

© Health Research and Educational Trust
DOI: 10.1111/j.1475-6773.2010.01132.x
RESEARCH ARTICLE

Examining the Relationship between Clinical Monitoring and Suicide Risk among Patients with Depression: Matched Case–Control Study and Instrumental Variable Approaches

Hyungjin Myra Kim, Daniel Eisenberg, Dara Ganoczy, Katherine Hoggatt, Karen L. Austin, Karen Downing, John F. McCarthy, Mark Ilgen, and Marcia Valenstein

Objective. To assess the relationship between closer monitoring of depressed patients during high-risk treatment periods and death from suicide, using two analytic approaches.

Data Source. VA patients receiving depression treatment between 1999 and 2004.

Study Design. First, a case–control design was used, adjusting for age, gender, and high-risk days (1,032 cases and 2,058 controls). Second, an instrumental variable (IV) approach ($N = 714,106$) was used, with IVs of (1) average monitoring rates in the VA facility of most use and (2) monitoring rates of VA facilities weighted inversely by distance from patients' residences.

Principal Findings. The case–control approach indicated a modest increase in suicide risk with each additional visit (odds ratio = 1.02; 95 percent confidence interval = 1.002, 1.04). The “facility used” IV estimate indicated near zero change in risk (0.0008 percent increase; $p = .97$) with each additional visit, while the distance-weighted IV estimate indicated a 0.032 percent decrease in risk ($p = .29$). An alternative analysis assuming a threshold effect of ≥ 4 visits during high-risk periods also showed a decrease (0.15 percent; $p = .08$) using the distance IV.

Conclusions. The IV approach appeared to address the selection bias more appropriately than the case–control analysis. Neither analysis clearly indicated that closer monitoring during high-risk periods was significantly associated with reduced suicide risks, but the distance-weighted IV estimate suggested a potentially protective effect.

Key Words. Suicide, HEDIS visit, depression treatment, case–control, instrumental variable

Suicide is the 11th leading cause of death in the United States (National Center for Injury Prevention and Control, Center for Disease Control [CDC]). To

reduce suicide-related mortality, the Institute of Medicine (2004) has called for improvements in suicide surveillance, monitoring, and prevention. Close clinical monitoring of individuals during high-risk periods is often suggested as one way to reduce suicide deaths, and the U.S. Food and Drug Agency (FDA) has recommended closer monitoring for children, adolescents, and young adults starting antidepressants or changing doses to reduce suicide risks at these times. The most stringent FDA recommendation for monitoring suggested that seven clinical visits be completed in the 12 weeks following these antidepressant treatment events (U.S. Food and Drug Administration). Close clinical monitoring has also been suggested for patients (of any age) who are recently discharged from psychiatric inpatient stays, another high-risk period for suicide (American Psychiatric Association 2003).

However, intensive monitoring following antidepressant changes does not appear to be the norm. Only 23 percent of patients enrolled in a large managed care organization and 6 percent of patients in the Veterans Health Administration (VHA) system who started new antidepressant treatment received seven or more visits in the first 12 weeks following these antidepressant starts (Stettin et al. 2006; Valenstein et al. 2009). Providing more intensive monitoring during these treatment periods would require substantial reorganization of mental health services. Health systems likely will require solid evidence for the effectiveness of close monitoring in reducing suicide before making these substantial service changes.

To date, no published studies have systematically evaluated the impact of intensive clinical monitoring on suicide deaths. Demonstrating such a link is difficult due to the low base rate of completed suicide and the resulting need for extremely large sample sizes (Gunnell, Saperia, and Ashby 2005; Simon

Address correspondence to Marcia Valenstein, M.D., M.S., Department of Psychiatry, University of Michigan Medical School, 4250 Plymouth Road, Ann Arbor, MI 48109; e-mail: marciav@umich.edu. H. Myra Kim, Sc.D., John F. McCarthy, Ph.D., Mark Ilgen, Ph.D., and Marcia Valenstein, M.D., M.S., are with the Department of Veterans Affairs Health Services Research and Development, Ann Arbor Center of Excellence (COE), Ann Arbor, MI. H. Myra Kim, Sc.D., is with the Center for Statistical Consultation and Research, University of Michigan, Ann Arbor, MI. Daniel Eisenberg, Ph.D., is with the Department of Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor, MI. Dara Ganoczy, M.P.H., Karen L. Austin, M.P.H., Karen Downing, B.A., John F. McCarthy, Ph.D., Mark Ilgen, Ph.D., and Marcia Valenstein, M.D., M.S., are with the Serious Mental Illness Treatment Research and Evaluation Center (SMITREC), Ann Arbor, MI. Katherine Hoggatt, Ph.D., is with the Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI. John F. McCarthy, Ph.D., Mark Ilgen, Ph.D., and Marcia Valenstein, M.D., M.S., are with the Department of Psychiatry, University of Michigan Medical School, Ann Arbor, MI.

et al. 2006). Sample sizes required to examine meaningful differences in suicide rates with alternative treatment interventions are estimated at 300,000 or more—numbers that preclude using randomized-controlled trials for this purpose (Gunnell, Saperia, and Ashby 2005; Simon et al. 2006). Administrative data from large health care organizations can provide the necessary sample size, but they raise an important methodological issue of confounding by treatment indication. In clinical settings, treatment receipt is influenced by a mixture of patient, provider, and organizational factors. If more severely ill patients are channeled to a particular treatment (e.g., increased monitoring) and are more likely to have adverse outcomes, this may result in spurious associations between increased monitoring and suicide (Prentice et al. 2005). A bias in the opposite direction may also arise if providers and organizations with more intensive monitoring provide higher quality care in other unmeasured dimensions.

In this study, we used data from a large national cohort of VHA patients in depression treatment to investigate the relationship between more intensive monitoring during high-risk treatment periods and suicide. Recognizing the potential for bias, we used two different analytic approaches to address confounding by treatment indication. We first used a case-control design, adjusting for all known characteristics available in the administrative databases that might be associated with both increased monitoring and suicide. The traditional case-control design allows matching of cases with controls in order to assess the relationship between the primary exposure variable and outcome of interest while controlling for the matching variables as well as the undefined variables associated with the matching variables. We subsequently used an instrumental variable (IV) approach to address potential biases due to variables not available in the administrative databases (Newhouse and McClellan 1998; Fortney et al. 2001; Salkever et al. 2004).

These two approaches each have potential advantages and disadvantages. The case-control approach is well suited for studying uncommon events and is intuitively straightforward. However, it is limited by the patient and facility characteristics available in the administrative or study databases, and it can produce biased results when factors unavailable in these databases influence both the level of monitoring and the likelihood of suicide. In theory, the IV approach can account for unmeasured confounders, but its validity depends on the appropriateness of the chosen IV. This approach typically involves a loss of statistical power, because it is based on variation in treatment due to specific factors (the IVs) rather than the full variation in treatment. The loss in statistical power may particularly limit the application of the IV

approach where the event of interest has a low base rate. We compared these two methods, with the ultimate goal of assessing whether increased monitoring during high-risk treatment periods reduces suicide deaths.

METHODS

Study Sample

Study data were drawn from a cohort of 887,859 patients receiving depression treatment between April 1, 1999 and September 30, 2004 in the VHA health system, but the size and composition of the samples differed slightly for the two approaches. Patients were included in the overall depression treatment cohort if they received both a diagnosis of a depressive disorder and an antidepressant medication fill from VHA providers, or if they had two encounters with diagnoses of depressive disorders. Patients were excluded if they received any diagnosis of bipolar I, schizophrenia, or schizoaffective disorder during the study period. Several prior studies and accrediting agencies have used similar definitions to define cohorts with depressive disorders when examining the quality of care for depression (Kerr et al. 2000; Charbonneau et al. 2003; Spettell et al. 2003; National Committee on Quality Assurance 2006). Because patients who died soon after entering the cohort had limited observation time, making monitoring rates difficult to calculate and important patient characteristics difficult to assess, patients with <84 observation days were excluded from the primary study analyses (although they were included in sensitivity analyses), resulting in a cohort of 835,944 patients.

Study Measures

Suicide. National Death Index (NDI) *Plus* includes national data regarding dates and causes of death for all U.S. residents, derived from death certificates filed in state vital statistics offices. As outlined in a prior report based on the same treatment cohort (Zivin et al. 2007), NDI queries were submitted for cohort patients with a date of death in the VA Beneficiary Identification and Records Locator System Death File during the study period and for patients who did not use VHA services in the year following the study period, resulting in a comprehensive assessment of all suicide deaths in the study population.

High-Risk Periods and Intensity of Clinical Monitoring. High-risk periods for suicide were defined as the 12-week (84 day) periods immediately following psychiatric hospitalizations, new antidepressant (AD) starts, other

antidepressant starts, or antidepressant dose changes (≥ 50 percent change in dose between two consecutive fills of a specified AD occurring within 6 months) (Valenstein et al. 2007). Monitoring visits were defined using Health and Employment Data and Information Set (HEDIS) criteria modified for the VHA setting (National Committee on Quality Assurance 2006). A HEDIS monitoring visit is an outpatient visit that has a psychiatric current procedural terminology (CPT) code or a nonpsychiatric CPT code accompanied by a mental health diagnosis. On any given high-risk day, only one monitoring visit was counted, even if more than one qualifying visit occurred. The intensity of monitoring was calculated as the HEDIS visit rate during the high-risk periods. Inpatient days were excluded from these high-risk periods, as we were interested in the relationship between outpatient monitoring and suicide deaths.

Covariates. Patient covariates included age categories, race, Hispanic ethnicity, sex, diagnoses of substance use disorder, posttraumatic stress disorder (PTSD), major depression, personality disorder, anxiety disorder and bipolar II, service connection (indicating disability from conditions newly occurring or exacerbated by military service), Charlson medical comorbidity index, use of services with Medicare claims, E-code indicating a suicide attempt, VA psychiatric hospitalization, psychiatric inpatient days, number of psychotropic medications, and number of years since cohort entry. The choice of these covariates was based on the prior literature regarding suicide risk factors and patient factors associated with increased monitoring levels, including our teams' prior paper on risk factors for suicide within this VA population (Simon 1992; Zivin et al. 2007). Facility-level variables included geographic region of the facility of most use in the year and an indicator of whether the facility was in an urban area. Psychiatric comorbidities were defined based on data from 6 months before cohort entry through the end of the study period. Covariates other than psychiatric comorbidities were defined based on data from the 12 months before cohort entry.

Study Analyses

Case-Control Design and Analyses

Matching. For each suicide case, we matched one to two control patients based on demographics and risk factors for suicide. The matching variables were chosen so that cases and controls were at risk in the same time period

with a similar duration of depression treatment and potentially similar illness severity. Control patients were matched individually to a case, randomly chosen with replacement from a pool of control patients who were alive on the date when the case died of suicide (*index date*), with similar numbers of total observation days (± 60 days) as the case before this index date, within ± 5 years of the case birth year, of the same sex, and with similar numbers of high-risk days within 12 months before the index date (± 30 high-risk days). When more than two possible controls satisfied the matching criteria, we randomly selected two controls. A total of 1,537 suicide cases were identified; 5 cases without a matching control were excluded. We then limited the case-control analyses to the 3,090 patients (1,032 cases and 2,058 controls) with at least one high-risk day in the year before their *index date*.

Statistical Analyses. Conditional logistic regression models were used to evaluate the relationship between suicide and intensity of monitoring during high-risk days. The primary predictor was the HEDIS visit rate calculated for the high-risk days in the year before the *index date*. Logistic models included the covariates listed above. The analysis was also conducted using a generalized linear model with case as the response variable with logit link, matched sets as the clusters, and HEDIS visit rate as the primary predictor. A Cox proportional hazard model was also used to assess time from cohort entry until suicide or the *index date*, with late entry at 1 year before *index date* for patients having > 1 year of observation. In these analyses, we categorized the HEDIS visit rate into five dummy variables to check for a nonlinear relationship, and we stratified by index year to assess whether the effect of monitoring changed across study years.

Sensitivity Analyses. All analyses were repeated using the subset of cases (and their matching controls) who survived at least a year after their depression diagnosis. We also completed analyses using a newly pulled case-control dataset without the restriction of patients having at least 84 observation days after cohort entry.

IV Design and Analysis

Analytic Dataset. All patients in the depression cohort from fiscal year (FY) 2000 to 2004 with at least 84 observation days and one high-risk day were included ($N = 706,280$). FY 1999 data were excluded because the data were available only for part of the year. The analytic dataset included 1,295,321 yearly patient-level observations, including HEDIS monitoring rates defined for each year. For the primary IV analysis, we excluded yearly

data with <14 high-risk days because HEDIS monitoring rates may not be reliable when patients have small numbers of high-risk days. As the research question pertains to policies regarding minimal levels of monitoring, the primary analytic dataset also excluded yearly data from those with HEDIS visit rates >7 visits per 84 high-risk days, although this was subjected to sensitivity analysis. This gave 1,168,589 observation years from 682,001 unique patients for the primary IV analysis.

IV Model. We used two-stage IV regression model. The first stage was

$$R_{it} = \alpha_{0t} + \alpha_{1t}V_{it} + \alpha_{2t}X_{it} + \varepsilon_{it}$$

where R_{it} is the monitoring level of patient i at year t (endogenous variable), V_{it} is the IV, and ε_{it} is the effect of the unmeasured factors that affect monitoring rate. X_{it} includes variables known to affect suicide risks and potentially associated with increased monitoring (Zivin et al. 2007) (listed earlier under “Covariates”) and dummy variables for each FY. In the second stage, we estimated the yearly completed suicide linear regression model using the predicted monitoring rate from the first-stage model for each patient, \hat{R}_{it} , and the X_{it} .

An IV approach requires identifying IVs that predict levels of clinical monitoring but are not correlated with suicide, conditional on the level of monitoring (Newhouse and McClellan 1998). We constructed two alternative IVs, both based on the idea that practice patterns of VA facilities closest to one’s residence may affect the type of treatment that one receives but may not be related to unmeasured factors that affect health outcomes. The first IV represented the average intensity of depression monitoring during high-risk periods for all patients at the facility used most often by the individual. For each individual, this IV was calculated as the average number of outpatient depression monitoring visits provided to all depressed VA patients (excluding the individual himself) during high-risk periods in the facility where the individual received the majority of his/her outpatient care during each FY.

The second IV was a distance inverse-weighted average of the monitoring rates of all VA facilities based on the facility’s distance from each individual. This IV was calculated for each individual as

$$\sum_{k=1}^n w_k \times M_k / \sum_{k=1}^n w_k,$$

where M_k was the average monitoring level of k th facility, the weights (w_k) were the inverse of the square root of distance to each facility from the centroid of each individual’s zip code, and n was the number of facilities in the

VA. IV approaches based on distance to facilities have been used in the health services research literature starting with McClellan, McNeil, and Newhouse (1994).

We used these two different IVs because we hypothesized that each would offer distinct advantages. We expected that the first IV would be more strongly predictive of the monitoring intensity that a patient receives (the first condition of a valid IV), because this IV is a function of practice patterns only at the facility that the patient attends most. On the other hand, we expected that the second IV would be less likely to be correlated with unmeasured factors that affect health outcomes (the second condition of a valid IV), because it is a distance-weighted practice patterns of surrounding facilities, determined solely by where one lives rather than the facility one attends the most (which may be affected by unmeasured severity level).

Statistical Analyses. To produce unbiased and consistent estimates, the IV should be (1) associated with the patient-level monitoring and (2) not associated with unmeasured patient or facility characteristics that are likely to affect suicide. To assess the first property for each of the two IVs, we checked the first-stage relationship between patient-level monitoring rate, R_{it} and each IV using linear regression, controlling for X_{it} . To assess the second property, we compared the distribution of individual and facility-level characteristics between individuals with lower versus higher values than the median IV value, and across the individuals of quartiles based on IV values. Finding groups to be similar for most measured characteristics would strengthen the case that the IV was not related to *unmeasured* differences. Our distance-weighted IV estimate might also be confounded if people move closer to facilities that practice more intensive (or less intensive) monitoring because of unmeasured severity of their condition. To explore this further, we looked at a subset of people who moved their residence (on a yearly basis) to see if those who move closer to facilities that practice more versus less intensive monitoring are different in their severity. Finally, we used a two-stage regression based on a linear probability model to examine the relationship between suicide and the IV, adjusting for the covariates X_{it} . The model was estimated using a two-stage model with robust standard errors to account for potential heterogeneity of the error terms within multiple years of data from a person. We used the Hausman test to assess whether there was a significant difference between the parameter estimates of the IV and the naïve analysis estimate using monitoring rate as the predictor (Hausman 1978).

Sensitivity Analyses. To confirm that results did not vary by year, we fit the model separately by each FY and assessed whether the direction of the

relationship remained consistent across years. IV estimates were also obtained using two-stage least random-effect estimators to account for multiple years of data from each person. Because excluding data with HEDIS visit rates >7 per 84 high-risk days may bias the sample toward less severe cases for suicide risk, sensitivity analyses was done including data with HEDIS visit rates >7 . Because treatment from non-VA providers is less common for those with the service connection, we also did subgroup analysis of the persons with a service connection to see whether the results are more reliable in this subgroup. Lastly, recognizing that an increase in monitoring may affect suicide only if it reaches a certain intensity, we conducted an analysis where a threshold monitoring effect was assumed at ≥ 4 visits, as the average increase to bring patients to 7 visits during high-risk periods was 3.5 visits.

RESULTS

Table 1 shows the mean HEDIS monitoring rate for the entire depression cohort with ≥ 84 observation days in the first 12 months following cohort entry ($N = 835,944$) and for the subsets of patients who had ≥ 1 or ≥ 14 high-risk days in the first 12 months following entry. Monitoring rates were higher among patients with indications of more severe psychiatric illness, such as having psychiatric hospitalizations or diagnoses of substance use disorder or posttraumatic stress disorder (PTSD). These characteristics are also associated with increased rates of suicide in this population, (Zivin et al. 2007) suggesting that confounding of the relationship between monitoring and completed suicide by treatment indication is plausible and may be an important source of bias.

Case-Control Analysis

For matched cases and controls, the monitoring rates during high-risk periods were similar. Cases had a mean HEDIS visit rate of 2.9 visits ($SD = 5.3$) per 84 high-risk days, while controls had 2.7 visits ($SD = 5.0$). Table 2 shows demographics and other patient characteristics for cases and controls. Despite matching on index date, sex, age, gender, period of observation, and numbers of high-risk days, cases were still more likely than controls to be white, non-Hispanic, to receive larger numbers of psychotropic medications, and less likely to have comorbid PTSD or a service connection. Unadjusted odds ratio (OR) of suicide associated with an increase of one HEDIS visit per 84 high-risk day period was 1.00 ($p = .27$), and covariate adjusted OR was 1.02 ($p = .03$, Table 3).

Table 1: Monitoring Levels by Measures of Patient Severity: Rate of HEDIS Visits on High-Risk Days in the First 12 Months Following Cohort Entry*

Patient Characteristic	Patients with Observation Days ≥ 84 (N = 835,944) N (%)	HEDIS Visit Rate Mean (SD) [†]	Patients with ≥ 1 High-Risk Days (N = 676,425) N (%)	HEDIS Visit Rate Mean (SD) [†]	Patients with ≥ 14 High-Risk Days (N = 667,030) N (%)	HEDIS Visit Rate Mean (SD) [‡]
Age group						
< 45 years	135,258 (16.2)	6.8 (11.2)	113,803 (16.8)	11.6 (17.3)	112,767 (16.9)	11.2 (15.6)
45-64 years	408,234 (48.8)	7.5 (12.5)	335,093 (49.5)	11.4 (17.8)	330,991 (49.6)	11.1 (15.8)
≥ 65 years	292,347 (35.0)	3.9 (6.4)	227,466 (33.6)	6.6 (11.0)	223,210 (33.5)	6.3 (7.9)
Male	767,861 (91.9)	6.1 (10.8)	621,516 (91.9)	9.8 (16.1)	612,702 (91.9)	9.5 (14.0)
Female	68,083 (8.1)	6.2 (9.1)	54,909 (8.1)	10.0 (13.9)	54,328 (8.1)	9.7 (11.8)
White	641,192 (76.7)	5.8 (9.8)	518,524 (76.7)	9.3 (14.8)	511,002 (76.6)	8.9 (12.6)
Black	103,564 (12.4)	8.5 (15.7)	83,753 (12.4)	13.4 (22.4)	82,547 (12.4)	13.0 (20.6)
Other	17,953 (2.2)	7.8 (12.3)	14,540 (2.2)	11.8 (17.2)	14,352 (2.2)	11.6 (15.7)
Unknown	73,235 (8.8)	5.4 (8.1)	59,608 (8.8)	9.1 (12.8)	59,129 (8.9)	8.9 (11.0)
Hispanic	39,209 (4.7)	7.1 (11.6)	31,667 (4.7)	11.2 (16.9)	31,318 (4.7)	10.9 (15.1)
Not Hispanic	796,735 (95.3)	6.1 (10.6)	644,758 (95.3)	9.7 (15.8)	635,712 (95.3)	9.4 (13.8)
Region						
Northeast	166,634 (19.9)	6.7 (12.6)	129,675 (19.2)	10.3 (17.1)	127,570 (19.1)	10.0 (15.5)
Central	185,546 (22.2)	6.1 (10.6)	150,332 (22.2)	9.9 (16.9)	148,152 (22.2)	9.6 (14.5)
South	314,081 (37.6)	5.3 (9.0)	260,239 (38.5)	8.8 (14.3)	256,987 (38.5)	8.4 (12.0)
West	169,499 (20.3)	7.1 (11.3)	136,053 (20.1)	11.2 (16.3)	134,197 (20.1)	10.9 (14.6)
Charlson index						
0	523,251 (62.6)	6.4 (10.7)	429,356 (63.5)	10.3 (16.2)	424,543 (63.7)	10.0 (14.2)
1	197,301 (23.6)	5.8 (10.7)	157,444 (23.3)	9.2 (15.6)	154,924 (23.2)	8.8 (13.5)
> 1	115,392 (13.8)	5.4 (10.1)	89,625 (13.3)	8.5 (15.0)	87,563 (13.1)	8.2 (12.8)
Psychiatric stay						
Yes	30,983 (3.7)	12.0 (19.2)	28,885 (4.3)	18.5 (27.8)	28,298 (4.2)	18.0 (25.8)
No	804,961 (96.3)	5.9 (10.1)	647,540 (95.7)	9.4 (15.0)	638,732 (95.8)	9.1 (12.9)
Substance abuse						
Yes	188,588 (22.6)	10.2 (17.4)	160,280 (23.7)	15.6 (24.2)	158,176 (23.7)	15.2 (22.4)
No	647,356 (77.4)	4.9 (7.2)	516,145 (76.3)	8.0 (11.7)	508,854 (76.3)	7.7 (9.0)

PTSD									
Yes	191,586 (22.9)	11.4 (15.4)	160,309 (23.7)	15.8 (21.1)	158,323 (23.7)	15.4 (18.9)			
No	644,358 (77.1)	4.5 (8.1)	516,116 (76.3)	7.9 (13.4)	508,707 (76.3)	7.7 (11.2)			
Personality disorder									
Yes	44,076 (5.3)	12.5 (18.2)	37,805 (5.6)	17.9 (24.4)	37,385 (5.6)	17.5 (22.5)			
No	791,868 (94.7)	5.8 (9.9)	638,620 (94.4)	9.3 (15.1)	629,645 (94.4)	9.0 (13.0)			
Bipolar disorder									
Yes	10,021 (1.2)	11.4 (15.7)	8,531 (1.3)	16.7 (21.9)	8,416 (1.3)	16.3 (20.0)			
No	825,923 (98.8)	6.1 (10.6)	667,894 (98.7)	9.7 (15.8)	658,614 (98.7)	9.4 (13.7)			
Other anxiety disorder									
Yes	234,677 (28.1)	7.9 (12.2)	193,061 (28.5)	12.0 (17.6)	190,630 (28.6)	11.7 (15.3)			
No	601,267 (71.9)	5.4 (9.9)	483,364 (71.5)	8.9 (15.1)	476,400 (71.4)	8.6 (13.1)			
Number of psychiatric meds									
0	388,655 (46.5)	5.5 (9.9)	326,514 (48.3)	8.8 (13.9)	325,286 (48.8)	8.6 (12.8)			
1	226,143 (27.1)	5.7 (10.2)	175,141 (25.9)	9.8 (16.4)	171,280 (25.7)	9.4 (13.6)			
2	127,623 (15.3)	6.8 (11.2)	99,620 (14.7)	11.0 (17.7)	97,100 (14.6)	10.5 (14.8)			
> 2	93,523 (11.2)	8.8 (13.2)	75,150 (11.1)	12.9 (19.5)	73,364 (11.0)	12.4 (16.6)			
Service connected									
Yes	264,432 (31.6)	7.1 (10.8)	208,521 (30.8)	10.7 (16.0)	205,407 (30.8)	10.4 (13.6)			
No	571,512 (68.4)	5.7 (10.6)	467,904 (69.2)	9.4 (15.8)	461,623 (69.2)	9.1 (13.9)			

*Excludes patients with <84 observation days.

[†]Wilcoxon rank-sum test or Kruskal-Wallis test—all significant at $p < .0001$.

HEDIS, Health and Employment Data and Information Set; PTSD, posttraumatic stress disorder.

Table 2: Case–Control Analysis—Demographic Characteristics and Other Variables by Suicide for Patients and Controls Who Had At Least One High-Risk Day in Prior Year

	<i>All Patients</i> (<i>N</i> = 3,090) <i>N</i> (%)	<i>Cases</i> (<i>N</i> = 1,032) <i>N</i> (%)	<i>Controls</i> (<i>N</i> = 2,058) <i>N</i> (%)	<i>p</i> -Value*
Male	2,983 (96.5)	996 (96.5)	1,987 (96.6)	.9561
Age (Mean [SD])	56.0 (13.4)	55.9 (13.6)	56.0 (13.2)	.8657
Race				
White	2,448 (79.2)	860 (83.3)	1,588 (77.2)	<.0001
Black	368 (11.9)	42 (4.1)	326 (15.8)	
Other	49 (1.6)	19 (1.8)	30 (1.5)	
Unknown	225 (7.3)	111 (10.8)	114 (5.5)	
Hispanic	142 (4.6)	30 (2.9)	112 (5.4)	.0018
Substance abuse	1,214 (39.3)	413 (40.0)	801 (38.9)	.5556
Bipolar II diagnosis	76 (2.5)	29 (2.8)	47 (2.3)	.3739
Other anxiety diagnosis	1,114 (36.1)	381 (36.9)	733 (35.6)	.4773
Personality disorder	355 (11.5)	120 (11.6)	235 (11.4)	.8635
PTSD	951 (30.8)	262 (25.4)	689 (33.5)	<.0001
Service connected	1,042 (33.7)	301 (29.2)	741 (36.0)	.0002
Charlson score				
0	1,633 (52.9)	572 (55.4)	1,061 (51.6)	.1342
1	648 (21.0)	205 (19.9)	443 (21.5)	
2	261 (8.5)	83 (8.0)	178 (8.7)	
3	297 (9.6)	97 (9.4)	200 (9.7)	
>3	251 (8.1)	75 (7.3)	176 (8.6)	
Had a psychiatric stay	702 (22.7)	232 (22.5)	470 (22.8)	.8232
Number of comorbid psychiatric diagnoses				
0	930 (30.1)	312 (30.2)	618 (30.0)	.2184
1	1,062 (34.4)	370 (35.9)	692 (33.6)	
>1	1,098 (35.5)	350 (33.9)	748 (36.4)	
Number of psychiatric medications				
0	291 (9.4)	85 (8.2)	206 (10.0)	.0003
1	1,165 (37.7)	362 (35.1)	803 (39.0)	
2	755 (24.4)	260 (25.2)	495 (24.1)	
3	447 (14.5)	153 (14.8)	294 (14.3)	
>3	432 (14.0)	172 (16.7)	260 (12.6)	
Medicare use	831 (26.9)	291 (28.2)	540 (26.2)	.2469
Suicide attempt	42 (1.4)	16 (1.6)	26 (1.3)	.5165
Major depression	1,017 (32.9)	352 (34.1)	665 (32.3)	.3165
Region northeast	544 (17.7)	175 (17.1)	369 (18.0)	.8970
Central	680 (22.1)	228 (22.2)	452 (22.0)	
South	1,190 (38.6)	395 (38.5)	795 (38.7)	
West	666 (21.6)	228 (22.2)	438 (21.3)	
Urban	2,775 (90.1)	905 (88.2)	1,870 (91.0)	.0133

*Comparing cases to controls.

PTSD, posttraumatic stress disorder.

Table 3: Case-Control Analysis—From Conditional Logistic Regression Model for the Relationship between Suicide and HEDIS Visit Rate

<i>Variable</i>	<i>Adjusted* Odds Ratios</i>	<i>95% CI</i>	<i>p-Value</i>
HEDIS visit rate	1.02	(1.002, 1.04)	.0253
Black [†]	0.24	(0.17, 0.34)	< .0001
Other race [†]	1.23	(0.67, 2.25)	.5101
Unknown race [†]	1.91	(1.40, 2.59)	< .0001
Hispanic	0.54	(0.35, 0.82)	.0041
PTSD	0.58	(0.44, 0.78)	.0003
Number of psychiatric medications	1.18	(1.11, 1.25)	< .0001
Service connection	0.79	(0.65, 0.95)	.0132
Number of psychiatric stays	0.86	(0.73, 0.998)	.0475
Number of days since depression diagnosis	0.996	(0.993, 0.999)	.0200

*In addition to the listed variables, the model is adjusted also for Charlson scores (0, 1, 2, 3, or > 3), any substance abuse, bipolar II diagnosis, personality disorder, anxiety disorder, major depression, Medicare use, number of psychiatric disorder diagnoses (0, 1, or > 1), number of psychiatric stay days, suicide attempts (E-code), and region and urban status of facility of most use in the year before study entry. None of these variables had a *p*-value < .05.

[†]Reference group is white race.

HEDIS, Health and Employment Data and Information Set; PTSD, posttraumatic stress disorder.

In sensitivity analyses, the positive relationship between monitoring and suicide persisted (OR = 1.03, *p* = .05) even when the analysis was limited to 1,547 patients (525 cases and 1,022 controls) who lived more than a year following their cohort entry. Similar results were obtained when the HEDIS visit rate was entered as five dummy categorical variables to check for non-linear relationships. The GENMOD model with logit link also resulted in an adjusted OR of 1.02 (*p* = .01) when all cases were included and an adjusted OR of 1.03 (*p* = .03) when the analysis was limited to those who lived more than a year following cohort entry. A Cox proportional hazard model with late entry did not give different results. The OR was also very similar when cases (and their matching controls) with HEDIS visit rate > 7 were excluded (OR = 1.07, confidence interval [CI] = 1.01, 1.13) and when using newly pulled case-control dataset that did not exclude those with < 84 observation days (OR = 1.02, CI = 1.01, 1.03).

IV Analysis

Table 4 shows patient-level monitoring is significantly higher for patients in high-intensity monitoring facilities (higher than median) than in low-intensity monitoring facilities, supporting the first criterion for a valid IV. Similarly, when patients were grouped into smaller increments of facility-level

Table 4: Instrumental Variable Analysis—Patient Characteristics by High Versus Low Facility-Level Monitoring (IV; Dichotomized at the Median Rate of 2.22 Visits per 84 High-Risk Days) and by High Versus Low Distance Inverse-Weighted Facility-Level Monitoring (Distance IV; Dichotomized at the Median of 2.31) Based on Patient Years Data from Fiscal Years 2000 to 2004 with at Least 14 High-Risk Days in Each Year

Facility Monitoring	IV		Distance IV	
	Low	High	Low	High
Number of observation	632,728	599,548	637,268	595,008
Yearly suicide probability*	53.0	56.1	53.7	56.3
Endogenous				
HEDIS visit rate [†]	1.74 ^{§§}	2.56	1.85 ^{§§}	2.45
Age	57.4 ^{§§}	56.8	57.0 ^{§§}	57.2
Male (%)	91.5 ^{§§}	91.3	91.2 ^{§§}	91.7
White (%)	79.6 ^{§§}	75.7	77.2 ^{§§}	78.2
PTSD (%)	25.9 ^{§§}	31.2	26.8 ^{§§}	30.3
Bipolar II (%)	1.2 ^{§§}	1.8	1.2 ^{§§}	1.9
Personality disorder (%)	5.0 ^{§§}	6.2	4.8 ^{§§}	6.3
Major depression (%)	26.3 ^{§§}	32.8	28.6 ^{§§}	30.4
Anxiety disorder (%)	32.5	32.5	31.3 ^{§§}	33.8
Substance use (%)	23.9 ^{§§}	30.6	24.7 ^{§§}	29.8
Suicide attempt (%)	0.65 [§]	0.60	0.72 ^{§§}	0.52
Psychiatric days [‡]	2.51 ^{§§}	4.15	2.67 ^{§§}	3.99
Psychiatric stays [‡]	0.13 ^{§§}	0.17	0.13 ^{§§}	0.17
Medications [¶]	1.86 ^{§§}	1.80	1.90 ^{§§}	1.76
Medicare use (%)	31.5 ^{§§}	26.3	29.9 ^{§§}	28.0
Charlson score ≥ 1 (%)	52.8 ^{§§}	50.9	53.7 ^{§§}	50.0
Service connected (%)	33.2 ^{§§}	33.7	33.1 ^{§§}	33.8
Region (%)				
Northeast	15.5 ^{§§}	22.7	11.0 ^{§§}	27.6
Central	19.1	24.9	23.3	20.5
South	56.1	20.5	54.4	22.0
West	9.4	31.9	11.4	29.9
Facility in urban area (%)	86.1 ^{§§}	94.0	89.1 ^{§§}	90.9
Facility bed size	644 ^{§§}	533	641 ^{§§}	535
Fiscal year (%)				
2000	6.0 ^{§§}	15.8	0.07 ^{§§}	21.6
2001	12.8	18.5	0.39	30.8
2002	22.6	17.9	20.0	20.4
2003	29.5	18.7	42.2	6.0
2004	28.9	22.3	37.3	13.9

Note. Values are means unless specified as %.

*Per 100,000 persons.

[†]HEDIS monitoring visit rate expressed as per 84 high-risk days.

[‡]Psychiatric days are inpatient psychiatric hospitalization days in the previous year, and psychiatric stays are the number of inpatient psychiatric hospitalizations in the previous year.

[§]*p*-value is < .05.

[¶]Number of psychotropic medications prescribed in the previous year.

^{§§}*p*-value is < .001 for comparison between patients with low versus high facility-level monitoring. HEDIS, Health and Employment Data and Information Set; PTSD, posttraumatic stress disorder.

monitoring, the mean patient-level HEDIS monitoring rates increased from 1.3 in patients with the lowest facility-level monitoring (between 1.1 and 1.3) to 2.4 in those with the highest facility monitoring level (between 5.7 and 5.9). In addition, the Kleibergen–Paap test of underidentification (Kleibergen and Paap 2006) testing for the correlation between the IVs and the patient-level monitoring was significant ($p < .001$) for both our IVs.

The second property for a valid IV cannot be tested directly, but examining how measured covariates correlate with the IV can give a sense of whether unmeasured differences are likely to be associated with suicide and thus consequential. Patient and facility characteristics in Table 4 for low-versus high-monitoring facilities based on each IV show similarity of several patient-level characteristics between these categories, particularly the nearly identical distributions of gender, age, anxiety, and comorbidity status. However, substantial differences between low- versus high-monitoring facilities are seen in the numbers of psychiatric inpatient days, psychiatric and substance use diagnoses, and geographic region, which may reflect idiosyncratic differences in diagnosis and practice patterns across geographic areas. We note that although these differences are substantial in some cases, they are much smaller than the differences between high versus low *patient-level* monitoring groups (not shown), suggesting that the IV estimate may be less prone to treatment indication bias than a standard single equation approach. We also did not find any patterns in severity such as more inpatient psychiatric days or higher percent of psychiatric illnesses in patients who moved closer to facilities practicing more intensive monitoring than those who moved closer to facilities practicing less intensive monitoring. Lastly, as an indirect test of the second property, we tested for the relationship between suicide and the IVs, conditional on the patient level of monitoring and other covariates and found that it was not significant ($p = .91$ for the IV and $.24$ for distance-weighted IV).

The first-stage model (Appendix SA2), with the endogenous patient-level monitoring as the dependent variable, indicated that 13.0 percent of variation in monitoring rate is explained by the average facility-level monitoring and other exogenous variables. Similarly, 12.5 percent of the variation in patient-level monitoring is explained by the distance inverse-weighted average facility-level monitoring and other exogenous variables. In particular, each additional facility-level average visit and distance inverse-weighted facility-level visit was associated with an unadjusted estimate of 0.28 ($p < .001$) and 0.70 ($p < .001$) more patient visits during high-risk periods, respectively. After adjusting for other patient and facility characteristics variables, the estimates were 0.20 (partial F -test, $p < .001$) for facility-level average monitoring

Table 5: Instrumental Variable (IV) Analyses—The Estimates Are of HEDIS Monitoring Rate from IV Analysis with Facility-Level Monitoring as an IV and with Distance Inverse-Weighted Facility Monitoring as an IV, Based on Linear Probability Model of Suicide

	<i>IV Estimate</i>		<i>Distance IV Estimate</i>	
	<i>Estimate (SE)</i>	<i>p-Value</i>	<i>Estimate (SE)</i>	<i>p-Value</i>
Primary analysis	0.000008 (0.000192)	.970	− 0.000323 (0.000307)	.293
Sensitivity analyses				
Using random effects model	0.000075 (0.000217)	.731	− 0.000165 (0.000526)	.753
Data includes HEDIS rate > 7	− 0.000014 (0.000056)	.797	− 0.000120 (0.000068)	.080
Service connected patients only	− 0.000392 (0.000308)	.203	− 0.000688 (0.000401)	.087

Note. All models were adjusted for age categories, male, race, Charlson comorbidity score (0, 1, 2, ≥3), psychiatric stays in prior year, psychiatric hospitalization days in prior year, number of psychotropic medication in prior year, posttraumatic stress disorder, service connection, substance abuse, bipolar II, personality disorder, major depression, anxiety, suicide attempt in prior 3 years, urban facility, geographical region of facility of most use in each year, and number of years since depression diagnosis.

Primary analysis used patient-year data in the national cohort of VA patients in depression treatment from fiscal years 2000–2004, included data when number of high-risk days was at least 14 in the year, and excluded data with HEDIS monitoring rate > 7.

HEDIS, Health and Employment Data and Information Set.

and 0.99 (partial *F*-test, $p < .001$) for distance inverse-weighted facility-level monitoring. Both IVs were strong predictors of patient-level monitoring; however, the IV based on average monitoring at the facility of most use had a standard deviation 3.5 times larger than that of the IV based on distance-weighted monitoring practices, indicating that it provides more variation and therefore more statistical power (though at the possible expense of higher bias, as noted earlier). Other patient characteristics, such as psychiatric hospitalization and comorbid psychiatric conditions, showed the expected positive relationships with monitoring.

An unadjusted “naïve” linear probability model estimate showed that suicide increased by 0.003 percent ($p = .03$) with each additional monitoring visit. After adjusting for the full set of patient and facility covariates, suicide still increased by 0.003 percent ($p = .05$) with each additional monitoring visit. The IV estimate (Table 5) using facility-level monitoring indicated a positive and insignificant association (0.0008 percent, $p = .97$) in suicide with each

additional monitoring visit. On the other hand, the distance-weighted IV estimate indicated a decrease in suicide of 0.032 percent with each additional visit, although still not statistically significant ($p = .29$). Hausman tests failed to reject that there is no endogeneity, indicating that the difference in coefficients between the naïve and IV analysis was not significant using either IV.

When a random-effects model was used instead of robust standard errors, we found the estimates to differ in size, but there was no change in these basic patterns. IV estimates from models fit separately by each year varied by year with both IVs, but with no consistent direction of association—that is, the estimates neither increased nor decreased in years. IV estimates based on data that included patients with HEDIS rates > 7 and estimates based on data that included only patients with a service connection both estimated a decrease in suicide risk with increased monitoring using either IV. Thus, all estimates, although not statistically significant, suggested a reduction in suicide (Table 5). Lastly, a model with a threshold effect of ≥ 4 visits during high-risk periods showed a decrease of 0.011 percent ($p = .88$) in suicide risks using facility monitoring IV and a decrease of 0.151 percent ($p = .08$) using the distance-based IV.

CONCLUSION

We used data from a large national cohort of VHA patients receiving depression treatment to complete the first comprehensive assessment of monitoring visits and suicide deaths. We show that in this observational dataset, patient-level characteristics associated with higher risks of suicide (Spettell et al. 2003) were also associated with higher levels of monitoring, suggesting there are likely substantial treatment selection biases when assessing the relationship between suicide and clinical monitoring.

We used two analytic approaches to examine the link between monitoring and suicide while addressing these treatment selection biases, the traditional case-control approach and an IV approach—and we found that these analytic approaches resulted in different conclusions.

Although widely used, traditional case-control analyses with the covariate data available in large administrative databases appeared unsuccessful in addressing treatment selection biases. Case-control analyses indicated that visits with clinicians were significantly associated with a slight *increase* in suicide deaths among patients during high-risk periods. This would seem an unlikely scenario unless one believes that clinical contacts increase patient

distress or impulsivity. Instead, this finding is likely due to residual treatment indication biases because of insufficient patient information in administrative data to adjust for factors associated with both monitoring and suicide.

Although some IV analyses based on facility of most use suggested no impact of monitoring on suicide risk, several IV analyses, particularly those using the distance-weighted IV, produced results more in line with expert opinion, suggesting that increased monitoring may lead to reduced suicide risks. However, none of the IV results were statistically significant.

IV estimates may be less subject to treatment indication biases than either the naïve single equation estimate or the case-control study estimates, and they may more closely address policy questions of whether to increase or decrease monitoring for defined populations during high-risk periods. IV analyses assess the *marginal* impact of increases in monitoring visits—the impact of increased monitoring for individuals who receive this closer monitoring only because of differences in facility practices (Newhouse and McClellan 1998). Some patients, because of high-risk behaviors, will receive high levels of monitoring regardless of usual facility practices, and IV estimates do not reflect the benefit these individuals receive from close monitoring. However, because IV approaches focus only on treatment variation that can be explained by the IV, these analyses tend to produce larger standard errors than more direct methods. Likely due to this imprecision and low base rate of completed suicide, IV analyses did not provide a definitive answer regarding the relationship between closer monitoring and suicide despite the use of a large administrative database (Sturm 1998).

We note that while insignificant findings in IV analyses might be due to residual bias or imprecision, it is also possible that clinical visits as currently practiced are not effective in reducing suicide (though they may be effective in addressing other concerns). Routine clinical visits may fail to include a careful assessment of suicidal ideation, plans, or access to lethal means. Alternatively, suicidal ideation or other clinical indicators of risk may not be present during routinely scheduled visits. Suicide attempts are often impulsive and are frequently planned for <30 minutes before being enacted (Simon et al. 2001). Therefore, even large increases in monitoring (e.g., more than four contacts) over the course of 84 days may not be sufficient to detect these short periods of acute risk. Finally to date, even if acute suicide risk is detected during a visit, few clinical interventions have been shown to be effective in reducing these risks (Mann et al. 2005). These possibilities, in combination with our lack of a robust finding for a protective effect, highlight a need to further refine and evaluate additional clinical approaches for reducing suicide risks.

We have reported previously that increasing monitoring from current levels to FDA suggested levels would mean a substantial reorganization of health services along with substantial incremental costs (Valenstein et al. 2009). As noted previously, RCTs of sufficient size to demonstrate a clear link between monitoring and suicide deaths are impractical, and we now show that using observational data (in which clinicians are following individuals they deem at greater risk more closely) to demonstrate *a clear link* between higher levels of monitoring and reduced suicide mortality is also difficult—even when using case-control and IV analyses to address treatment biases. Given the difficulty in demonstrating this link, we believe health care organizations and clinicians will remain unwilling to change current behaviors and press forward with implementing broad policy recommendations regarding monitoring—unless a public consensus develops that these activities should proceed without firm evidence. Indeed, because of the lack of evidence for effectiveness and the large investment that would be required to implement a blanket policy of close monitoring during high-risk periods for all patients, an argument could be made that treatment resources might be better used for mental health interventions with stronger evidence for effectiveness. Future research using large datasets with more detailed information on potential confounders and the development of new methodologies to address treatment selection biases in observational data are clearly needed.

Study Limitations

Diagnoses, demographics, and cause of death may not be completely accurate in administrative databases. However, VHA administrative data quality is generally considered good, with high levels of concordance between VHA administrative data and medical record data (Kashner 1998; Cowper et al. 1999). The NDI is also considered the “gold standard” of U.S. mortality databases (Cowper et al. 2002). Findings for VHA patients also may not generalize to non-VHA patient populations, and findings within the VHA may change as greater numbers of younger veterans enter the health care system following their return from Iraq or Afghanistan.

In summary, although expert and governmental recommendations have urged closer monitoring for depressed patients during high-risk periods to prevent suicide, strong evidence for this recommendation may be difficult to generate. When using observational datasets to address this issue, IV analyses appear less biased than case-control approaches; however, even with a very large database, the application of the IV method appears limited due to low

rate of suicide and the resulting increased variability of its estimate, making it difficult to arrive at definite answers regarding this relationship.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: This research was supported by grants from the Department of Veterans Affairs, Health Services Research and Development Service, IIR 04-211-1 and MRP 03-320, and by the National Institute of Mental Health, R01-MH078698-01.

Disclosures: None.

Disclaimers: None.

REFERENCES

- American Psychiatric Association. 2003. "Practice Guideline for the Assessment and Treatment of Patients with Suicidal Behaviors." *American Journal of Psychiatry* 160 (11, suppl): 1-60.
- Charbonneau, A., A. K. Rosen, A. S. Ash, R. R. Owen, B. Kader, A. Spiro, C. Hankin, L. R. Herz, M. Jo, V. Pugh, L. Kazis, D. R. Miller, and D. R. Berlowitz. 2003. "Measuring the Quality of Depression Care in a Large Integrated Health System." *Medical Care* 41 (5): 669-80.
- Cowper, D. C., D. M. Hynes, J. D. Kubal, and P. A. Murphy. 1999. "Using Administrative Databases for Outcomes Research: Select Examples from VA Health Services Research and Development." *Journal of Medical System* 23 (3): 249-59.
- Cowper, D. C., J. D. Kubal, C. Maynard, and D. M. Hynes. 2002. "A Primer and Comparative Review of Major US Mortality Databases." *Annals of Epidemiology* 12 (7): 462-8.
- Fortney, J., K. Rost, M. Zhang, and J. Pyne. 2001. "The Relationship between Quality and Outcomes in Routine Depression Care." *Psychiatric Services* 52 (1): 56-62.
- Gunnell, D., J. Saperia, and D. Ashby. 2005. "Selective Serotonin Reuptake Inhibitors (SSRIs) and Suicide in Adults: Meta-Analysis of Drug Company Data from Placebo Controlled, Randomised Controlled Trials Submitted to the MHRA's Safety Review." *British Medical Journal* 330 (7488): 385-8.
- Hausman, J. A. 1978. "Specification Tests in Econometrics." *Econometrica* 46 (6): 1251-71.
- Institute of Medicine. ed. 2004. *Reducing Suicide: A National Imperative*. Washington, DC: National Academies Press.
- Kashner, T. M. 1998. "Agreement between Administrative Files and Written Medical Records: A Case of the Department of Veterans Affairs." *Medical Care* 36 (9): 1324-36.

- 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 of Quality Improvement* 26 (4): 203–16.
- Kleibergen, F., and R. Paap. 2006. "Generalized Reduced Rank Tests Using the Singular Value Decomposition." *Journal of Econometrics* 133: 97–126.
- Mann, J. J., A. Apter, J. Bertolote, A. Beautrais, D. Currier, A. Haas, U. Hegerl, J. Lonnqvist, K. Malone, A. Marusic, L. Mehlum, G. Patton, M. Phillips, W. Rutz, Z. Rihmer, A. Schmidtke, D. Shaffer, M. Silverman, Y. Takahashi, A. Varnik, D. Wasserman, P. Yip, and H. Hendin. 2005. "Suicide Prevention Strategies: A Systematic Review." *Journal of American Medical Association* 294 (16): 2064–74.
- McClellan, M., B. J. McNeil, and J. P. Newhouse. 1994. "Does More Intensive Treatment of Acute Myocardial Infarction in the Elderly Reduce Mortality? Analysis Using Instrumental Variables." *Journal of American Medical Association* 272 (11): 859–66.
- National Center for Injury Prevention and Control, Center for Disease Control (CDC). 2006. "Web-Based Injury Statistics Query and Reporting System" [accessed on October 18, 2006]. Available at <http://www.cdc.gov/ncipc/wisqars/default.htm>
- National Committee on Quality Assurance. 2006. "State of Health Care Quality Report, Antidepressant Medication Management" [accessed on October 17, 2006]. Available at http://www.ncqa.org/sohc2003/antidepressant_medication_management.htm
- Newhouse, J. P., and M. McClellan. 1998. "Econometrics in Outcomes Research: The Use of Instrumental Variables." *Annual Review of Public Health* 19: 17–34.
- Prentice, R. L., R. Langer, M. L. Stefanick, B. V. Howard, M. Pettinger, G. Anderson, D. Barad, J. D. Curb, J. Kotchen, L. Kuller, M. Limacher, and J. Wactawski-Wende for the Women's Health Initiative Investigators. 2005. "Combined Postmenopausal Hormone Therapy and Cardiovascular Disease: Toward Resolving the Discrepancy between Observational Studies and the Women's Health Initiative Clinical Trial." *American Journal of Epidemiology* 162 (5): 404–14.
- Salkever, D. S., E. P. Slade, M. Karakus, L. Palmer, and P. A. Russo. 2004. "Estimation of Antipsychotic Effects on Hospitalization Risk in a Naturalistic Study with Selection on Unobservables." *Journal of Nervous and Mental Disease* 192 (2): 119–28.
- Simon, G. E. 1992. "Psychiatric Disorder and Functional Somatic Symptoms as Predictors of Health Care Use." *Psychiatric Medicine* 10 (3): 49–59.
- Simon, G. E., J. Savarino, B. Operskalski, and P. S. Wang. 2006. "Suicide Risk during Antidepressant Treatment." *American Journal of Psychiatry* 163 (1): 41–7.
- Simon, T. R., A. C. Swann, K. E. Powell, L. B. Potter, M. Kresnow, and P. W. O'Carroll. 2001. "Characteristics of Impulsive Suicide Attempts and Attempters." *Suicide and Life-Threatening Behavior* 32 (suppl): 49–59.
- Spettell, C. M., T. C. Wall, J. Allison, J. Calhoun, R. Kobylinski, R. Fargason, and C. I. Kiefe. 2003. "Identifying Physician-Recognized Depression from Admin-

- istrative Data: Consequences for Quality Measurement.” *Health Service Research* 38 (4): 1081–102.
- Stettin, G. D., J. Yao, R. R. Verbrugge, and R. E. Aubert. 2006. “Frequency of Follow-Up Care for Adult and Pediatric Patients during Initiation of Antidepressant Therapy.” *American Journal of Managed Care* 12 (8): 453–61.
- Sturm, R. 1998. “Instrumental Variable Methods for Effectiveness Research.” *International Journal of Methods in Psychiatric Research* 7 (1): 17–26.
- U.S. Food and Drug Administration. 2004. “FDA Public Health Advisory: Suicidality in Children and Adolescents Being Treated with Antidepressant Medications” [accessed on September 10, 2006]. Available at <http://www.fda.gov/cder/drug/antidepressants/SSRIPHA200410.htm>
- U.S. Food and Drug Administration. 2005. “FDA Public Health Advisory: Suicidality in Adults Being Treated with Antidepressant Medication” [accessed on September 10, 2006]. Available at <http://www.fda.gov/cder/drug/antidepressants/default.htm>
- Valenstein, M., D. Eisenberg, J. F. McCarthy, K. L. Austin, D. Ganoczy, H. M. Kim, K. Zivin, J. D. Piette, M. Olfson, and F. C. Blow. 2009. “Service Implications of Providing Intensive Monitoring during High-Risk Periods for Suicide among VA Patients with Depression.” *Psychiatric Services* 60 (4): 439–44.
- Valenstein, M., H. K. Kim, D. Ganoczy, J. F. McCarthy, K. Zivin, K. L. Austin, K. Hoggatt, D. Eisenberg, J. D. Piette, F. C. Blow, and M. Olfson. 2007. *Periods of Increased Suicide Risks among VA Patients Receiving Depression Treatment: System and Policy Implications*. Nineteenth NIMH Research Conference on Mental Health Services MHSR. Washington, DC.
- Zivin, K., H. M. Kim, J. F. McCarthy, K. L. Austin, K. J. Hoggatt, H. M. Walters, and M. Valenstein. 2007. “Suicide Mortality among Individuals Receiving Treatment for Depression in the Veterans Affairs Health System: Associations with Patient and Treatment Setting Characteristics.” *American Journal of Public Health* 97 (12): 2193–8.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix SA2: First-Stage Regression Results from IV Analysis.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.