Predicting Clinically Significant Response to Primary Care Treatment for Depression from Electronic Health Records of Veterans

Mary C. Vance[1,2], M. Myron Chang[3], Jeremy B. Sussman[3,4], Kara Zivin[3,5], Paul N. Pfeiffer[3,5]

[1] Center for the Study of Traumatic Stress, Department of Psychiatry, Uniformed Services University, Bethesda, MD
[2] Henry M. Jackson Foundation for the Advancement of Military Medicine, Bethesda, MD
[3] VA Center for Clinical Management Research, VA Ann Arbor Healthcare System, Ann Arbor, MI
[4] Department of Internal Medicine, University of Michigan, Ann Arbor, MI
[5] Department of Psychiatry, University of Michigan, Ann Arbor, MI

Abstract
To reduce delays in referral to specialty mental health care, we evaluated clinical prediction models estimating the likelihood of response to primary care treatment of depression in the VA healthcare system. We included patients with a primary care depression diagnosis between October 1, 2015 and December 31, 2017, an initial Patient Health Questionnaire (PHQ-9) score ≥ 10 within 30 days, a follow-up PHQ-9 score within 2-8 months, and no specialty mental health care within three months prior to depression diagnosis. We evaluated eight ordinary least squares regression models, each with a different procedure for selecting predictors of percentage change in PHQ-9 score from baseline to follow-up. Predictors included patient characteristics from electronic health records and neighborhood characteristics from US census data. We repeated each modeling procedure 1,000 times, using different training and validation sets of patients. We used R^2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) to evaluate model performance. The final cohort included 3,461 patients. The two best performing models included multiple iterations of backwards stepwise variable selection with R^2 of 0.063, RMSE of 41.56, MAE of 33.44; and R^2 of 0.064, RMSE of 41.55, MAE of 33.46. Model performance did not suggest its use as a guide in clinical decision-making. Future research should explore whether obtaining additional risk factor data from patients (e.g., duration of symptoms) or modeling PHQ-9 scores over a narrower time interval improves performance of clinical risk prediction tools for depression.

Key Words: Depression, predictive modeling, electronic health records

1. Introduction

Patients typically receive care for depression in primary care settings, and some models of integrating mental health specialists into primary care (e.g., collaborative care) improve depression outcomes compared to usual primary care (Butler et al., 2011, Gilbody et al., 2006 and Unützer et al., 2002). However, accurately identifying which patients with

depression may improve with primary care treatment (including collaborative care and other forms of integrated care, such as coordinated or co-located care) and which patients may not improve and therefore require more intensive treatment, such as in a specialty mental health clinic remains challenging. Currently, primary and integrated care providers decide to treat depression in primary care or refer to specialty care based on clinical judgment, weighing factors such as the patient's mental illness severity, complexity, and treatment history. However, reliance on clinical judgment alone could result in wide practice variations, with both unnecessary referrals to specialty care for patients likely to improve in primary care treatment and delays in referral to specialty care for patients unlikely to improve in primary care.

Using a clinical prediction tool could improve triage decisions in primary care depression management. Clinical prediction tools can guide clinical decision-making by stratifying patients into different risk categories using statistical models that use large data sources. Although several such tools predict trajectories of depression in currently depressed patients (Rubenstein et al., 2007, Dowrick et al., 2011 and Chondros et al 2018), prior studies have not examined the likelihood of response to treatment in primary care settings that include integrated care. These studies have also not typically used data from electronic health records (EHR), but rather from longitudinal cohorts. Tools using EHR data reflect real-world patient populations and treatment patterns which practices could implement without additional data collection.

The Department of Veterans Affairs (VA) health system routinely screens veterans for depression in primary care settings. Clinicians may decide either to initiate treatment for patients diagnosed with depression in the primary care or primary care-mental health integration (PCMHI) setting (a co-located mental health clinic) or refer the patient to a specialty mental health clinic. To inform the future development of a clinical prediction tool, we evaluated various clinical risk prediction models using VA health system EHR data to estimate the likelihood that depression treatment initiation in a VA primary care or PCMHI setting would result in clinically significant symptom improvement among veterans with depression. We hypothesized that variables derived from EHR data will help predict which patients with depression will improve in primary care.

2. Methods

2.1 Data sources

We used data from the VA Corporate Data Warehouse (CDW), a data repository which contains nationwide clinical, enrollment, financial, administrative, utilization, and benefits information for all patients who receive care through the VA (Department of Veterans Affairs, 2014). Additionally, we obtained characteristics of each US census tract from the US Census Bureau 2013-2017 American Community Survey 5-year estimates (US Census Bureau(a). 2019).

2.2 Patient cohort selection

The patient cohort included all patients who had:

(a) A diagnosis of a unipolar depressive disorder (ICD-10-CM codes F32.0-F32.5, F32.8, F32.9-F33.3, F33.40-F33.42, F33.8, F33.9, F34.1, F43.21, F43.23) recorded during an outpatient primary care/PCMHI encounter between October 1, 2015 and December 31, 2017; this represents the qualifying diagnosis for the patient,

(b) A baseline Patient Health Questionnaire-9 (PHQ-9) (Kroenke et al., 2001) score ≥ 10 (which indicates probable major depression) recorded within 30 days of the most recent qualifying diagnosis,

(c) An associated treatment initiation including antidepressant medication fills or psychotherapy that occurs within 30 days of the diagnosis,

(d) No prior antidepressant medication fills, psychotherapy, or specialty mental health treatment in the three months prior to diagnosis,

(e) A follow-up PHQ-9 score between two and eight months following treatment initiation,

(f) No completed specialty mental health clinic appointment prior to first follow-up PHQ-9 score,

(g) Not more than one skipped answer on the PHQ-9, and

(h) Neighborhood characteristics available from US census data based on patient residence census tract information within the CDW.

2.3 Outcome variables

For each patient, we computed the percentage depressive symptom change from the baseline PHQ-9 to the follow-up PHQ-9 survey, which defined the continuous PHQ-9 score outcome variable. We also determined whether each patient had a significant response (≥50% improvement) in their depression symptoms between baseline and follow-up, which defined the dichotomous PHQ-9 score outcome variable.

2.4 Predictors

We extracted the following patient demographic data from VA EHR: age, gender, race, Hispanic or Latino ethnicity, census tract of home residence, and marital status. We extracted patients' answers to individual items on the baseline PHQ-9, whether the patient's diagnosis was for major depressive disorder (yes/no), facility type of qualifying diagnosis (e.g., VA medical center, VA outpatient clinic), the patient's branch of military service, and the patient's service-connected VA disability rating expressed in multiples of ten (0% to 100%).

Clinical characteristics included the patient's most recent alcohol use disorder screening result as reported on the Alcohol Use Disorders Identification Test-Concise (AUDIT-C) (Bush et al., 1998), the patient's most recent posttraumatic stress disorder (PTSD) screening result as reported on the Primary Care PTSD Screen (PC-PTSD) (Prins, et al., 2003), and the number of outpatient mental health or substance use visits between six months and two years prior to qualifying diagnosis. We looked into prior antidepressant use over two different time frames: at least six months prior to qualifying diagnosis, and within two years prior to the initial PHQ-9 survey date. We also examined the number of homelessness service visits, number of inpatient residential stays, number of inpatient residential days, number of outpatient medical visits, and number of psychiatric or substance emergency department visits in the two years prior to the qualifying depression diagnosis. Finally, we included medical comorbidities for the 31 Elixhauser categories, represented as 31 separate binary indicators, along with the patient's total Elixhauser score for the past two years computed as the sum of the 31 indicators (Elixhauser et al., 1998 and Quan et al., 2005); psychiatric comorbidities (PTSD, other anxiety disorder, serious mental illness, nicotine use disorder, alcohol use disorder, and substance use disorder) for the prior two years (Zivin et al., 2015, Zivin et al., 2016); and prior non-antidepressant psychotropic medication fills for the past two years.

The neighborhood characteristics included the census tract percentages for: male, veteran, black or African-American, age 65 years or older, age 25 years and over with less

than a high school education, unemployed, below the federal poverty level for the past 12 months, in a female-headed household with no married spouse and with family, in an owner-occupied housing unit, and in housing units lacking complete plumbing facilities (US Census Bureau(b), 2019).

2.5 Modeling methods

We began with four linear models, three of which used benchmark models and one main model. The benchmark models included: 1) with all possible predictors, 2) using no predictors, and 3) using four predictors predetermined for inclusion due to their likely clinical significance, based on previous literature (Panaite et al., 2019). These predictors included black race, presence of PTSD comorbidity, service-connected VA disability rating, and initial PHQ-9 score. The main model involved a backward stepwise selection approach to identify which predictors were significantly associated with PHQ-9 percentage improvement. Using this method, we aimed to have both a better-performing prediction model and a listing of significant variables for further study.

During the modeling process, we noticed that the backward stepwise selection approach was not consistent in its results. The significant group of predictors in one training set may vary substantially from a group of predictors in another training set. In light of this issue, we created a second main model, which contained only the predictors that were consistently significant predictors across multiple training sets; we hypothesized that these would likely successfully serve as standalone predictors of depression improvement.

To address the possibility of non-linear relationships across the predictor variables, we included all 77 predictors (Appendix 1) in the construction of a regression tree model. A regression tree model iteratively selects a single variable that best splits the data into two groups, until no further improvement can be made, and then "prunes" the tree back based on a cost function of model complexity. This process highlighted three predictors that were likely to have significant pairwise interaction effects: initial PHQ-9 score, service-connected VA disability rating, and PC-PTSD score. We incorporated these interactions into our modeling procedures for the benchmark model with all possible predictors and the two main models, resulting in three additional models for analysis. We evaluated these three non-linear models to determine whether including these interactions led to increased model performance over the other linear models.

In total, we created eight ordinary least squares (OLS) regression models, each with a different procedure for determining which predictor variables to use. The models are as follows:

Model One (Full Model) – We used all 77 available predictors to predict percentage improvement in the PHQ-9.

Model Two (Intercept-Only Model) –We used a single intercept equal to the mean percentage of PHQ-9 improvement.

Model Three (Model with Pre-Specified Predictors) – We used four predictors identified from prior work (Panaite et al., 2019) as significantly associated with depression improvement: black race, presence of PTSD comorbidity, service-connected VA disability rating, and initial PHQ-9 score.

Model Four (Model from Backward Stepwise Selection) – We used a backward stepwise selection approach following Dowrick and colleagues' approach (Dowrick et al., 2011). Starting with a model containing all 77 predictors, we removed each single predictor and compared the resulting model's Akaike's Information Criterion (AIC) to the AIC of the starting model. Then we removed the predictor associated with the largest reduction in AIC. We reiterated this process until the removal of any additional predictor caused the AIC to increase. We modified the backward

stepwise selection approach to always retain the Model Three predictors regardless of their effects on AIC.

Model Five (Model with Common Predictors) –To address the notion that predictors for Model Four may have depended on which patients we included in the training set, we ran five preliminary iterations of backward stepwise selection, using a different training set each time, and recorded what predictors Model Four selected. Model Five included predictors that remained in all five iterations of Model Four.

Model Six (Full Model with Interaction Terms) – Using forwards selection after exploratory modeling, we found three interaction terms that retained statistical significance: an interaction between the patient's VA disability rating and initial PHQ-9 score, an interaction between the patient's PC-PTSD score and VA disability rating, and an interaction between the patient's PC-PTSD score and initial PHQ-9 score. We added these terms to the full list of predictors from Model One, resulting in a model with 80 predictors.

Model Seven (Model from Backward Stepwise Selection with Interaction Terms) – We developed this analogously to Model Four, except starting with Model Six's 80 predictors instead of Model One's 77 predictors.

Model Eight (Model with Common Predictors and Interaction Terms) – We used the same process to create Model Eight from Model Seven as we did to select predictors for Model Five based on Model Four's results.

2.6 Model evaluation

From our patient cohort, we identified and kept only those patients with complete non-missing data for all the predictors. Then we split our patient cohort into an 80% training set and a 20% validation set. We did this 1,000 times, resulting in 1,000 different training sets and 1,000 different validation sets of patients. With each training set, we developed the eight OLS regression models using our continuous outcome variable. Then, we used each model to predict PHQ-9 percentage change for the patients in each respective validation set, using R^2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) to evaluate model performance. With 1,000 repetitions, we ended up with 1,000 recorded R^2, RMSE, and MAE values for each model. Finally, we took the averages of these values to obtain the performance measures for the eight models.

In further analysis, we developed logistic regression analogues of Models One through Six and Eight using the dichotomous outcome variable. The process was similar to above: we split the patient cohort into an 80% training set and a 20% validation set 1,000 times, we developed the logistic models using each training set, and we used each model to predict significant response in each respective validation set. To evaluate performance, we recorded Area Under the Curve (AUC), Brier score, calibration intercept, calibration slope, and calibration adjusted R^2 for each model, and took their averages.

We conducted all analyses using R version 3.5.1.

3. Results

3.1 Descriptive statistics

Table 1 shows selected descriptive statistics for our final sample (N = 3,461), which consisted of all patients with complete non-missing data. We repeatedly split this sample into an 80% training set (N = 2,769) and a 20% validation set (N = 692) for further analysis.

3.2 Modeling results

Table 2 shows each OLS model's average R^2, RMSE, and MAE over 1,000 training runs. The R^2 for the OLS models were low, ranging from 0.041 to 0.064. The error measures, RMSE and MAE, were correspondingly high and ranged from 41.55 to 42.90 and 33.44 to 34.42, respectively. Models Five (R^2 of 0.06, RMSE of 41.56, MAE of 33.44) and Eight (R^2 of 0.06, RMSE of 41.55, MAE of 33.46) were modestly more effective than the other models.

Figure 1 shows the calibration for the OLS models. Flexible regression lines for each model were plotted using their respective validation set predictions. Additionally, Model 5's individual predictions, which are representative of all the models, are plotted as red circles. This demonstrates that the vast majority of the linear models' predictions lay between 0% and 50% improvement, even though the observed improvements ranged between -150% to 100%.

For the logistic models, AUCs varied from 0.50 to 0.61, Brier scores varied from 0.22 to 0.25, calibration intercepts varied from 0.02 to 0.28, calibration slopes varied from 0.18 to 0.98, and calibration adjusted R^2 varied from 0.51 to 0.86. Models Five and Eight had the best performance by a small margin, with higher AUCs, lower Brier scores, and desirable calibration plots (Table 3).

We also tested for significant differences depending on type of treatment (medications, psychotherapy, or both). Treatment type did not significantly alter our results.

4. Discussion

In this study, we sought to develop a clinical prediction model to determine the likelihood that depression treatment in a VA primary care/PCMHI setting would result in clinically significant symptom improvement among veterans with depression who had not recently been treated, using EHR and census data. We developed a primary model based on backward stepwise selection and compared its predictive performance to three benchmark models. In the process, we noted the instability of the stepwise selection process and constructed a simpler, more consistent fifth model to address that issue. Then we built three models that allowed for non-linear interactions.

We found that our models for predicting percentage improvement in PHQ-9 scores after primary care/PCMHI treatment for depression within 30 days of diagnosis performed modestly, with R^2 for the models never higher than 6%. Models based on a simple but not predetermined set of variables did not substantially improve model performance. Our tool's ability to predict depression improvement was poor compared to previously developed prognostic models predicting depression onset, trajectory, and treatment outcome (Chondros et al 2018, King et al., 2008, Zuithoff et al., 2009, Bellon et al., 2011, King et al., 2013, Perlis, 2013, Vohringer et al., 2013 and Chekroud et al., 2016) and also compared to most commonly reported prognostic models (Steyerberg, 2009). In fact, our models performed marginally or no better than chance. These models proved inadequate to serve as clinical risk prediction tools to inform depression treatment in the primary care/PCMHI setting.

We suggest several possible reasons for the modest results. The clinical data used in this study may be missing valuable predictors that are not included in the EHR, such as duration of current depression symptoms and psychosocial supports and stressors. Although screening for depression is required across VA primary care settings, only a minority of patients diagnosed with depression had sufficient baseline and follow-up PHQ-9 data in extractable EHR data fields to be included in this study. Furthermore, because our study aimed to predict depression improvement with treatment in primary care/PCMHI alone, we excluded veterans with completed specialty mental health clinic visits prior to

their follow-up PHQ-9, which could have effectively excluded veterans with more severe depression. These factors produced a highly-selected sample that may have removed the variation that would be needed to predict outcomes. Additionally, follow-up PHQ-9 scores occurred across a wide time interval (i.e., 2 to 8 months), which may also have reduced our models' ability to predict outcomes.

Certain other limitations and cautions about our study design should be noted. In model development, we used approaches that examined all predictors and selected significant predictors based on the training set in order to improve model performance. This could lead to the models overfitting on the training data. The backward stepwise selection approach kept identifying different predictors, which validated this issue. To address potential overfitting, we only evaluated model performance on validation data not used in model development. This way, the reported results would better reflect how well each model could predict PHQ-9 percentage improvement for the general population of VA patients with depression.

As our sample was drawn from VA users, this model cannot be extrapolated to the non-VA population. Because our study relied on EHR data, some degree of missingness and misclassification error is unavoidable, and psychiatric and medical diagnoses could not be verified. We obtained PHQ-9 data from CDW data fields that are populated by manually entering each PHQ-9 item via an EHR note template designed for this purpose or a separate data entry application. We were not able to extract PHQ-9 data collected outside of these systems, including data recorded only in the free text of notes. As such, some diagnostic information for depression may have been missed. Finally, as referenced above, information on some potential confounders could not be obtained from EHR.

These limitations and cautions notwithstanding, our study provides a foundation for the future development of clinical prediction tools to guide mental health treatment using EHR data. Further refinement of our clinical prediction models could improve model performance and utility for clinical practice. For example, additional patient-reported variables (e.g., functional impairment) can be collected using structured data fields in existing EHR, and their inclusion in predictive models may improve results. Additionally, as the use of measurement-based care increases and PHQ-9 data as well as other systematically collected patient data become more readily available, clinical prediction tools developed with such data are likely to improve.

Acknowledgements

References

Bellon, J. A., J. de Dios Luna, M. King, B. Moreno-Kustner, I. Nazareth, C. Monton-Franco, M. J. GildeGomez-Barragan, M. Sanchez-Celaya, M. A. Diaz-Barreiros, C. Vicens, J. A. Cervilla, I. Svab, H. I. Maaroos, M. Xavier, M. I. Geerlings, S. Saldivia, B. Gutierrez, E. Motrico, M. T. Martinez-Canavate, B. Olivan-Blazquez, M. S.

Sanchez-Artiaga, S. March, M. del Mar Munoz-Garcia, A. Vazquez-Medrano, P. Moreno-Peral and F. Torres-Gonzalez (2011). "Predicting the onset of major depression in primary care: international validation of a risk prediction algorithm from Spain." Psychol Med 41(10): 2075-2088.

Bush, K., D. R. Kivlahan, M. B. McDonell, S. D. Fihn and K. A. Bradley (1998). "The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test." Arch Intern Med 158(16): 1789-1795.

Butler, M., R. L. Kane, D. McAlpine, R. Kathol, S. S. Fu, H. Hagedorn and T. Wilt (2011). "Does integrated care improve treatment for depression? A systematic review." J Ambul Care Manage 34(2): 113-125.

Chekroud, A. M., R. J. Zotti, Z. Shehzad, R. Gueorguieva, M. K. Johnson, M. H. Trivedi, T. D. Cannon, J. H. Krystal and P. R. Corlett (2016). "Cross-trial prediction of treatment outcome in depression: a machine learning approach." Lancet Psychiatry 3(3): 243-250.

Chondros, P., S. Davidson, R. Wolfe, G. Gilchrist, C. Dowrick, F. Griffiths, K. Hegarty, H. Herrman and J. Gunn (2018). "Development of a prognostic model for predicting depression severity in adult primary patients with depressive symptoms using the diamond longitudinal study." J Affect Disord 227: 854-860.

Department of Veterans Affairs (2014). "Corporate data warehouse (CDW)." Retrieved from https://www.hsrd.research.va.gov/for_researchers/vinci/cdw.cfm on December 28, 2019,.

Dowrick, C., C. Flach, M. Leese, J. Chatwin, R. Morriss, R. Peveler, M. Gabbay, R. Byng, M. Moore, A. Tylee and T. Kendrick (2011). "Estimating probability of sustained recovery from mild to moderate depression in primary care: evidence from the THREAD study." Psychol Med 41(1): 141-150.

Elixhauser, A., C. Steiner, D. R. Harris and R. M. Coffey (1998). "Comorbidity measures for use with administrative data." Med Care 36(1): 8-27.

Gilbody, S. , Bower, P. , Fletcher, J. , Richards, D. & Sutton, A. J. (2006). Collaborative Care for Depression. Archives of Internal Medicine, 166(21), 2314–2321.

King, M., C. Bottomley, J. Bellon-Saameno, F. Torres-Gonzalez, I. Svab, D. Rotar, M. Xavier and I. Nazareth (2013). "Predicting onset of major depression in general practice attendees in Europe: extending the application of the predictD risk algorithm from 12 to 24 months." Psychol Med 43(9): 1929-1939.

King, M., C. Walker, G. Levy, C. Bottomley, P. Royston, S. Weich, J. A. Bellon-Saameno, B. Moreno, I. Svab, D. Rotar, J. Rifel, H. I. Maaroos, A. Aluoja, R. Kalda, J. Neeleman, M. I. Geerlings, M. Xavier, I. Carraca, M. Goncalves-Pereira, B. Vicente, S. Saldivia, R. Melipillan, F. Torres-Gonzalez and I. Nazareth (2008). "Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees: the PredictD study." Arch Gen Psychiatry 65(12): 1368-1376.

Kroenke, K., R. L. Spitzer and J. B. Williams (2001). "The PHQ-9: validity of a brief depression severity measure." J Gen Intern Med 16(9): 606-613.

Panaite, V., N. W. Bowersox, K. Zivin, D. Ganoczy, H. M. Kim and P. N. Pfeiffer (2019). "Individual and neighborhood characteristics as predictors of depression symptom response." Health Serv Res 54(3): 586-591.

Perlis, R. H. (2013). "A clinical risk stratification tool for predicting treatment resistance in major depressive disorder." Biol Psychiatry 74(1): 7-14.

Prins, A., Ouimette, P., Kimerling, R., Cameron, R. P., Hugelshofer, D. S., Shaw-Hegwer, J., Thrailkill, A., Gusman, F.D., Sheikh, J. I. (2003). "The Primary Care PTSD Screen (PC-PTSD): Development and operating characteristics." Primary Care Psychiatry 9: 9-14.

Quan, H., V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J. C. Luthi, L. D. Saunders, C. A. Beck, T. E. Feasby and W. A. Ghali (2005). "Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data." Med Care 43(11): 1130-1139.

Rubenstein, L. V., N. R. Rayburn, E. B. Keeler, D. E. Ford, K. M. Rost and C. D. Sherbourne (2007). "Predicting outcomes of primary care patients with major depression: development of a depression prognosis index." Psychiatr Serv 58(8): 1049-1056.

Steyerberg, E. W. (2009). Clinical prediction models: a practical approach to development, validation, and updating. New York, Springer.

Unützer, J., Katon, W., Callahan, C. M., Williams, J. W., Jr, Hunkeler, E., Harpole, L., Hoffing, M., Della Penna, R. D., Noël, P. H., Lin, E. H., Areán, P. A., Hegel, M. T., Tang, L., Belin, T. R., Oishi, S., Langston, C., & IMPACT Investigators. Improving Mood-Promoting Access to Collaborative Treatment (2002). Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. JAMA, 288(22), 2836–2845.

US Census Bureau(a). "About the American Community Survey." Retrieved from https://www.census.gov/programs-surveys/acs/about.html on November 5, 2019.

US Census Bureau(b). "American Community Survey and Puerto Rico Community Survey 2018 Subject Definitions." Retrieved November 5, 2019, from https://www2.census.gov/programs-surveys/acs/tech_docs/subject_definitions/2018_ACSSubjectDefinitions.pdf.

Vohringer, P. A., M. I. Jimenez, M. A. Igor, G. A. Fores, M. O. Correa, M. C. Sullivan, N. S. Holtzman, E. A. Whitham, S. A. Barroilhet, K. Alvear, T. Logvinenko, D. M. Kent and S. N. Ghaemi (2013). "A clinical predictive score for mood disorder risk in low-income primary care settings." J Affect Disord 151(3): 1125-1131.

Zivin, K., H. M. Kim, M. Yosef, D. T. Maust, M. Valenstein, E. G. Smith and D. S. Davydow (2016). "Antidepressant Medication Treatment and Risk of Death." J Clin Psychopharmacol 36(5): 445-452.

Zivin, K., M. Yosef, E. M. Miller, M. Valenstein, S. Duffy, H. C. Kales, S. Vijan and H. M. Kim (2015). "Associations between depression and all-cause and cause-specific risk of death: a retrospective cohort study in the Veterans Health Administration." J Psychosom Res 78(4): 324-331.

Zuithoff, N. P., Y. Vergouwe, M. King, I. Nazareth, E. Hak, K. G. Moons and M. I. Geerlings (2009). "A clinical prediction rule for detecting major depressive disorder in primary care: the PREDICT-NL study." Fam Pract 26(4): 241-250.

Table 1: Selected Descriptive Statistics for Patient Cohort (N = 3,461)

	Numerical Variables		Categorical Variables
	Mean (SD)	No. > 0 (%)	No. (%)
PHQ-9			
Baseline score	15.51 (4.11)	--	--
Follow-up score	10.81 (6.55)	--	--
% improvement	28.76 (42.84)	2,531 (73.13%)	--
Demographics			
Age	52.35 (15.30)	--	--
Male	--	--	2,792 (80.67%)
White	--	--	2,068 (59.75%)
Black	--	--	1,043 (30.14%)
Hispanic or Latino	--	--	414 (11.96%)
Married	--	--	1,844 (53.28%)
Comorbidities (past 2 years)			
Elixhauser	--	--	3,119 (90.12%)
PTSD	--	--	636 (18.38%)
Anxiety disorder	--	--	990 (28.60%)
Serious mental illness	--	--	384 (11.10%)
Nicotine use disorder	--	--	351 (10.14%)
Alcohol use disorder	--	--	227 (6.56%)
Substance use disorder	--	--	165 (4.77%)
Hospital Visits (past 2 years)			
Inpatient residential stays	0.14 (0.55)	293 (8.47%)	--
ED visits	0.04 (0.43)	52 (1.50%)	--
Outpatient medical visits	12.94 (14.18)	3,349 (96.76%)	--
Mental health/substance use visits*	1.38 (5.56)	937 (27.07%)	--
Homelessness services	0.16 (1.55)	145 (4.19%)	--
Clinical Information			
Major depressive disorder	--	--	2,730 (78.88%)
VA disability rating	48.79 (37.09)	2,631 (76.02%)	--
AUDIT-C score (max. score of 12)**	1.62 (2.16)	2,059 (59.49%)	--
PC-PTSD score (max. score of 4)**	1.35 (1.60)	1,698 (49.06%)	--
Prior Psychotropic Usage (past 2 years)			
Alcohol/drug addiction treatments	--	--	22 (0.64%)
Antiparkinson/antihistaminic agents	--	--	298 (8.61%)
Benzodiazepines	--	--	263 (7.60%)
Other anti-anxiety agents	--	--	70 (2.02%)
Other hypnotics	--	--	132 (3.81%)
Mood stabilizers	--	--	790 (22.83%)
Antidepressants***	--	--	1,646 (47.56%)
Antipsychotics	--	--	73 (2.11%)
Stimulants/ADHD agents	--	--	16 (0.46%)
Anticolinesterases/dementia meds	--	--	17 (0.49%)
Census Tract Characteristics			
% veteran	9.99 (5.59)	--	--
% ≥ 25 with less than high school	13.00 (9.60)	--	--
% female-headed household	14.70 (7.87)	--	--
% unemployed	7.69 (5.24)	--	--
% below the poverty line	16.33 (11.75)	--	--
% with owner-occupied housing units	63.11 (20.38)	--	--
% lacking complete plumbing facilities****	0.37 (0.86)	--	--
% male	48.90 (3.84)	--	--
% black	18.47 (23.44)	--	--
% over 65 years old	14.24 (7.49)	--	--

* Mental health/substance use visits are measured from 6 months to 2 years prior to the qualifying diagnosis; all other hospital visit types are measured from the qualifying diagnosis date to 2 years prior to the qualifying diagnosis

** AUDIT-C and PC-PTSD scores are derived from multiple-choice questionnaires, where each choice adds a value towards the total score; AUDIT-C scores ranged from 0 to 12 and PC-PTSD scores ranged from 0 to 4

*** Prior antidepressant usage is counted if the use occurs within 30 days of qualifying diagnosis, or if the use occurs from 3 months to 2 years prior to the qualifying diagnosis

**** Only 1,082 patients (31.26% of the cohort) had a positive value for the census tract characteristic "% lacking complete plumbing facilities"; all other census tract characteristics had over 90% positive values

Table 2: Performance Measures for Linear Regression Models Predicting % Improvement in PHQ-9 Scores

	R^2	RMSE	MAE
Model One (Full Model)	0.041	42.38	33.97
Model Two (Intercept-Only Model)	NA	42.90	34.42
Model Three (Model With Pre-Specified Predictors)	0.042	42.03	33.99
Model Four (Model From Backward Stepwise Selection)	0.048	42.01	33.77
Model Five (Model With Common Predictors)	0.063	41.56	33.44
Model Six (Full Model With Interaction Terms)	0.041	42.39	33.99
Model Seven (Model From Backward Stepwise Selection With Interaction Terms)	0.048	42.01	33.78
Model Eight (Model With Common Predictors and Interaction Terms)	0.064	41.55	33.46

Abbreviations: RMSE, Root Mean Squared Error; MAE, Mean Absolute Error

Figure 1: Calibration Plot for Linear Regression Models Predicting % Improvement in PHQ-9 Scores
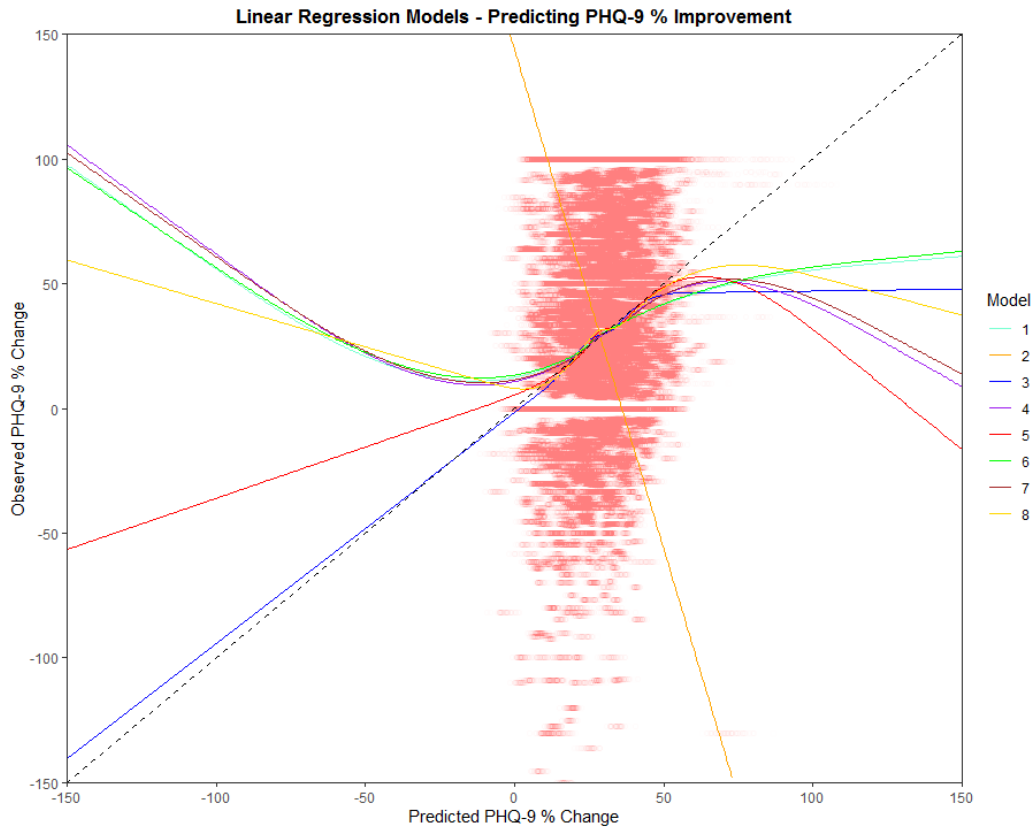
Table 3: Performance Measures for Logistic Regression Models Predicting 50% or Greater Improvement in PHQ-9 Scores

	AUC	Brier Score	Calibration Intercept (SD)	Calibration Slope (SD)	Calibration Adj. R^2 (SD)
Model One (Full Model)	0.58	0.23	0.23 (0.08)	0.35 (0.25)	0.59 (0.31)
Model Two (Intercept-Only Model)	0.5	0.23	NA	NA	NA
Model Three (Model With Pre-Specified Predictors)	0.59	0.22	0.05 (0.11)	0.87 (0.32)	0.61 (0.23)
Model Four (Model From Backward Stepwise Selection)	0.59	0.22	0.18 (0.11)	0.51 (0.32)	0.66 (0.30)
Model Five (Model With Common Predictors)	0.61	0.22	0.02 (0.14)	0.98 (0.42)	0.86 (0.19)
Model Six (Full Model With Interaction Terms)	0.57	0.25	0.28 (0.04)	0.18 (0.10)	0.51 (0.29)
Model Eight (Model With Common Predictors and Interaction Terms)	0.61	0.22	0.11 (0.12)	0.74 (0.34)	0.80 (0.25)

Abbreviations: AUC, Area Under the Curve

Appendix 1: List of 77 Predictors

1. The patient's age on the date of diagnosis
2. The patient's gender
3. Binary indicator if the patient's self-identified race is "Black or African American" (multiracial patients are not included)
4. The patient's ethnicity
5. The patient's marriage status closest to the date of diagnosis, or if not available then the patient's current marriage status
6. The patient's disability rating using the VA's procedure, in 10% intervals[1]
7. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Congestive Heart Failure"
8. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Cardiac Arrhythmias"
9. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Valvular Disease"
10. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Pulmonary Circulation Disorders"
11. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Peripheral Vascular Disorders"
12. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Hypertension, uncomplicated"
13. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Hypertension, complicated"
14. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Paralysis"
15. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Other Neurological Disorders"
16. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Chronic Pulmonary Disease"
17. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Diabetes, uncomplicated"
18. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Diabetes, complicated"
19. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Hypothyroidism"
20. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Renal Failure"
21. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Liver Disease"
22. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Peptic Ulcer Disease Excluding Bleeding"
23. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "AIDS/HIV"
24. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Lymphoma"

---

[1] U.S. Department of Veterans Affairs, Veterans Benefits Administration. Compensation – Benefit Rates. Accessed at https://www.benefits.va.gov/compensation/rates-index.asp on February 27, 2020.

25. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Metastatic Cancer"
26. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Solid Tumor Without Metastasis"
27. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Rheumatoid Arthritis/Collagen Vascular Diseases"
28. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Coagulopathy"
29. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Obesity"
30. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Weight Loss"
31. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Fluid and Electrolyte Disorders"
32. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Blood Loss Anemia"
33. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Deficiency Anemia"
34. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Alcohol Abuse"
35. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Drug Abuse"
36. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Psychoses"
37. Binary indicator if the patient had an ICD code within two years prior to diagnosis for the Elixhauser category "Depression"
38. Percent of veterans out of the civilian population 18 years and over within the patient's census tract
39. Percent of 25-year-olds and over with less than a high school education within the patient's census tract
40. Percent of total family households with female householder, no husband present, and family within the patient's census tract
41. Percent of civilian labor force that is unemployed within the patient's census tract
42. Percent of all people whose income in the past 12 months is below the poverty level within the patient's census tract
43. Percent of occupied housing units that are owner-occupied within the patient's census tract
44. Percent of occupied housing units lacking complete plumbing facilities within the patient's census tract
45. Percent of total population that is male within the patient's census tract
46. Percent of total population that is black (one race) within the patient's census tract
47. Percent of total population that is 65 years or older within the patient's census tract
48. Binary indicator if the patient had a psychiatric comorbidity for PTSD within two years prior to diagnosis
49. Binary indicator if the patient had a psychiatric comorbidity for other anxiety disorders within two years prior to diagnosis
50. Binary indicator if the patient had a psychiatric comorbidity for serious mental illness within two years prior to diagnosis

51. Binary indicator if the patient had a psychiatric comorbidity for nicotine use disorder within two years prior to diagnosis
52. Binary indicator if the patient had a psychiatric comorbidity for alcohol use disorder within two years prior to diagnosis
53. Binary indicator if the patient had a psychiatric comorbidity for other substance use disorders within two years prior to diagnosis
54. Number of inpatient hospitalizations within two years prior to diagnosis
55. Number of days spent in inpatient hospitalizations/stays within two years prior to diagnosis
56. Number of psychiatric/substance ED visits within two years prior to diagnosis
57. Number of outpatient mental health/substance use visits within two years prior to diagnosis
58. Number of outpatient medical visits within two years prior to diagnosis
59. The patient's score on the initial/qualifying PHQ-9 survey
60. The patient's answer on the tenth (unscored) question in the initial PHQ-9 survey
61. The patient's AUDIT-C score closest to the date of diagnosis
62. The patient's PC PTSD score closest to the date of diagnosis
63. Binary indicator if the patient has prior antidepressant use 6+ months before diagnosis
64. Binary indicator if the patient has had prior alcohol and drug addiction treatment psychotropic medication use within two years prior to initial PHQ-9 survey
65. Binary indicator if the patient has had prior antiparkinson/Antihistamine psychotropic medication use within two years prior to initial PHQ-9 survey
66. Binary indicator if the patient has had prior benzodiazepine psychotropic medication use within two years prior to initial PHQ-9 survey
67. Binary indicator if the patient has had prior antianxiety psychotropic medication use within two years prior to initial PHQ-9 survey
68. Binary indicator if the patient has had prior hypnotics psychotropic medication use within two years prior to initial PHQ-9 survey
69. Binary indicator if the patient has had prior mood stabilizer psychotropic medication use within two years prior to initial PHQ-9 survey
70. Binary indicator if the patient has had prior antidepressant psychotropic medication use within two years prior to initial PHQ-9 survey
71. Binary indicator if the patient has had prior antipsychotic psychotropic medication use within two years prior to initial PHQ-9 survey
72. Binary indicator if the patient has had prior stimulant or ADD drug psychotropic medication use within two years prior to initial PHQ-9 survey
73. Binary indicator if the patient has had prior anticholinesterase psychotropic medication use within two years prior to initial PHQ-9 survey
74. The type of facility at the patient's diagnosis visit
75. Binary indicator if the patient's diagnosis is classified as Major Depressive Disorder
76. Number of outpatient visits utilizing homelessness services within two years prior to diagnosis
77. The patient's branch of service closest to the date of diagnosis