulfillment data and PHQ-9 data to create progressively more data-enriched subsets of patients. First, we collected an array of demographic and clinical variables from DARTNet EHRs. The data were then standardized as previously described.¹⁵ Variables collected included patient age, sex, diagnoses, history of present illness, family history, social history, laboratory tests and results, and medications ordered. Overall, we extracted all records for up to 4 years (2006 to 2010, as available per individual practice) for all patients with at least one depression-related diagnosis or at least one total PHQ-9 score. From these eligible patients, a retrospective, open cohort of “new” adolescent (ages 13 to 18 years) and

Conflict of interest: Drs. Valuck, Anderson, and Libby and Mr. Allen have received grant funding from the federal government and foundation and industry sources, including antidepressant manufacturers (Eli Lilly and Company, Forest Laboratories, Lundbeck A/S); Mr. Brandt, Ms. Bryan, and Drs. West and Pace have received grant funding from federal government, foundation, and industry sources.

Disclaimer: Statements in this article should not be construed as endorsement by the Agency for Healthcare Research and Quality or the US Department of Health and Human Services.

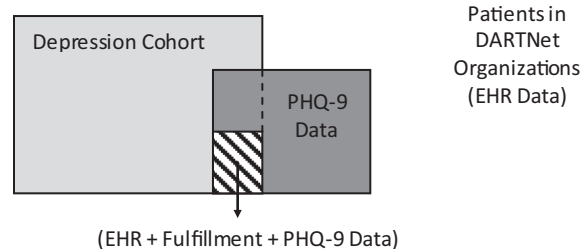
Corresponding author: Robert J. Valuck, PhD, RPh, Department of Clinical Pharmacy, Skaggs School of Pharmacy and Pharmaceutical Sciences, 12850 E. Montview Blvd., Mail Stop C238, Aurora, CO, 80045 (E-mail: robert.valuck@ucdenver.edu).

adult (age ≥ 19 years) depression episodes (referred to as the *depressed cohort*) was created using the following criteria: (1) at least one diagnostic code (*International Classification of Diseases, Ninth Revision, Clinical Modification* [ICD-9-CM]) of 296.2, 296.3, 300.4, or 311; (2) at least 90 days without an antidepressant prescription before the index diagnosis date; and (3) at least 120 days without depression diagnoses. These criteria are based on the Healthplan Employer Data Information System (HEDIS) criteria for defining and measuring new episodes of depression, which are used by the National Committee for Quality Assurance^{16,17} and have been used in prior work.^{18,19}

Second, we obtained prescription fulfillment data (ie, records of paid pharmacy claims) for patients who had such data available (for this study, approximately 55% of patients were in DARTNet practices that had access to fulfillment data). Prescription fulfillment data are important because they indicate that a patient filled a prescription, which is not always the case for medications ordered and noted in the EHR. Prescription fulfillment data also include the quantity of and days supplied for each prescription and thus represent the best available source of information for drug exposure.

Third, we worked with the participating DARTNet practices to create PHQ-9 data collection processes. We electronically prompted practice staff to administer a PHQ-9 at specific intervals for specific patients based on depression diagnoses, other diagnoses used for depression in a given practice, and length of time since last PHQ-9. The PHQ-9 data were collected a number of ways, ranging from printed forms given to the patient in the waiting room or completed by nurses, to online systems used by nurses or clinicians. All practice sites received training on the interpretation and use of PHQ-9 scores to diagnose and monitor depression. Total PHQ-9 scores and item 9 (suicidality) scores were abstracted from the HER, standardized, then added to other clinical data from steps 1 and 2, above. PHQ-9 data collection for this study occurred for a period of 3 to 4 months in most of the participating organizations. Some participating organizations were already collecting PHQ-9 data, so we also collected their existing data. Figure 1 provides a conceptual depiction of the DARTNet population, with the subsets of patients for whom data from various sources was extracted or collected for this study.

Figure 1. Conceptual diagram of the DARTNet population, data sources, and study cohorts. EHR, electronic health record; PHQ-9, 9-item Patient Health Questionnaire.



Measures

For each episode of depression, we identified patients' age at time of index diagnosis, sex, episode length (according to HEDIS criteria), primary diagnosis code used for the episode, severity of illness (among patients whose depression was coded with ICD-9-CM code 296.2 or 296.3, which allows for severity coding), episode start type (new vs recurrent), and episode end type (among patients whose depression was coded with ICD-9-CM code 296.2 or 296.3, which allows for coding of partial or full remission). We reviewed PHQ-9 data from all patients in DARTNet participating practices, patients in the depression cohort, and patients in the depression cohort for whom prescription fulfillment data also were available.

Statistical Analysis

Descriptive statistics (N, percent, mean, median, range) were used to characterize all measures. Spearman's rank-order correlation (2-tailed α level of 0.05) was used to test the strength and direction, if any, of the relationship between baseline severity of illness and baseline suicidality levels according to the specified PHQ-9 score ranges. All data management and analyses were performed using SAS version 9.2 (SAS, Inc., Cary, NC). The study protocol and data management and analysis plan were approved by the Colorado Multiple Institution Review Board (Aurora, CO) and the American Academy of Family Physicians Institutional Review Board (Leawood, KS), on behalf of individual DARTNet practices.

Results

EHR and PHQ-9 data were obtained from 14 participating DARTNet organizations for 117,878 pa-

tients with either a depression-related diagnosis or at least one total PHQ-9 score. A total of 2969 patients had at least one PHQ-9 total score recorded and 2469 had at least one item-9 (suicidal-ity) score recorded. From the 117,878 eligible patients, the criteria described above were used to identify a cohort of 81,028 adolescent and adult depression episodes, representing 61,464 unduplicated patients. From this cohort, 1019 patients had at least one PHQ-9 total score recorded and 908 had at least one item-9 score recorded. Fulfillment data obtained to enhance the EHR data were provided by 3 organizations for a total of 67,391 patients, of which 37,416 matched to a patient in the depression cohort. A match across all 3 data sources (EHR, PHQ-9, and fulfillment) identified 259 patients with data from each source.

Table 1 illustrates demographic and clinical characteristics of the depression episodes identified for study. Using only EHR data, patients ($n = 61,464$) had an average of 1.3 episodes (median, 1.0 episode; range, 1–11 episodes) of depression, a mean age of 57.51 years, and approximately 70% were women. Relatively few (2.04%) of the episodes identified were in adolescents, but older adults (aged ≥ 60 years in this analysis) were well represented (43.41%). Median episode length was 436 days (approximately 14 months) using the National Committee for Quality Assurance/HEDIS criteria for defining episode length. Nearly 80% of episodes were coded using either ICD-9-CM code 311 (65.78%) or 300.4 (13.97%); only 20% were coded as major depressive disorder using ICD-9-CM codes 296.2 or 296.3. Among the 16,412 episodes diagnosed with a major depressive disorder code, 59% had either unspecified or missing severity information. Of the 41% ($n = 6,741$) with severity of illness indicated, 17.88% were coded as mild, 18.54% as moderate, and 4.65% as severe depression. Overall, only 8.3% of all episodes identified using EHR data had severity of illness information available. Finally, most episodes (75%) were index episodes (vs recurrent), meaning the first episode seen in the available data. Indications of partial or full remission (ie, the end of an episode according to ICD-9-CM coding) were used in less than 0.2% of episodes, so the characterization of episode ends and overall length is tenuous at best using this data element. Results were generally similar for patients with both EHR and prescription fulfillment data as well as those with EHR plus fulfillment plus PHQ-9 data

available; exceptions (eg, mean age being higher in the EHR plus fulfillment group) because the practices for which fulfillment data were available have more older adults in their patient populations.

Table 2 describes utilization patterns during antidepressant exposure for episodes for which patients ordered or filled an antidepressant prescription within 30 days of index depression diagnosis. Slightly more than 21% of episodes had EHR-detected orders for antidepressants, but only 2% to 8% of episodes had antidepressant prescriptions filled. This low rate could be because of incomplete or delayed fulfillment data being fed back to practices; we relied on the data available at the time the study. Among those episodes with an antidepressant order or fill, selective serotonin reuptake inhibitors were the most commonly used (73% to 75%), followed by selective noradrenergic reuptake inhibitors (11% to 17%), bupropion (5% to 9%), and various other agents ($< 5\%$ each). Mean time to first prescription fill was approximately 7 days (median, 2.0 days), according to fulfillment data; thus, most antidepressant prescriptions seem to be filled within a week of being ordered, if they are filled at all. According to fulfillment data, typical persistence with therapy was about 90 days, corresponding with guideline-recommended time windows for initial management of acute depressive episodes. Adherence with antidepressant prescriptions was high, with mean and median medication possession ratio values of 1.0, indicating that patients had medication available on essentially all days for which they remained on therapy. Concomitant psychotropic medication use was fairly common, with anxiolytics used in 13% to 29% of episodes and antipsychotics used in 7% to 17% of episodes for which patients received antidepressants. Stimulants and narcotic analgesics were used less frequently ($< 5\%$ of episodes for which patients received an antidepressant).

Table 3 describes PHQ-9 data collected for this study for all patients in participating DARTNet practices over a period of 6 to 9 months. Nearly 5000 PHQ-9 instruments (4900 instruments in 2969 patients) were administered and entered into patients' EHRs, representing the capability of DARTNet to collect direct, patient-reported outcomes on a large scale. An average of 1.2 PHQ-9s were collected per patient (median, 1.0), although the relatively brief data collection period likely limited the number of response (follow-up) measurements that

Table 1. Overview of Demographic and Clinical Characteristics of Depression Episodes

	EHR	EHR + Fulfillment	EHR + Fulfillment + PHQ-9
Episodes, n	81,028	46,131	284
Subjects, n	61,464	37,416	259
Episodes per subject			
Mean	1.3	1.2	1.1
Median	1.0	1.0	1.0
Range	1–11	1–6	1–5
Demographic characteristics of episodes			
Age at start of episode (years)			
Mean	57.71	63.18	52.47
Median	59.00	67.00	53.00
Range	13–89	13–89	16–88
Age group			
Adolescent, 13–18 years	1649 (2.04)	472 (1.02)	7 (2.46)
Adult, 19–64 years	44,208 (54.56)	19,197 (41.61)	192 (67.61)
Older adult, ≥65 years	35,171 (43.41)	26,462 (57.36)	85 (29.93)
Female sex	56,547 (69.79)	30,818 (66.81)	204 (71.83)
Clinical characteristics of episodes			
Length of episode (days)			
Mean	655.53	701.03	759.47
Median	436.00	477.00	626.00
Primary diagnosis of episode*			
296.2x	9305 (11.48)	5437 (11.79)	9 (3.17)
296.3x	7107 (8.77)	5528 (11.98)	7 (2.46)
300.4x	11,319 (13.97)	6922 (15.01)	22 (7.75)
311.x	53,297 (65.78)	28,244 (61.23)	246 (86.62)
Episode severity [†]			
Unspecified	7819 (47.64)	5365 (48.93)	3 (18.75)
Mild	2934 (17.88)	2436 (22.22)	4 (25.00)
Moderate	3043 (18.54)	1394 (12.71)	5 (31.25)
Severe without psychosis	539 (3.28)	378 (3.45)	1 (6.25)
Severe with psychosis	225 (1.37)	200 (1.82)	0 (0.00)
Severity missing	1852 (11.28)	1192 (10.87)	3 (18.75)
Type of episode at start			
Index	60,361 (74.49)	37,222 (80.69)	258 (90.85)
Recurrent	20,667 (25.51)	8909 (19.31)	26 (9.15)
Episode end type			
Full	70 (0.09)	68 (0.15)	0 (0.00)
Partial	72 (0.09)	47 (0.10)	0 (0.00)
Continuation/unknown	80,886 (99.82)	46,016 (99.75)	284 (100.0)

All data provided as n (%) unless otherwise indicated.

*Code from the *International Classification of Diseases, Ninth Revision (ICD-9)*.

[†]Among ICD-9 codes 296.2 or 296.3; severity levels defined as 0–4 (no depression), 5–9 (minimal), 10–14 (mild), 15–19 (moderate), and 20–27 (major).²¹ Scores from item 9 (suicidality) of the 9-item Patient Health Questionnaire (PHQ-9) were characterized by raw scores indicating frequency of suicidal thoughts in the past 2 weeks: 0 (none), 1 (several days), 2 (more than half the days), or 3 (nearly every day).

EHR, electronic health record; PHQ-9, 9-item Patient Health Questionnaire.

could be made for each patient. Overall, the 4900 PHQ-9s collected exhibited a considerable amount of variation in both severity of illness (32.05% with no depression, 26.89% with minimal, 19.54% with mild,

12.04% with moderate, and 9.47% with severe depression) and suicidality (69.43% with a score of 0, 22.58% with a score of 1, 4.97% with a score of 2, and 3.02% with a score of 3 on item 9 of the PHQ-9).

Table 2. Treatment Utilization Patterns during Antidepressant Exposure for Episodes Receiving an Antidepressant within 30 Days of Depression Diagnosis

	EHR	EHR + Fulfillment	EHR + Fulfillment + PHQ-9
Episodes, n	81,028	46,131	284
Episodes receiving an AD within 30 days, n	17,385 (21.46)	996 (2.16)	23 (8.10)
AD Utilization			
AD used to treat episode			
SSRI	12,697 (73.03)	750 (75.30)	17 (73.91)
SNRI	2146 (12.34)	111 (11.14)	4 (17.39)
TCA	393 (2.26)	29 (2.91)	0 (0.00)
Bupropion	1581 (9.09)	52 (5.22)	0 (0.00)
MAOI	5 (0.03)	0 (0.00)	0 (0.00)
PP	316 (1.82)	16 (1.61)	1 (4.35)
Tetracyclics	247 (1.42)	38 (3.82)	1 (4.35)
Days to first AD Rx			
Mean	1.14	6.98	7.61
Median	0.00	2.00	0.00
Persistence*			
Mean	37.87	158.67	150.04
Median	30.00	91.00	90.00
Adherence†			
Mean	1.26	1.00	0.95
Median	1.00	1.00	1.00
Other psychotropic medications prescribed			
Anxiolytics	2243 (12.90)	290 (29.12)	4 (17.39)
Antipsychotics	1806 (10.39)	77 (7.73)	4 (17.39)
Stimulants	456 (2.62)	11 (1.10)	0 (0.00)
Narcotic analgesics	311 (1.79)	40 (4.02)	0 (0.00)

Data provided as n (%) unless otherwise indicated.

*Persistence: days between first and last prescription.

†Adherence: medication possession ratio.

AD, antidepressant; EHR, electronic health record; MAOI, monoamine oxidase inhibitor; PHQ-9, 9-item Patient Health Questionnaire; PP, phenylpiperazine antidepressant; Rx, prescription; SNRI, serotonin-noradrenergic reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

Patients from the depressed cohort with PHQ-9 data available ($n = 1019$) had similar severity levels but slightly higher suicidality levels (27.85% with a score of 1, 5.71% with a score of 2, and 2.73% with a score of 3 on item 9 of the PHQ-9) than the overall population ($n = 2969$) from whom PHQ-9 data were collected. Importantly, regardless of the existence of a depression diagnosis or the timing of administration, 25% to 30% of patients had some degree of suicidal thought at some point in time, according to the PHQ-9.

Table 4 presents baseline severity of illness and suicidality information for depression episodes based on both ICD-9-CM codes and PHQ-9 scores. There were 1440 PHQ-9s collected in the depression cohort, representing 738 episodes and 670 patients with at least one PHQ-9. One hun-

dred thirty-five episodes (18.3%) had a baseline PHQ-9 available, according to our time window definition for a baseline measure. Most baseline PHQ-9s (73.33%) were administered on the day of depression diagnosis; 19.26% were administered before diagnosis (mean, 8.9 days before) and 7.41% during the week after diagnosis. Using ICD-9-CM codes, only 19 episodes (2.5%) had severity of illness information available; 8 were indicated as mild and 11 were indicated as moderate. Using PHQ-9 data, 25 episodes (18.52%) were classified as having no depression at baseline, 39 (28.89%) had minimal depression, 31 (22.96%) had mild depression, 23 (17.04%) had moderate depression, and 17 (12.59%) had severe depression. In terms of baseline suicidality, none of the 738 episodes in the depression

Table 3. Nine-Item Patient Health Questionnaire (PHQ-9) Instruments Collected in DARTNet Practices

	All DARTNet Patients	Depression Cohort	Depression Cohort + Fulfillment Data Available
PHQ-9s collected, n	4900	1892	309
Patients with at least one PHQ-9, n	2969	1019	259
PHQ-9s collected per patient, n			
Mean	1.7	1.9	1.2
Median	1.0	1.0	1.0
Range	1–32	1–32	1–5
Age at time of first PHQ-9 (years)			
Mean	49.1	51.1	53.2
Median	49.0	51.0	54.0
Range	1–89	15–89	16–89
Age group			
Adolescent (13–18 years)	76 (2.56)	22 (2.16)	8 (3.09)
Adult (≥ 19 years)	2768 (93.23)	997 (97.84)	251 (96.91)
Female sex	2005 (67.53)	707 (69.38)	186 (71.81)
PHQ-9 total score information			
Mean	8.9	9.1	9.2
Median	8.0	8.0	8.0
Range	0–27	0–27	0–27
Score ranges			
0–4 (no depression)	1,448 (32.05)	544 (30.49)	67 (29.13)
5–9 (minimal)	1,215 (26.89)	517 (28.98)	63 (27.39)
10–14 (mild)	883 (19.54)	349 (19.56)	47 (20.43)
15–19 (moderate)	544 (12.04)	202 (11.32)	29 (12.61)
20–7 (severe)	428 (9.47)	172 (9.64)	24 (10.43)
PHQ-9 item 9 (suicidality) information			
Score			
Mean	0.4	0.5	0.3
Median	0.0	0.0	0.0
Range	0–3	0–3	0–3
Specific scores			
0 (no thoughts in past 2 weeks)	2,807 (69.43)	1,027 (63.71)	243 (81.00)
1 (thoughts on several days)	913 (22.58)	449 (27.85)	32 (10.67)
2 (thoughts on more than half the days)	201 (4.97)	92 (5.71)	18 (6.00)
3 (thoughts nearly every day)	122 (3.02)	44 (2.73)	7 (2.33)

Data provided as n (%) unless otherwise indicated.
PHQ-9, 9-item Patient Health Questionnaire.

cohort had any form of suicidality recorded by ICD-9-CM coding; however, 28 episodes (3.8%) had some degree of suicidality according to the PHQ-9 item 9 score (≥ 1). Correlation analysis performed on the 2 baseline PHQ-9 measures (severity of illness and suicidality) revealed a strong relationship (Spearman $\rho = 0.54$; $P < .001$) between these 2 dimensions of depression. As baseline depression severity increased, so did the degree of reported suicidality.

Table 5 presents depression response and “emergent” (diagnosis after depression) suicidal-

ity information for depression episodes, based on both ICD-9-CM codes and PHQ-9 scores. Among the 738 episodes (670 patients) with at least one PHQ-9, there were 684 episodes (92.6%) with at least one “response” (follow-up) PHQ-9 score recorded more than 7 days after the depression diagnosis date but before the start of any subsequent episode(s). Most episodes had a single response PHQ-9 score recorded, although the mean (1.8) and range (1–28) of response PHQ-9s collected per episode indicate that DARTNet practices successfully collected multi-

Table 4. Baseline Severity of Illness and Suicidality Measurement Using the 9-Item Patient Health Questionnaire (PHQ-9) and Electronic Health Record Data among Depressed Patients in DARTNet Practices

	Depression Cohort	Depression Cohort + Fulfillment Data Available
PHQ-9s collected, n	1440	192
Episodes with at least one PHQ-9, n	738	169
Patients with at least one PHQ-9, n	670	149
Baseline severity of illness measurement		
Episodes with a baseline PHQ-9,* n	135	22
Days between index depression diagnosis and closest PHQ-9, n		
Mean	-8.9	-5.3
Median	0.0	0.0
Range	-116 to 6	-116 to 5
Baseline PHQ-9s collected		
Before index diagnosis date (T-120 to T-1)	26 (19.26)	2 (9.09)
On index diagnosis date	99 (73.33)	16 (72.73)
After index diagnosis date (T+1 to T+7)	10 (7.41)	4 (18.18)
Baseline severity of illness levels		
ICD-9 severity level		
Level 1 (mild)	8 (10.96)	—
Level 2 (moderate)	11 (15.07)	—
Level 3 (severe without psychosis)	0 (0.00)	—
Level 4 (severe with psychosis)	0 (0.00)	—
PHQ-9 severity level		
0-4 (no depression)	25 (18.52)	0 (0.00)
5-9 (minimal)	39 (28.89)	5 (22.73)
10-14 (mild)	31 (22.96)	8 (36.36)
15-19 (moderate)	23 (17.04)	3 (13.64)
20-27 (severe)	17 (12.59)	6 (27.27)
Baseline suicidality measurement		
ICD-9 suicidality		
Ideation	0 (0.00)	0 (0.00)
Suicide attempt	0 (0.00)	0 (0.00)
None	135 (100.0)	22 (100.0)
PHQ-9 item 9 (suicidality) score		
Mean	0.4	0.6
Median	0.0	0.0
Range	0-3	0-3
Specific scores		
0 (no thoughts in past 2 weeks)	76 (73.08)	14 (63.64)
1 (thoughts on several days)	20 (19.23)	4 (18.18)
2 (thoughts on more than half the days)	6 (5.77)	3 (13.64)
3 (thoughts nearly every day)	2 (1.92)	1 (4.55)

Data provided as n (%) unless otherwise indicated.

*Baseline PHQ-9 was defined as the closest total PHQ-9 score within 120 days before and 7 days after the index depression diagnosis date.

ICD-9, *International Classification of Diseases, Ninth Revision*.

ple response PHQ-9s from many patients both before and during the study period. Most (81.58%) response PHQ-9s were collected >90 days after the index depression diagnosis date; 15.74% were collected 7 to 60 days after diagno-

sis, whereas only 2.78% were collected 61 to 90 days after diagnosis.

In terms of measuring depression response, Table 5 also illustrates the rarity with which ICD-9-CM resolution codes are used (0.2% of episodes),

Table 5. Depression Response and Suicidality Measurement Using 9-Item Patient Health Questionnaire (PHQ-9) and Electronic Health Records (EHR) Data among Depressed Patients in DARTNet Practices

	Depression Cohort	Depression Cohort + Fulfillment Data Available
PHQ-9s collected, n	1,440	192
Episodes with at least one PHQ-9, n	738	169
Patients with at least one PHQ-9, n	670	149
Depression response measurement		
Episodes with a response PHQ-9, n	684	149
Response PHQ-9s per episode, n		
Mean	1.8	1.1
Median	1.0	1.0
Range	1–28	1–4
Days between index depression diagnosis and closest PHQ-9, n		
Mean	796.8	802.5
Median	545	630
Range	8–3327	2–2934
Response PHQ-9s collected after depression diagnosis (days)		
7–60	107 (15.64)	17 (11.41)
61–90	19 (2.78)	1 (0.67)
>90	558 (81.58)	131 (87.92)
Depression response measurement		
ICD-9 severity level		
Level 5 (partial response)	1 (0.10%)	0 (0.00%)
Level 6 (full/complete response)	1 (0.10%)	0 (0.00%)
PHQ-9 severity level (lowest response)		
<10 (partial response)	464 (67.84%)	96 (64.43%)
<5 (full/complete response)	272 (39.77%)	53 (35.57%)
Days between baseline and lowest response PHQ-9		
Mean	71.1	42.0
Median	37.0	42.0
Range	10–378	29–55
Change in PHQ-9 scores (baseline to lowest)		
Mean	5.1	3.0
Median	3.0	3.0
Range	–3 to 19	1–5
Ranges of change		
–5 to –1 point	8 (9.88%)	0 (0.00%)
0–5 points	42 (51.85%)	2 (100.0%)
6–10 points	14 (17.28%)	0 (0.00%)
>10 points	17 (20.99%)	0 (0.00%)
≥50%	45 (55.56%)	0 (0.00%)
Baseline ≥10 and response <10	29 (35.80%)	2 (100.0%)
Emergent suicidality measurement		
ICD-9 suicidality		
Ideation	2 (0.29%)	1 (0.67%)
Suicide attempt	1 (0.15%)	1 (0.67%)
None	681 (99.56%)	147 (98.66%)
PHQ-9 item 9 (suicidality)		
Mean	0.6	0.2
Median	0.0	0.0
Range	0–3	0–3

Continued

Table 5. Continued

	Depression Cohort	Depression Cohort + Fulfillment Data Available
Specific scores		
0 (no thoughts in past 2 weeks)	317 (53.01%)	124 (84.35%)
1 (thoughts on several days)	212 (35.45%)	14 (9.52%)
2 (thoughts on more than half the days)	48 (8.03%)	5 (3.40%)
3 (thoughts nearly every day)	21 (3.51%)	4 (2.72%)
Emergence vs. resolution of suicidality (for episodes with baseline and response PHQ-9 item 9)		
Episodes		
Episodes with baseline suicidality = 0 and emergent suicidality >0	10 (16.67)	0 (0.00)
Episodes with baseline suicidality >0 and emergent suicidality = 0	8 (13.33)	0 (0.00)

Data provided as n (%) unless otherwise indicated.
ICD-9, *International Classification of Diseases, Ninth Revision*.

even among those episodes for which resolution codes are available. Using the lowest total PHQ-9 score after diagnosis as an indicator of response, 68% of episodes achieved total PHQ-9 scores <10, and 40% achieved total scores <5 (2 commonly used thresholds for treatment success).

Using the change between baseline and lowest total PHQ-9 score as an alternative method of measuring response, we applied several specifications for classifying changes in depression severity after diagnosis: In terms of raw total score change, only 9.88% of episodes showed clinical worsening, 51.85% showed an improvement of 0 to 5 points, 17.28% showed an improvement of 6 to 10 points, and 20.99% showed an improvement of more than 10 points; more than half (55.56%) of episodes showed at least a 50% reduction from baseline to lowest PHQ-9 total score, and 35.80% of episodes had a baseline total score ≥ 10 and a lowest total score <10.

In terms of “emergent” suicidality (ie, changes in suicidality between the baseline and response/follow-up periods), Table 5 again illustrates the rarity with which suicide-related behaviors can be detected via ICD9-CM codes: Only 2 instances of suicidal ideation and 1 suicide attempt were detected, a rate of approximately 0.44% among the episodes studied. Using item 9 of the PHQ-9 to measure “emergent” suicidality, 281 instances of suicidal ideation were detected, in addition to 317 cases where “no suicidal thoughts” were affirmatively reported (which differs from failure to detect such thoughts via medical record coding). Finally, we compared baseline and “emergent” suicidality scores from item 9 of the PHQ-9 and observed 10

instances (1.4% of episodes) where suicidality was denied at baseline but emerged after diagnosis and 8 instances (1.1% of episodes) where suicidality was reported at baseline but resolved after diagnosis. These 2 indicators hold promise, as sample sizes continue to increase, for detecting and determining the natural course of suicidality among depressed patients and ultimately how it may be influenced by different forms of treatment.

Discussion

This study of patients cared for in geographically diverse primary care sites across the United States found that depressed patients exhibited substantial variation in baseline severity of illness and suicidality. These findings are consistent with those of other research in both clinical trial populations²⁰ and epidemiologic surveys.²¹ Twenty-five to 30% of patients in our study expressed some degree of suicidal ideation according to the PHQ-9, and suicidality was correlated with severity. Practices were able to implement severity monitoring in a variety of ways and were prompted through their EHR to collect the data at clinically useful intervals. Combining severity assessments with data on prescription fill rates provides an opportunity to identify prescription drug adherence and severity or suicidality changes in relation to treatment changes.

Clinically, our study has several important implications. First, we showed the viability of integrating EHR, prescription fulfillment, and PHQ-9 data to provide clinicians with an enriched medical record, which may enable them to provide higher quality care for depression. Knowledge of baseline

severity of illness and suicidality provides a starting point for evaluating the effectiveness of treatment, the need to adjust treatment, and ultimately recognizing patients who have treatment-resistant depression. Better decisions at these key points in the treatment of depression have been shown to improve patient outcomes.^{4,5} Integration of prescription fulfillment data are important because it enables clinicians to differentiate between therapeutic failure (ie, a drug not working as well as expected) and poor outcomes that may be due to adherence issues (ie, not filling antidepressant prescriptions regularly or at all) and thus incorporate these factors into their patients' treatment plans. Many DARTNet practices continued to use the PHQ-9 and fulfillment data after the study has been completed, despite the implementation challenges that had to be overcome to make the additional data collection a part of routine care (data on file). Importantly, the study demonstrates that measurement-based care for depression is both possible and potentially useful in primary care settings.

This study has a number of limitations. First, DARTNet practices are not fully representative of primary care practices in the United States,¹⁴ and the sites participating in this study are only a subset of the research network, thus limiting the generalizability of the results. Second, these findings do not apply to those clinic patients who are not identified as depressed, a well-described group that may represent as much as 50% of those in primary care settings who are depressed.^{20,22} Third, the PHQ-9 has been validated in primary care settings but is still an imperfect tool for measuring depression severity and response, particularly near the lower end of the score range.⁹ Fourth, data were available from all 3 of our measurement sources (EHR, PHQ-9, and fulfillment data) for only a small number of subjects. This limited our ability to validate and compare the individual measures or determine the extent to which any one measure may be superior to another. Finally, the data presented are hypothesis generating in nature. Although a number of trends seemed apparent in the results, these must be confirmed in larger, subsequent analyses of the DARTNet population.

Conclusions

This study provides further evidence of the relative frequency and severity of depression and suicidality

in primary care settings, and it demonstrates the capability of DARTNet to use existing EHR data and supplement it with both prescription fulfillment data and PHQ-9 data collected at the point of care to enhance clinical care for depression and enable more robust comparative effectiveness research.

The authors thank Elizabeth Staton, MS, for assistance with the preparation and editing of the manuscript for submission for publication.

References

1. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003;289:3095–105.
2. Murray CJL, Lopez AD, eds. *A Comprehensive Assessment of Mortality and Disability from Diseases, Injuries, and Risk Factors in 1990 and Projected to 2020. The Global Burden of Disease and Injury Series, no. 1.* Cambridge, MA: Harvard School of Public Health, on behalf of the World Health Organization and the World Bank, Harvard University Press; 1996.
3. Nease DE Jr, Volk RJ, Cass AR. Does the severity of mood and anxiety symptoms predict health care utilization? *J Fam Pract* 1999;48:769–77.
4. Blom MB, Spinhoven P, Hoffman T, et al. Severity and duration of depression, not personality factors, predict short term outcome in the treatment of major depression. *J Affect Disord* 2007;104:119–26.
5. Thase ME, Simons AD, Cahalane J, McGeary J, Harden T. Severity of depression and response to cognitive behavior therapy. *Am J Psychiatry* 1991; 148:784–9.
6. Kroenke K, Spitzer RL. The PHQ-9: a new depression and diagnostic severity measure. *Psychiatr Ann* 2002;32:509–21.
7. Goldsmith S, Pellmar T, Kleinman A, Bunney W, eds. *Reducing Suicide: A National Imperative.* Washington, DC: Institute of Medicine National Academies Press; 2002.
8. Harris EC, Barraclough B. Suicide as an outcome for mental disorders: a meta-analysis. *Br J Psychiatry* 1997;170:205–28.
9. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med* 2001;16:606–13.
10. Trivedi MH, Rush AJ, Ibrahim HM, et al. The Inventory of Depressive Symptomatology, Clinician Rating (IDS-C) and Self-Report (IDS-SR), and the Quick Inventory of Depressive Symptomatology, Clinician Rating (QIDS-C) and Self-Report (QIDS-SR) in public sector patients with mood disorders: a psychometric evaluation. *Psychol Med* 2004;34:73–82.

11. O'Connor EA, Whitlock EP, Beil TL, Gaynes BN. Screening for depression in adult patients in primary care settings: a systematic evidence review. *Ann Intern Med* 2009;151:793–803.
12. Agency for Healthcare Research and Quality. Comparative Effectiveness of Second- Generation Antidepressants in the Pharmacologic Treatment of Adult Depression. In: *Effective Health Care*. Pub. No. 07-EHC007–1. Rockville, MD: AHRQ; 2007.
13. Pace W, et al. Distributed Ambulatory Research in Therapeutics Network (DARTNet): summary report. July 2009, Prepared by University of Colorado DEcIDE Center under Contract No. HHSA 29020050037I-TO2. Rockville, MD: Agency for Healthcare Research and Quality. Available at: http://effectivehealthcare.ahrq.gov/ehc/products/53/151/2009_0728DEcIDE_DARTNet.pdf. Accessed December 4, 2010.
14. Pace WD, Cifuentes M, Valuck RJ, et al. An electronic practice-based network for observational comparative effectiveness research. *Ann Intern Med* 2009;151:338–40.
15. Agency for Health Care Research and Quality. Distributed Ambulatory Research in Therapeutics Network (DARTNet): Summary Report, Appendix A, pages A4–A8. Available at: http://effectivehealthcare.ahrq.gov/ehc/products/53/152/2009_0728DEcIDE_DARTNet_Appendices.pdf. Accessed December 4, 2010.
16. National Committee for Quality Assurance. HEDIS volume 2: technical specifications. Washington, D.C.: National Committee for Quality Assurance; 2004.
17. Scholle SH. NCQA behavioral health measurement efforts. *J Managed Care Pharm* 2005;11:S9–S11.
18. Libby AM, Brent DA, Morrato EH, Orton HD, Allen R, Valuck RJ. Decline in treatment of pediatric depression after FDA advisory on risk of suicidality with SSRIs. *Am J Psychiatry* 2007;164:884–91.
19. Valuck R, Libby AM, Orton HD, Morrato EH, Allen R, Baldessarini RJ. Spillover effects on treatment of adult depression in primary care after the FDA advisory on risk of pediatric suicidality with SSRIs. *Am J Psychiatry* 2007;164:1198–205.
20. Gaynes BN, Rush AJ, Trivedi MH, et al. Major depression symptoms in primary care and psychiatric care settings: a cross-sectional analysis. *Ann Fam Med* 2007;5:126–34.
21. Kessler RC, Berglund P, Demler O, et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003;289:3095–105.
22. Ormel J, Koeter MW, van den Brink W, et al. Recognition, management, and course of anxiety and depression in general practice. *Arch Gen Psychiatry* 1991;48:700–6.