Word count:
Abstract: 250
Text: 4,498
Tables: 1
Figures: 4
Supplementary Tables: 8
Supplementary Figures: 1

Development of a model to predict antidepressant treatment response for depression among Veterans

June 2022

Victor Puac-Polanco[1], Hannah N. Ziobrowski[1], Eric L. Ross[2,3,4], Howard Liu[1,5], Brett Turner[1,5,6], Ruifeng Cui[7,8], Lucinda B. Leung[9,10], Robert M. Bossarte[5,11], Corey Bryant[12], Jutta Joormann[13], Andrew A. Nierenberg[4,14], David W. Oslin[15,16], Wilfred R. Pigeon[5,17], Edward P. Post[12,18], Nur Hani Zainal[1], Alan M. Zaslavsky[1], Jose R. Zubizarreta[1,19,20], Alex Luedtke[21,22], Chris J. Kennedy[3,4], Andrea Cipriani[23], Toshiaki A. Furukawa[24], Ronald C. Kessler[1]

[1] Department of Health Care Policy, Harvard Medical School, Boston, MA, USA; [2] Department of Psychiatry, McLean Hospital, Belmont, MA, USA; [3] Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA; [4] Department of Psychiatry, Harvard Medical School, Boston, MA, USA; [5] Center of Excellence for Suicide Prevention, Canandaigua VA Medical Center, Canandaigua, NY, USA; [6] Harvard T.H. Chan School of Public Health, Boston, MA, USA; [7] VISN 4 Mental Illness Research, Education and Clinical Center, VA Pittsburgh Health Care System, Department of Veterans Affairs, Pittsburgh, PA, USA; [8] Department of Psychiatry, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA; [9] Center for the Study of Healthcare Innovation, Implementation, and Policy, VA Greater Los Angeles Healthcare System, Los Angeles, CA, USA; [10] Division of General Internal Medicine and Health Services Research, UCLA David Geffen School of Medicine, Los Angeles, CA, USA; [11] Department of Behavioral Medicine and Psychiatry, West Virginia University, Morgantown, WV, USA; [12] Center for Clinical Management Research, VA Ann Arbor, Ann Arbor, MI, USA; [13] Department of Psychology, Yale University, New Haven, CT, USA; [14] Department of Psychiatry, Dauten Family Center for Bipolar Treatment Innovation, Massachusetts General Hospital, Boston, MA, USA; [15] VISN 4 Mental Illness Research, Education, and Clinical Center, Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA; [16] Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA; [17] Department of Psychiatry, University of Rochester Medical Center, Rochester, NY, USA; [18] Department of Medicine, University of Michigan Medical School, Ann Arbor, MI, USA; [19] Department of Statistics, Harvard University, Cambridge, MA, USA; [20] Department of Biostatistics, Harvard University, Cambridge, MA, USA; [21] Department of Statistics, University of Washington, Seattle, WA, USA; [22] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA; [23] Department of Psychiatry, University of Oxford, Oxford, UK; [24] Department of Health Promotion and Human Behavior, School of Public Health, Kyoto University Graduate School of Medicine, Kyoto, Japan

Author for correspondence: Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Ave, Boston, MA, 02115, USA, Email: kessler@hcp.med.harvard.edu

Financial Support:

This research was supported by the Office of Mental Health Services and Suicide Prevention and Center of Excellence for Suicide Prevention (Bossarte), the National Institute of Mental Health of the National Institutes of Health (R01MH121478, Kessler), the United States Department of Veterans Affairs Health Services Research & Development Service Career Development Award (IK2 HX002867, Leung), the PCORI Project Program Award (ME-2019C1-16172, Zubizarreta), and the Advanced Fellowship from the VISN 4 Mental Illness Research, Education, & Clinical Center (MIRECC, Cui, Oslin). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the U.S. Department of Veterans Affairs, or the United States Government.

ABSTRACT

Background: Only a limited number of patients with major depressive disorder (MDD) respond to a first course of antidepressant medication (ADM). We investigated the feasibility of creating a baseline model to determine which these would be among patients beginning ADM treatment in the US Veterans Health Administration (VHA).

Methods: A 2018-2020 national sample of n=660 VHA patients receiving ADM treatment for MDD completed an extensive baseline self-report assessment near the beginning of treatment and a 3-month self-report follow-up assessment. Using baseline self-report data along with administrative and geospatial data, an ensemble machine learning method was used to develop a model for 3-month treatment response defined by the Quick Inventory of Depression Symptomatology Self-Report and a modified Sheehan Disability Scale. The model was developed in a 70% training sample and tested in the remaining 30% test sample.

Results: 35.7% of patients responded to treatment. The prediction model had an area under the ROC curve (SE) of 0.66 (0.04) in the test sample. A strong gradient in probability (SE) of treatment response was found across three subsamples of the test sample using training sample thresholds for high [45.6% (5.5)], intermediate [34.5% (7.6)], and low [11.1% (4.9)] probabilities of response. Baseline symptom severity, comorbidity, treatment characteristics (expectations, history, and aspects of current treatment), and protective/resilience factors were the most important predictors.

Conclusions: Although these results are promising, parallel models to predict response to alternative treatments based on data collected before initiating treatment would be needed for such models to help guide treatment selection.

Keywords: Antidepressant medication; Clinical decision support; Depression; Machine learning; Treatment response; Veterans Health Administration.

INTRODUCTION

Major depressive disorder (MDD) has high prevalence and high impairment (GBD 2019 Diseases and Injuries Collaborators, 2020). The two primary first-line MDD treatments are psychotherapy and antidepressant medication (ADM; Qaseem, Barry, & Kansagara, 2016). ADM is the more common treatment despite most patients preferring psychotherapy (McHugh, Whitton, Peckham, Welge, & Otto, 2013) due to lower cost and wider availability (Hockenberry, Joski, Yarbrough, & Druss, 2019). But some MDD patients do not respond to ADMs (Cipriani et al., 2018; Kazdin et al., 2021; Little, 2009) but do to psychotherapy or an ADM-psychotherapy combination. However, the latter treatments often are provided only after months of unsuccessful ADM treatment (Day et al., 2021). A meaningful proportion of patients drop out before receiving other treatments (Larson et al., 2021). A strategy to predict likelihood of responding before initiating ADM treatment could be of value.

Many multivariable models have been developed, typically using machine learning (ML) methods (Chekroud et al., 2021; Ermers, Hagoort, & Scheepers, 2020; Lee et al., 2018), to predict depression treatment response. Most such models can be faulted, though, for (i) low external validity because of restriction to clinical trial samples; (ii) focus on biomarkers infeasible to use in routine clinical practice; (iii) including many fewer predictors than documented in the literature; or (iv) suboptimal analytic methods.

The current report presents results of a study designed to address these problems by analyzing an observational sample of patients recruited near beginning ADM treatment and administered an extensive baseline battery of self-report questions to assess predictors of ADM treatment response found in previous studies. The patients were followed for 3 months to assess

treatment response. The data were analyzed using a state-of-the-art stacked generalization ML method.

MATERIALS AND METHODS

Sample

As detailed elsewhere (Puac-Polanco et al., 2021), a probability sample of patients beginning MDD outpatient treatment was selected from Veterans Health Administration (VHA) EHRs December 2018-June 2020. Inclusion criteria were: (i) beginning first outpatient MDD treatment in the past year; and (ii) receiving ADM prescription and/or psychotherapy referral. Exclusions were: (i) 12-month suicide attempt; (ii) lifetime diagnoses of bipolar disorder, nonaffective psychosis, dementia, intellectual disability, autism, Tourette's disorder, stereotyped movement disorder, or borderline intellectual functioning; (iii) lifetime prescriptions of mood stabilizers or antipsychotic medications (Supplementary Table 1). The exclusion of 12-month suicide attempts was made because such patients in VHA are placed on a high-risk list that leads to intensive case management, making the experiences of these patients unrepresentative of the more general patient population.

Recruitment letters were sent to 55,106 provisionally eligible patients the day after their first outpatient visit. The letter described the study purposes and the requirements of self-report web- or phone-based baseline assessment taking 45 minutes and at 3-months follow-up taking 20 minutes, with compensation of $50 and $25, respectively, for the two assessments. A team member then attempted to contact each patient over the next week (3 call attempts). Of the 17,000 reached, 6,298 agreed to participate and 4,164 completed the baseline assessment (Supplementary Figure 1). At baseline, 809 respondents had received an ADM prescription

without referral to psychotherapy and were otherwise eligible. The 660 of these 809 who completed the 3-month assessment are the focus of this report. The protocol was approved by the Institutional Review Board of Syracuse VA Medical Center, Syracuse, New York.

Measures

Treatment response: Two-week depressive symptoms were assessed with the 16-item Quick Inventory of Depression Symptomatology Self-Report (QIDS-SR; Rush et al., 2003). A modified version of the Sheehan Disability Scale (SDS; Leon, Olfson, Portera, Farber, & Sheehan, 1997) was used to assess role impairment by asking patients how much depression interfered with the ability to work, participate in family and home life, or participate in social activities in the past 2 weeks on a 0-10 visual analog scale with response options of *not at all* (0), *mildly* (1-3), *moderately* (4-6), *markedly* (7-9), and *extremely* (10) (Cronbach's α=0.85).

Treatment response was defined as either (i) a 3-month QIDS-SR score no more than half its baseline value or (ii) a baseline SDS score of 4-10 in *any* role impairment domain along with a 3-month SDS score of 0-3 in *all* such domains. A similar composite definition of ADM treatment response was used in previous research (Huang et al., 2018; Wang et al., 2018; Zilcha-Mano et al., 2021).

Predictors: Numerous recent reports have carried out reviews or meta-analyses of research on baseline predictors of response to individual types of depression treatment (e.g., Furukawa et al., 2021; Noma et al., 2019) or treatment in general pooled across multiple treatment types (e.g., Buckman et al., 2021a; Buckman et al., 2022; Buckman et al., 2021b; Buckman et al., 2021c), which are referred to collectively as *prognostic* predictors. Other reviews have examined baseline variables that interact significantly with treatment type to predict outcomes (Maj et al., 2020; Perlman et al., 2019; Perna, Alciati, Daccò, Grassi, &

Caldirola, 2020), which are referred to as *prescriptive* predictors. Predictors from all important domains of either prescriptive or prognostic predictors were included in our baseline questionnaire or abstracted from EHRs or government small-area geospatial databases linked to patient residential addresses. Included here were six domains involving the episodes (symptom frequency, severity, subtypes, clinical staging, psychiatric comorbidities, functioning and quality of life), two others involving stressors (early environmental exposures, recent environmental stressors), and three involving personality/cognition (personality scales, neurocognition, dysfunctional cognitive schemas). A separate domain of "protective/resilience factors" assessed patient psychological characteristics (e.g., coping styles, self-reported psychological resilience) and environmental resources (e.g., access to supportive social relationships; access to material resources). Two other domains included information about comorbid physical disorders and family history of psychopathology.

We also included information about socio-demographics and treatment characteristics associated in previous research with differential depression treatment response (Constantino, Vîslă, Coyne, & Boswell, 2018; Kraus, Kadriu, Lanzenberger, Zarate Jr, & Kasper, 2019). Treatment characteristics included self-reports about current expectations, which were assessed as of the time of baseline rather than asking patients to recall their expectations prior to making their initial treatment contact. This time frame is relevant because, as noted above, the baseline assessment was carried out only after treatment started. Treatment characteristics also included patient self-reports about past treatment experiences and EHR data on treatment histories and ADM types prescribed in the current treatment. The latter were classified as norepinephrine-dopamine reuptake inhibitors (NDRI), serotonin antagonist reuptake inhibitors (SARI), serotonin modulator and stimulators (SMS), serotonin-norepinephrine reuptake

inhibitors (SNRI), selective serotonin reuptake inhibitors (SSRI), tricyclic antidepressants (TCA), and tetracyclic antidepressants (TeCA). We also included a dummy variable for ADMs suggested as most effective in controlled trials (i.e., escitalopram, mirtazapine, paroxetine, sertraline, venlafaxine) (Cipriani et al., 2018; Kazdin et al., 2021; Little, 2009). Other dummy variables were included for typical combinations of ADMs with baseline symptoms (e.g., trazodone with sleep disturbance, duloxetine with severe physical pain). We also recorded whether the treatment provider was the patient's regular primary care physician, someone else in the same primary care office, or someone at a mental health specialty clinic.

Categorical variables were coded as dummy indicators. Quantitative variables were standardized to a mean of 0 and variance of 1 and discretized into quintiles to create stabilized predictors and nested dichotomies. These transformations resulted in 2,768 potential predictors (Supplementary Tables 2-4). Item-level missingness was handled by single imputation carried out in the total sample before defining separate training and test samples, with missing values imputed to the mode for dichotomous and categorical variables and to the mean for ordinal and interval variables.

Analysis methods

The R program *sbw* (Zubizarreta, Li, Allouah, & Greifer, 2021) was used to make weighting adjustments for: (i) discrepancies in baseline EHR variables between eligible VHA patients and the 809 baseline respondents and (ii) discrepancies in baseline survey variables between the 660 3-month follow-up respondents and nonrespondents (Zubizarreta, 2015).

The Super Learner (SL) stacked generalization ML method (Polley, LeDell, Kennedy, Lendle, & van der Laan, 2021) was used to develop a prediction model in the weighted sample of 3-month respondents. SL generates predictions from a weighted combination of conventional

and flexible ML algorithms in an ensemble. Our SL specification used 10-fold cross-validation (10F-CV) to generate a weighted composite that performs at least as well in expectation as the best algorithm in the ensemble (Polley, Rose, & van der Laan, 2011). The appeal of stacked generalization over single algorithms is improved predictive accuracy by virtue of combining results across algorithms that include a wide range of functional forms (Polley et al., 2011). Consistent with recommendations (LeDell, van der Laan, & Petersen, 2016), a diverse set of algorithms was included in the SL ensemble (Supplementary Table 5). Some prior computational psychiatric studies have used similar stacked generalization procedures (Karrer et al., 2019; Ziobrowski et al., 2021a).

We estimated the SL model in a stratified (by the outcome variable) random 70% training sample n=462) and validated it in the remaining 30% test sample (n=198). Prediction strength, defined as area under the receiver operating characteristic curve [AUC (ROC)], was compared across a wide range of hyperparameter settings for each algorithm in the 10F-CV sample (Supplementary Table 5). Predictors were selected independently in each 10F-CV fold with a range of constraints on predictor number using lasso regression (Park & Casella, 2008) for linear models and Bayesian additive regression trees (BART) (Chipman, George, & McCulloch, 2010) for nonadditive models. Comparisons of AUC (ROC) estimated in the full training sample and 10F-CV allowed determination of how much each learner (i.e., combination of number of allowed predictors and hyperparameter values for a given algorithm) was overfitting and CV prediction strength. A subset of learners with balance between these two criteria was selected for the final SL ensemble. Once the final SL model was estimated, 10F-CV was used for model calibration in the 10F-CV sample based on isotonic regression (Lindhiem, Petersen, Mentch, & Youngstrom, 2020).

Models were assessed in the test sample by how well predicted probability of treatment response ranked patients on observed response (i.e., discrimination). The AUC (ROC) and the AUC of the precision recall curve [AUC (PRC)] were compared for the SL and a simpler benchmark lasso regression model whose penalty parameter was selected via internal cross-validation, both estimated in the training sample and applied to the test sample. Operating characteristics in the test sample were then inspected across quantiles of predicted probability of response. Operating characteristics included conditional and cumulative *sensitivity* (SN; the proportion of all patients responding to treatment who were in the quantile) and *positive predictive value* (PPV; the prevalence of treatment response in the decile). A locally estimated scatterplot smoothed calibration curve (Austin & Steyerberg, 2014) with .75 bandwidth was used to quantify model calibration in the test sample using the integrated calibration index (ICI) and expected calibration error (ECE) (Austin & Steyerberg, 2019). Model fairness (Yuan, Kumar, Ahmad, & Teredesai, 2021) was evaluated by examining variation in the association of predicted probability of response with observed response across socio-demographic subgroups related to health disparities (age, sex, race/ethnicity, education) using robust Poisson regression models (Zou, 2004).

Predictor importance was examined using the model-agnostic kernel Shapley Additive Explanations (SHAP) method (Lundberg & Lee, 2017), which generates a predicted difference in outcome score for each patient based on changing one and only one predictor at a time from its observed score to the mean across all logically possible permutations of other predictors. The mean of this "SHAP value" for a given predictor across all patients is 0. However, the mean *absolute* SHAP value provides useful information about the average importance of the predictor. A bee swarm plot of the association between the individual-level SHAP value and the observed

score for a given predictor was used to describe dominant direction of association. Mean absolute SHAP values can also be aggregated across subsets of predictors by summing SHAP values across the predictors at the individual level and then calculating the mean of the absolute value of this sum. Such aggregate scores estimate the expected change in prevalence of treatment response if all predictors for all patients changed from their observed values to the mean values.

SAS statistical software, version 9.4 (SAS Institute Inc, 2013), was used for data management, estimating prevalence of treatment response, and calculating SN, PPV, and AUC. R, version 4.0.5 (R Core Team, 2021) was used to estimate the SL model and SHAP values.

Reporting

We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guidelines (Collins, Reitsma, Altman, & Moons, 2015) in presenting results.

RESULTS

Sample characteristics and treatment response

Baseline QIDS-SR scores were transformed to approximate Hamilton Rating Scale of Depression (HRSD) categories using published transformation rules (Table 3 in Rush et al., 2003) to give a sense of the baseline symptom severity distribution. 30.1% of patients were classified as having baseline mild depression, 35.6% moderate, 21.4% severe, and 12.9% very severe (Supplementary Table 6). Given that the baseline assessment was not administered until after the initiation of treatment, there is a possibility that these severities were lower than if assessments had been carried out prior to beginning treatment. However, the Pearson correlation between the baseline QIDS-SR score and number of days between beginning treatment and

taking the baseline assessment (Median=21 days; inter-quartile range=14-30 days) was nonsignificant (r=.013, p=.74).

The great majority (80.7%) of patients were prescribed a single ADM, most commonly SSRIs (57.0%), SNRIs (16.8%), NDRIs (15.7%), and SARIs (15.0%). The modal socio-demographic categories were between 35-49 years of age, male, non-Hispanic White, married, and living in a Major Metropolitan Area. Except for age (p = 0.009), no statistically significant differences were found in baseline socio-demographics, clinical severity, or ADM classes between the weighted baseline sample and the doubly weighted analytic sample. 35.7% of doubly weighted test sample patients responded to treatment as of the 3-month assessment.

Model performance

The SL test sample AUC (ROC) was 0.66 compared to 0.62 for the benchmark lasso model (Figure 1). SL had better SN than lasso for all values of specificity (SP) above 0.25. SL had much higher PPV than the lasso model for SN below 0.10 and somewhat higher PPV than lasso across most of the remaining SN range (Figure 2). SL test sample AUC (ROC) remained 0.66 when the analysis was limited to patients classified as having at least moderate baseline symptom severity compared to AUC (ROC) of 0.60 among patients classified as having mild baseline symptom severity.

(Figures 1 and 2 about here)

Calibration based on the isotonic regression transformation was (ICI=0.28 and ECE=0.34). The SL model also had comparable fairness across subgroups defined by age, sex, race/ethnicity, and education (Supplementary Table 7).

(Table 1 about here)

A monotonic gradient was found in the proportion of test sample patients that responded to treatment [i.e., PPV (SE)] across SL model quantiles defined in the training sample. These quantiles could be collapsed without meaningful loss of information into three groups of patients (Table 1). Among patients in the first group, those with high predicted probabilities of response, 45.7% (5.5) responded to treatment. In the intermediate predicted probably group, 34.5% (7.6) responded. In the low predicted probability group, 11.1% (4.9) responded. These predicted probabilities did not vary significantly between patients whose baseline symptom severity was mild versus more severe (Table 1). Although the thresholds to define these groups were for quintiles in the training sample, 50.4% of patients in the test sample fell into the high group, 30.1% the intermediate group, and 19.4% the low group.

Predictor importance

The 2,768 predictors were highly redundant, as indicated by 750 (27.1%) of them having significant univariable associations with the outcome in the training sample at the .05 level but only 53 (1.9%) being selected by SL (Figure 3). Forty-six of these 53 were patient self-reports, 4 EHR variables, and 3 geospatial variables. The aggregate mean absolute SHAP value across all these predictors was 4.3%. This means the probability of treatment response would have changed by an estimated average of 4.3% if each patient's scores on all selected predictors changed from observed to sample-wide mean values.

(Figure 3 about here)

The most important predictors were features of the episode (10 of 53 predictors), with an aggregate mean absolute SHAP value of 3.5% (81% of the total). This included the most important predictor, overall depressive symptom frequency in the 2 weeks before treatment, in addition to two other important symptom measures, frequency of being happy or at peace (3rd

most important) and anhedonia (reverse coded, 14th most important), along with five indicators of current or recent comorbidity (4th, 15th, 22nd, 34th, 44th). These predictors were for the most part associated with reduced probability of treatment response. However, SHAP value distributions (Figure 4) show that some associations were nonmonotonic. For example, lower-than-average but not lowest overall depressive symptom frequency was associated with highest probability of treatment response.

(Figure 4 about here)

The 2nd most important predictor domain involved treatment characteristics (22 of 53 selected predictors), with an aggregate mean absolute SHAP value of 1.2% (25% of the total). Included were 10 indicators of positive treatment expectation/preference (e.g., 8th most important, expectation of having a good relationship with treatment provider), all positively associated with treatment response. Another 7 treatment-related predictors involved current treatment (e.g., 6th most important, referral to a mental health specialist or psychologist carried out intake). None of the ADM types was among the important predictors. The remaining treatment-related predictors involved treatment history (e.g., 12th most important, past psychotherapy was not helpful), all positively associated with treatment response.

There were only two other important predictor domains: recent stressors and protective/resilience factors, with aggregate mean absolute SHAP values of 0.9% (26% of the total) and 0.7% (17% of the total), respectively. The most important stressors were financial (2nd) and high mortality rate due to drug overdose in the patient's county of residence (5th), both negatively associated with treatment response. The protective/resilience factors included 3 indicators of psychological resilience and 3 of social support. As shown in the bee swarm plot, most of these predictors had nonmonotonic associations with the outcome due to patients with

higher-than-average but not highest reported protective/resilience scores having highest probabilities of treatment response.

DISCUSSION

The 35.7% 3-month ADM treatment response rate is comparable to previous VHA studies (Katz, Liebmann, Resnick, & Hoff, 2021) but lower than most civilian studies (Cuijpers et al., 2020), presumably reflecting the greater severity/complexity of depressed Veterans than civilians (Ziobrowski et al., 2021b). This highlights the potential importance of patients in the group with highest predicted probability of ADM response being more than four times as likely to respond as patients in the lowest group (45.7% versus 11.1%). Accurate discrimination of this sort is valuable as a first step in determining optimal treatments. However, multiple treatment-specific models need to be developed and combined to create a precision treatment rule for optimal assignment of patients across interventions (Kessler & Luedtke, 2021). For instance, psychotherapy or ADM plus psychotherapy might be prioritized for patients with low predicted probabilities of ADM treatment response, but only in the subset of patients with higher predicted probabilities of response in treatment-specific models for psychotherapy or combined therapy.

In this respect, our model might be compared to pharmacogenomic models used to determine pre-emptively whether ADMs are likely to be effective for individual patients (Greden et al., 2019). Our model performed at least as well as these pharmacogenomic models and could be implemented at a fraction of the cost of pharmacogenetic testing. The largest pharmacogenomic testing trial to date for ADM selection found that patients receiving test-congruent medications had a 12% higher probability of treatment response than patients receiving test-incongruent medications (29% versus 17%; Greden et al., 2019), whereas the

differences we found were 34.6% (45.7% versus 11.1%) between our high and low groups, 11.2% (45.7% versus 34.5%) between our top and intermediate groups, and 23.4% (34.5% versus 11.1%) between our intermediate and low groups.

Caution is needed in interpreting results regarding predictor importance because predictor importance rankings can be very unstable when, as in our dataset, many predictors are highly correlated (Leeuwenberg et al., 2022). Several broad results about predictor importance are nonetheless noteworthy. The most striking is that baseline clinical characteristics of the episode were by far the most important predictors. This is consistent with a recent individual-level meta-analysis of over 6,000 patients in primary care depression treatment across 12 trials, where baseline depression symptom severity was by far the single most important predictor of treatment response independent of treatment type (Buckman et al., 2021b). Other significant clinical predictors in that recent meta-analysis included duration of the depressive episode before beginning treatment, comorbid panic, and duration of comorbid anxiety. We found a different set of important clinical predictors, including two secondary depressive symptom factors and several measures of psychiatric comorbidity, but, as in the meta-analysis, these were all much less important than overall baseline depression symptom frequency.

The secondary symptom factors (absence of positive emotions, anhedonia) are both central aspects of melancholic depression. Evidence in previous studies has been mixed for melancholic depression being less responsive to treatment than others (Maj et al., 2020). It is noteworthy that the baseline assessment included the other indicators of melancholia (i.e., deep feelings of despair, mood worse in the morning, early morning awakening, psychomotor changes, weight loss, excessive guilt), but we did not attempt to define this or any other theoretical (Benazzi, 2006) or data-driven (Buckman et al., 2021a) MDD subtype beyond those

that emerged in an exploratory factor analysis of symptoms in the baseline assessment. The nonadditive models in the SL ensemble would have been expected to detect interactions across these factors if a strong data-driven episode subtype existed. Nonetheless, it might be useful in future investigations to use unsupervised ML methods to explore the possibility of detecting such clusters.

The importance of treatment characteristics, the next most important predictor domain in our sample, was striking in two ways. First, ADM type was unrelated to treatment response. Second, multiple aspects of treatment history and current treatment expectations were important. Although the literature on treatment expectations is inconsistent in its measures and controls for prior experiences, our finding that both process and outcome expectations were important predictors is broadly consistent with previous studies (Laferton, Kube, Salzmann, Auer, & Shedden-Mora, 2017). This is striking given that we controlled for and found significant associations of several measures of past treatment experiences that presumably underlie expectations. Taken together, these results argue for the potential value of shared decision making and patient-centered care for depression (Rush & Thase, 2018), for the potential value of expanding interventions to influence treatment expectations (Gruszka, Burger, & Jensen, 2019) and for the importance of including psychometrically sound and conceptually cohesive questions about treatment expectations and past treatment experiences in baseline patient assessments (e.g., Barth, Kern, Lüthi, & Witt, 2019).

The finding that recent stressors were important is broadly consistent with evidence documenting effects of stressful life experiences on depression treatment response (Buckman et al., 2022). The fact that financial stress was the 2nd most important predictor was especially striking and is consistent with prior studies showing that unemployment and low household

income are top predictors of low ADM treatment response (Lee et al., 2018). The findings that baseline protective/resilience factors were important is also in line with much previous research (Buckman et al., 2021c; Laird, Lavretsky, St Cyr, & Siddarth, 2018). The fact that some of these associations were nonmonotonic is consistent with naturalistic evidence that moderate, compared to extremely low or high, levels of emotional reactivity to stress predict low future depression severity (Santee & Starr, 2021) and that baseline self-reported resilience is sometimes significant in predicting depression treatment response only in interaction with other predictors (Choi et al., 2021; Min, Lee, Lee, Lee, & Chae, 2012). These specifications might reflect the greater importance of protective/resilience factors in the subset of patients whose depressive episodes are triggered by stressful life experiences, which could be the subject of future investigation (Chromik, 2021).

Limitations: The study had several noteworthy limitations. Three of these involve external validity. First, the baseline response rate was low, although comparable to response rates in other VHA studies examining mental health outcomes (King, Beehler, Buchholz, Johnson, & Wray, 2019; Stolzmann et al., 2019). However, as shown in a previous report (Puac-Polanco et al., 2021), there are minimal differences between our responders and non-responders on baseline administrative variables and equally modest differences in baseline self-reports between patients followed versus lost to follow-up, although response bias might nonetheless exist with respect to unmeasured variables. Second, we did not account for possible disruptions in care due to the COVID-19 pandemic, which involved 7.6% of study patients who completed assessments after February 2020. Third, although the model was validated in a separate test sample, it was not tested in an external validation sample. Nor is it clear whether findings would generalize to non-VHA patients.

A separate set of limitations involve design decisions that could have biased results. One of these is that study recruitment and assessment occurred only after the initial visit, during which time symptoms might have decreased, leading to distortion in our estimates of associations between baseline symptoms and treatment response. As reported above, the association between time between initiating treatment and completing the baseline assessment was unrelated to baseline QIDS-SR scores, somewhat reducing this concern, but it is nonetheless important that future replications and extensions of our work are carried out with baseline assessment administered before treatment selection is made. Another limitation that might have biased results was the use of a very large set of predictors, which could have resulted in over-fitting even though we used procedures to minimize this possibility.

A final set of noteworthy limitations involves the measures. The predictors excluded information about military experiences that might have led to the depression. And the outcomes were based on self-reports rather than clinical interviews.

Strengths: The study also had several strengths, including an observational sample with greater external validity than clinical trial samples, a rich baseline predictor set that included a wide range of variables found in previous research to be prognostic predictors of depression treatment response, and use of a rigorous ML method to develop the model.

CONCLUSIONS

Within the context of these limitations, we found that a model to predict ADM treatment response could be developed based largely on a battery of self-report questions along with some administrative variables from EHRs and geospatial databases. The model had modest overall prediction strength but nonetheless provided enough discrimination across three broad groups of

patients to have potential value in informing depressed patients pre-emptively about their likelihood of responding to ADM as part of a patient-centered shared decision-making process. The model had good calibration and fairness with respect to key indicators of health disparities. Our findings would need to be replicated in a sample where the baseline assessment occurred before the beginning of treatment, the model streamlined, and parallel models built for predicted response to other types of treatment before results could be useful. In addition, parallel models combined across different treatments would be needed to determine best treatment options for particular patients (Kessler & Luedtke, 2021).

ACKNOWLEDGMENTS

Conflict of Interest:

In the past 3 years, Dr. Kessler was a consultant for Datastat, Inc., Holmusk, RallyPoint Networks, Inc., and Sage Therapeutics. He has stock options in Mirah, PYM, and Roga Sciences. Dr. Pigeon consulted for CurAegis Technologies and received clinical trial support from Pfizer, Inc. and Abbvie, Inc. Dr. Zubizarreta consulted for Johnson & Johnson Real World Data Analytics. Dr. Cipriani is supported by the National Institute for Health Research (NIHR) Oxford Cognitive Health Clinical Research Facility, by an NIHR Research Professorship (grant RP-2017-08-ST2-006), by the NIHR Oxford and Thames Valley Applied Research Collaboration and by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005); he has also received research, educational and consultancy fees from INCiPiT (Italian Network for Paediatric Trials), CARIPLO Foundation and Angelini Pharma. The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the UK Department of Health. The remaining authors report no conflict of interest.

Ethical Standards:

All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

REFERENCES

Austin, P. C., & Steyerberg, E. W. (2014) Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Statistics in Medicine, 33*(3), 517-535. doi:10.1002/sim.5941

Austin, P. C., & Steyerberg, E. W. (2019) The Integrated Calibration Index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Statistics in Medicine, 38*(21), 4051-4065. doi:10.1002/sim.8281

Barth, J., Kern, A., Lüthi, S., & Witt, C. M. (2019) Assessment of patients' expectations: development and validation of the Expectation for Treatment Scale (ETS). *BMJ Open, 9*(6), e026712. doi:10.1136/bmjopen-2018-026712

Benazzi, F. (2006) Various forms of depression. *Dialogues in Clinical Neuroscience, 8*(2), 151-161. doi:10.31887/DCNS.2006.8.2/fbenazzi

Buckman, J. E. J., Cohen, Z. D., O'Driscoll, C., Fried, E. I., Saunders, R., Ambler, G., . . . Pilling, S. (2021a) Predicting prognosis for adults with depression using individual symptom data: a comparison of modelling approaches. *Psychological Medicine*, Advance online publication. doi:10.1017/s0033291721001616

Buckman, J. E. J., Saunders, R., Arundell, L. L., Oshinowo, I. D., Cohen, Z. D., O'Driscoll, C., . . . Pilling, S. (2022) Life events and treatment prognosis for depression: A systematic review and individual patient data meta-analysis. *Journal of Affective Disorders, 299*, 298-308. doi:10.1016/j.jad.2021.12.030

Buckman, J. E. J., Saunders, R., Cohen, Z. D., Barnett, P., Clarke, K., Ambler, G., . . . Pilling, S. (2021b) The contribution of depressive 'disorder characteristics' to determinations of

prognosis for adults with depression: an individual patient data meta-analysis. *Psychological Medicine, 51*(7), 1068-1081. doi:10.1017/s0033291721001367

Buckman, J. E. J., Saunders, R., O'Driscoll, C., Cohen, Z. D., Stott, J., Ambler, G., . . . Pilling, S. (2021c) Is social support pre-treatment associated with prognosis for adults with depression in primary care? *Acta Psychiatrica Scandinavica, 143*(5), 392-405. doi:10.1111/acps.13285

Chekroud, A. M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., . . . Choi, K. (2021) The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*, *20*(2), 154-170. doi:10.1002/wps.20882

Chipman, H. A., George, E. I., & McCulloch, R. E. (2010) BART: Bayesian additive regression trees. *The Annals of Applied Statistics, 4*(1), 266-298. doi:10.1214/09-AOAS285

Choi, W., Kim, J. W., Kang, H. J., Kim, H. K., Kang, H. C., Lee, J. Y., . . . Kim, J. M. (2021) Synergistic effects of resilience and serum ghrelin levels on the 12-week pharmacotherapeutic response in patients with depressive disorders. *Journal of Affective Disorders, 295*, 1489-1493. doi:10.1016/j.jad.2021.09.039

Chromik, M. (2021) Making SHAP Rap: Bridging local and global insights through interaction and narratives. In *Human-Computer Interaction – INTERACT 2021* (ed. C Ardito, R Lanzilotti, A Malizia, H Petrie, A Piccinno, G Desolda and K Inkpen), pp. 641-651. Springer: Cham.

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., . . . Geddes, J. R. (2018) Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and

network meta-analysis. *Lancet (London, England), 391*(10128), 1357-1366. doi:10.1016/s0140-6736(17)32802-7

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015) Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD Statement. *The British Journal of Surgery, 102*(3), 148-158. doi:10.1002/bjs.9736

Constantino, M. J., Vîslă, A., Coyne, A. E., & Boswell, J. F. (2018) A meta-analysis of the association between patients' early treatment outcome expectation and their posttreatment outcomes. *Psychotherapy (Chicago, Ill.), 55*(4), 473-485. doi:10.1037/pst0000169

Cuijpers, P., Noma, H., Karyotaki, E., Vinkers, C. H., Cipriani, A., & Furukawa, T. A. (2020) A network meta-analysis of the effects of psychotherapies, pharmacotherapies and their combination in the treatment of adult depression. *World Psychiatry*, *19*(1), 92-107. doi:10.1002/wps.20701

Day, E., Shah, R., Taylor, R. W., Marwood, L., Nortey, K., Harvey, J., . . . Strawbridge, R. (2021) A retrospective examination of care pathways in individuals with treatment-resistant depression. *British Journal of Psychiatry Open, 7*(3), e101-e101. doi:10.1192/bjo.2021.59

Ermers, N. J., Hagoort, K., & Scheepers, F. E. (2020) The Predictive Validity of Machine Learning Models in the Classification and Treatment of Major Depressive Disorder: State of the Art and Future Directions. *Frontiers in Psychiatry, 11*, 472. doi:10.3389/fpsyt.2020.00472

Furukawa, T. A., Suganuma, A., Ostinelli, E. G., Andersson, G., Beevers, C. G., Shumake, J., . . . Cuijpers, P. (2021) Dismantling, optimising, and personalising internet cognitive

behavioural therapy for depression: a systematic review and component network meta-analysis using individual participant data. *The Lancet Psychiatry, 8*(6), 500-511. doi:10.1016/s2215-0366(21)00077-8

GBD 2019 Diseases and Injuries Collaborators (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: A systematic analysis for the Global Burden of Disease Study 2019. *The Lancet, 396*(10258), 1204-1222. doi:10.1016/s0140-6736(20)30925-9

Greden, J. F., Parikh, S. V., Rothschild, A. J., Thase, M. E., Dunlop, B. W., DeBattista, C., . . . Dechairo, B. (2019) Impact of pharmacogenomics on clinical outcomes in major depressive disorder in the GUIDED trial: A large, patient- and rater-blinded, randomized, controlled study. *Journal of Psychiatric Research, 111*, 59-67. doi:10.1016/j.jpsychires.2019.01.003

Gruszka, P., Burger, C., & Jensen, M. P. (2019) Optimizing Expectations via Mobile Apps: A New Approach for Examining and Enhancing Placebo Effects. *Frontiers in Psychiatry, 10*, 365. doi:10.3389/fpsyt.2019.00365

Hockenberry, J. M., Joski, P., Yarbrough, C., & Druss, B. G. (2019) Trends in Treatment and Spending for Patients Receiving Outpatient Treatment of Depression in the United States, 1998-2015. *JAMA Psychiatry, 76*(8), 810-817. doi:10.1001/jamapsychiatry.2019.0633

Huang, J., Wang, Y., Chen, J., Zhang, Y., Yuan, Z., Yue, L., . . . Fang, Y. (2018) Clinical outcomes of patients with major depressive disorder treated with either duloxetine, escitalopram, fluoxetine, paroxetine, or sertraline. *Neuropsychiatric Disease and Treatment, 14*, 2473-2484. doi:10.2147/ndt.s159800

Karrer, T. M., Bassett, D. S., Derntl, B., Gruber, O., Aleman, A., Jardri, R., . . . Bzdok, D. (2019) Brain-based ranking of cognitive domains to predict schizophrenia. *Human Brain Mapping, 40*(15), 4487-4507. doi:10.1002/hbm.24716

Katz, I. R., Liebmann, E. P., Resnick, S. G., & Hoff, R. A. (2021) Performance of the PHQ-9 across conditions and comorbidities: Findings from the Veterans Outcome Assessment survey. *Journal of Affective Disorders, 294*, 864-867. doi:10.1016/j.jad.2021.07.108

Kazdin, A. E., Wu, C.-S., Hwang, I., Puac-Polanco, V., Sampson, N. A., Al-Hamzawi, A., . . . Kessler, R. C. (2021) Antidepressant use in low- middle- and high-income countries: a World Mental Health Surveys report. *Psychological Medicine*, Advance online publication. doi:10.1017/S0033291721003160

Kessler, R. C., & Luedtke, A. (2021) Pragmatic Precision Psychiatry-A New Direction for Optimizing Treatment Selection. *JAMA Psychiatry, 78*(12), 1384-1390. doi:10.1001/jamapsychiatry.2021.2500

King, P. R., Beehler, G. P., Buchholz, L. J., Johnson, E. M., & Wray, L. O. (2019) Functional concerns and treatment priorities among veterans receiving VHA Primary Care Behavioral Health services. *Families, Systems & Health*, *37*(1), 68-73. doi:10.1037/fsh0000393

Kraus, C., Kadriu, B., Lanzenberger, R., Zarate Jr, C. A., & Kasper, S. (2019) Prognosis and improved outcomes in major depression: a review. *Translational Psychiatry, 9*(1), 127. doi:10.1038/s41398-019-0460-3

Laferton, J. A. C., Kube, T., Salzmann, S., Auer, C. J., & Shedden-Mora, M. C. (2017) Patients' Expectations Regarding Medical Treatment: A Critical Review of Concepts and Their Assessment. *Frontiers in Psychology, 8*, 233-233. doi:10.3389/fpsyg.2017.00233

Laird, K. T., Lavretsky, H., St Cyr, N., & Siddarth, P. (2018) Resilience predicts remission in antidepressant treatment of geriatric depression. *International Journal of Geriatric Psychiatry, 33*(12), 1596-1603. doi:10.1002/gps.4953

Larson, S., Nemoianu, A., Lawrence, D. F., Troup, M. A., Gionfriddo, M. R., Pousti, B., . . . Touya, M. (2021) Characterizing primary care for patients with major depressive disorder using electronic health records of a US-based healthcare provider. *Journal of Affective Disorders, 300*, 377-384. doi:10.1016/j.jad.2021.12.096

LeDell, E., van der Laan, M. J., & Petersen, M. (2016) AUC-Maximizing Ensembles through Metalearning. *The International Journal of Biostatistics, 12*(1), 203-218. doi:10.1515/ijb-2015-0035

Lee, Y., Ragguett, R. M., Mansur, R. B., Boutilier, J. J., Rosenblat, J. D., Trevizol, A., . . . McIntyre, R. S. (2018) Applications of machine learning algorithms to predict therapeutic outcomes in depression: A meta-analysis and systematic review. *Journal of Affective Disorders, 241*, 519-532. doi:10.1016/j.jad.2018.08.073

Leeuwenberg, A. M., van Smeden, M., Langendijk, J. A., van der Schaaf, A., Mauer, M. E., Moons, K. G. M., . . . Schuit, E. (2022) Performance of binary prediction models in high-correlation low-dimensional settings: a comparison of methods. *Diagnostic and Prognostic Research, 6*(1), 1. doi:10.1186/s41512-021-00115-5

Leon, A. C., Olfson, M., Portera, L., Farber, L., & Sheehan, D. V. (1997) Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *International Journal of Psychiatry in Medicine, 27*(2), 93-105. doi:10.2190/t8em-c8yh-373n-1uwd

Lindhiem, O., Petersen, I. T., Mentch, L. K., & Youngstrom, E. A. (2020) The Importance of Calibration in Clinical Psychology. *Assessment, 27*(4), 840-854. doi:10.1177/1073191117752055

Little, A. (2009) Treatment-resistant depression. *American Family Physician, 80*(2), 167-172.

Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. Retrieved from https://proceedings.neurips.cc//paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf.

Maj, M., Stein, D. J., Parker, G., Zimmerman, M., Fava, G. A., De Hert, M., . . . Wittchen, H. U. (2020) The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry*, *19*(3), 269-293. doi:10.1002/wps.20771

McHugh, R. K., Whitton, S. W., Peckham, A. D., Welge, J. A., & Otto, M. W. (2013) Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. *The Journal of Clinical Psychiatry, 74*(6), 595-602. doi:10.4088/JCP.12r07757

Min, J. A., Lee, N. B., Lee, C. U., Lee, C., & Chae, J. H. (2012) Low trait anxiety, high resilience, and their interaction as possible predictors for treatment response in patients with depression. *Journal of Affective Disorders, 137*(1-3), 61-69. doi:10.1016/j.jad.2011.12.026

Noma, H., Furukawa, T. A., Maruo, K., Imai, H., Shinohara, K., Tanaka, S., . . . Cipriani, A. (2019) Exploratory analyses of effect modifiers in the antidepressant treatment of major depression: Individual-participant data meta-analysis of 2803 participants in seven

placebo-controlled randomized trials. *Journal of Affective Disorders, 250*, 419-424. doi:10.1016/j.jad.2019.03.031

Park, T., & Casella, G. (2008) The Bayesian Lasso. *Journal of the American Statistical Association, 103*(482), 681-686. doi:10.1198/016214508000000337

Perlman, K., Benrimoh, D., Israel, S., Rollins, C., Brown, E., Tunteng, J. F., . . . Berlim, M. T. (2019) A systematic meta-review of predictors of antidepressant treatment outcome in major depressive disorder. *Journal of Affective Disorders, 243*, 503-515. doi:10.1016/j.jad.2018.09.067

Perna, G., Alciati, A., Daccò, S., Grassi, M., & Caldirola, D. (2020) Personalized Psychiatry and Depression: The Role of Sociodemographic and Clinical Variables. *Psychiatry Investigation, 17*(3), 193-206. doi:10.30773/pi.2019.0289

Polley, E., LeDell, E., Kennedy, C., Lendle, S., & van der Laan, M. J. (2021). Superlearner: Super learner prediction, version 2.0-28. Retrieved from https://CRAN.R-project.org/package=SuperLearner.

Polley, E. C., Rose, S., & van der Laan, M. J. (2011) Super learning. In *Targeted learning: Casual inference for observational and experimental data* (ed. MJ van der Laan and S Rose), pp. 43-66. Springer: New York.

Puac-Polanco, V., Leung, L. B., Bossarte, R. M., Bryant, C., Keusch, J. N., Liu, H., . . . Kessler, R. C. (2021) Treatment Differences in Primary and Specialty Settings in Veterans with Major Depression. *Journal of the American Board of Family Medicine*, *34*(2), 268-290. doi:10.3122/jabfm.2021.02.200475

Qaseem, A., Barry, M. J., & Kansagara, D. (2016) Nonpharmacologic Versus Pharmacologic Treatment of Adult Patients With Major Depressive Disorder: A Clinical Practice

Guideline From the American College of Physicians. *Annals of Internal Medicine, 164*(5), 350-359. doi:10.7326/m15-2570

R Core Team. (2021). R: A language and environment for statistical computing. Retrieved from https://www.R-project.org/.

Rush, A. J., & Thase, M. E. (2018) Improving Depression Outcome by Patient-Centered Medical Management. *American Journal of Psychiatry, 175*(12), 1187-1198. doi:10.1176/appi.ajp.2018.18040398

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., . . . Keller, M. B. (2003) The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573-583. doi:10.1016/s0006-3223(02)01866-8

Santee, A. C., & Starr, L. R. (2021) Examining Linear and Nonlinear Associations Between Negative Emotional Reactivity to Daily Events and Depression Among Adolescents. *Clinical Psychological Science*, Advance online publication. doi:10.1177/21677026211045684

SAS Institute Inc (2013) SAS ®Software 9.4 edn. Cary, NC.

Stolzmann, K., Meterko, M., Miller, C. J., Belanger, L., Seibert, M. N., & Bauer, M. S. (2019) Survey Response Rate and Quality in a Mental Health Clinic Population: Results from a Randomized Survey Comparison. *The Journal of Behavioral Health Services & Research, 46*(3), 521-532. doi:10.1007/s11414-018-9617-8

Wang, G., You, X., Wang, X., Xu, X., Bai, L., Xie, J., . . . Hu, C. (2018) Safety and effectiveness of escitalopram in an 8-week open study in Chinese patients with depression and anxiety. *Neuropsychiatric Disease and Treatment, 14*, 2087-2097. doi:10.2147/ndt.s164673

Yuan, M., Kumar, V., Ahmad, M. A., & Teredesai, A. (2021). Assessing Fairness in Classification Parity of Machine Learning Models in Healthcare. Retrieved from https://arxiv.org/abs/2102.03717.

Zilcha-Mano, S., Wang, X., Wajsbrot, D. B., Boucher, M., Fine, S. A., & Rutherford, B. R. (2021) Trajectories of Function and Symptom Change in Desvenlafaxine Clinical Trials: Toward Personalized Treatment for Depression. *Journal of Clinical Psychopharmacology, 41*(5), 579-584. doi:10.1097/jcp.0000000000001435

Ziobrowski, H. N., Kennedy, C. J., Ustun, B., House, S. L., Beaudoin, F. L., An, X., . . . van Rooij, S. J. H. (2021a) Development and Validation of a Model to Predict Posttraumatic Stress Disorder and Major Depression After a Motor Vehicle Collision. *JAMA Psychiatry, 78*(11), 1228-1237. doi:10.1001/jamapsychiatry.2021.2427

Ziobrowski, H. N., Leung, L. B., Bossarte, R. M., Bryant, C., Keusch, J. N., Liu, H., . . . Kessler, R. C. (2021b) Comorbid mental disorders, depression symptom severity, and role impairment among Veterans initiating depression treatment through the Veterans Health Administration. *Journal of Affective Disorders, 290*, 227-236. doi:10.1016/j.jad.2021.04.033

Zou, G. (2004) A Modified Poisson Regression Approach to Prospective Studies with Binary Data. *American Journal of Epidemiology, 159*(7), 702-706. doi:10.1093/aje/kwh090

Zubizarreta, J. R. (2015) Stable Weights that Balance Covariates for Estimation With Incomplete Outcome Data. *Journal of the American Statistical Association, 110*(511), 910-922. doi:10.1080/01621459.2015.1023805

Zubizarreta, J. R., Li, Y., Allouah, A., & Greifer, N. (2021). sbw: Stable balancing weights for causal inference and estimation with incomplete outcome data (Version 1.1.1). Retrieved from https://cran.rstudio.com/web/packages/sbw/

Figure Titles and Footnotes

Figure 1 Title:

Receiver Operating Characteristic (ROC) curve comparing Super Learner with benchmark lasso in the test sample

Figure 2 Title:

Precision Recall Curve (PRC) comparing Super Learner with benchmark lasso in the test sample

Figure 3 Title:

Predictor importance as determined by Shapley Additive Explanation (SHAP) values for the Super Learner Model in the test sample[a]

Figure 3 Footnotes:

Abbreviations: sx, symptoms; freq, frequency; (D), dummy variable; (S), stabilized variable; ADM, antidepressant medication; tx, treatment; (RS), reverse stabilized; MDE, major depressive episode; px, patient; PCP, primary care provider; PTSD, post-traumatic stress disorder; BMI, body mass index; HRSD, Hamilton Rating Scale of Depression.
[a]See Supplementary Table 8 for descriptions of the predictor labels.

Figure 4 Title:

Bee swarm plot of individual-level predictor-specific SHAP values for the most important predictors in the Super Learner model[a]

Figure 4 Footnotes:

Abbreviations: sx, symptoms; freq, frequency; (D), dummy variable; (S), stabilized variable; ADM, antidepressant medication; tx, treatment; (RS), reverse stabilized; MDE, major depressive episode; px, patient; PCP, primary care provider; PTSD, post-traumatic stress disorder; BMI, body mass index; HRSD, Hamilton Rating Scale of Depression.
[a]See Supplementary Table 8 for descriptions of the predictor labels.

Table 1. Prediction of 3-Month ADM treatment response in the test sample in three group defined by predicted probabilities in the training sample (n=198)

	Distribution		PPV		Cumulative PPV		Sensitivity		Cumulative Sensitivity	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	%	(SE)
High	50.4[a]	(4.1)	45.7[b]	(5.5)	45.7	(5.5)	64.7	(6.7)	64.7	(6.7)
Intermediate	30.1[c]	(3.9)	34.5[d]	(7.6)	41.5	(4.5)	29.2	(6.6)	93.9	(2.7)
Low	19.4	(3.3)	11.1	(4.9)	35.6	(3.9)	6.0	(2.7)	100.0	--

Abbreviations: ADM, antidepressant medication; PPV, positive predictive value (i.e., predicted proportion with treatment response); SE, standard error; SN, sensitivity (i.e., proportion of all treatment responders).

[a]80.9% among patients with mild baseline symptom severity versus 35.3% among other patients.

[b]47.5% among patients with mild baseline symptom severity versus 45.7% among other patients.

[c]13.7% in the subsample of patients with mild baseline symptom severity versus 38.3% among other patients.

[d]30.3% among patients with mild baseline symptom severity versus 38.4% among other patients.

Figure 1. Receiver Operating Characteristic (ROC) curve comparing Super Learner with benchmark lasso in the test sample

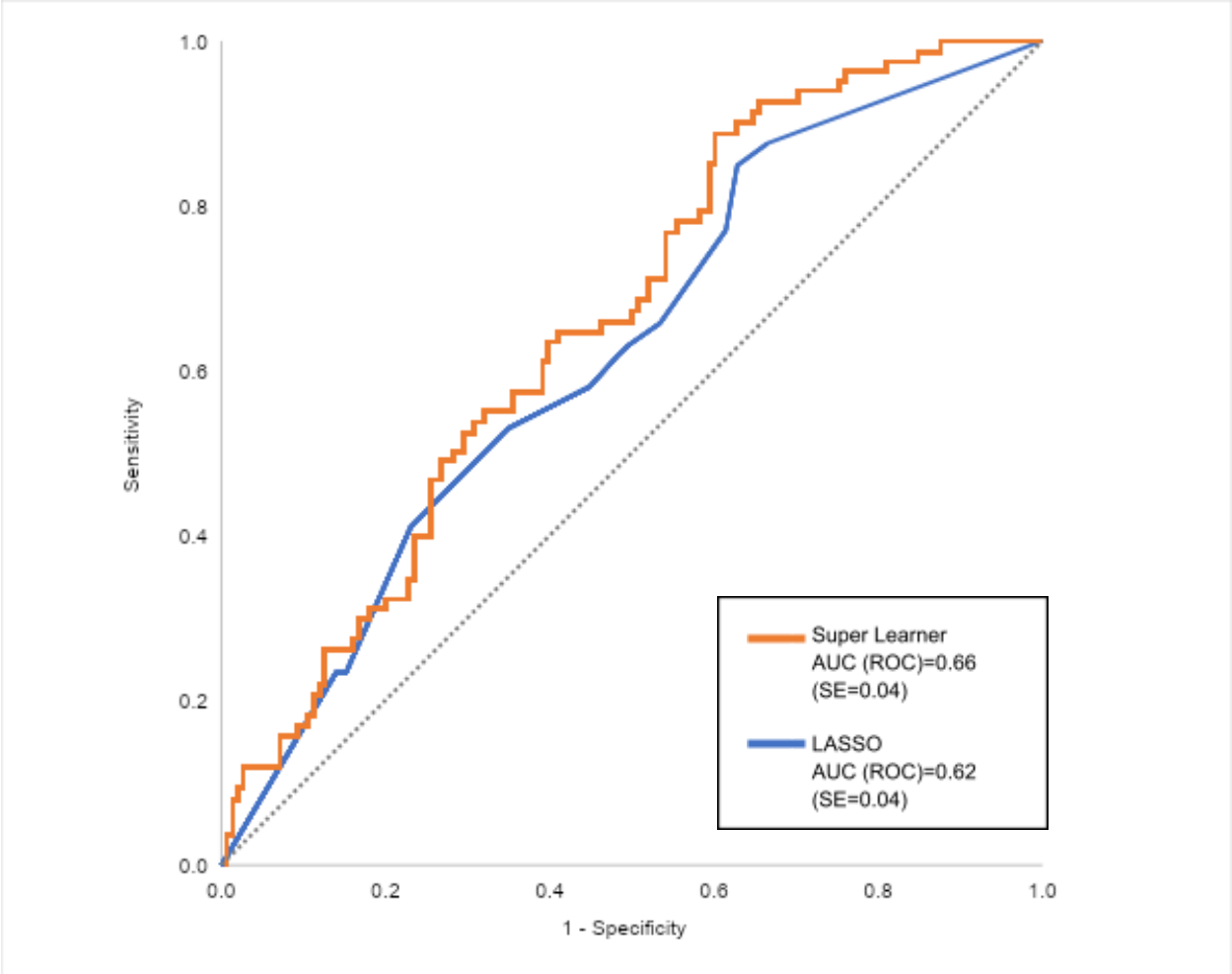


Figure 2. Precision Recall Curve (PRC) comparing Super Learner with benchmark lasso in the test sample

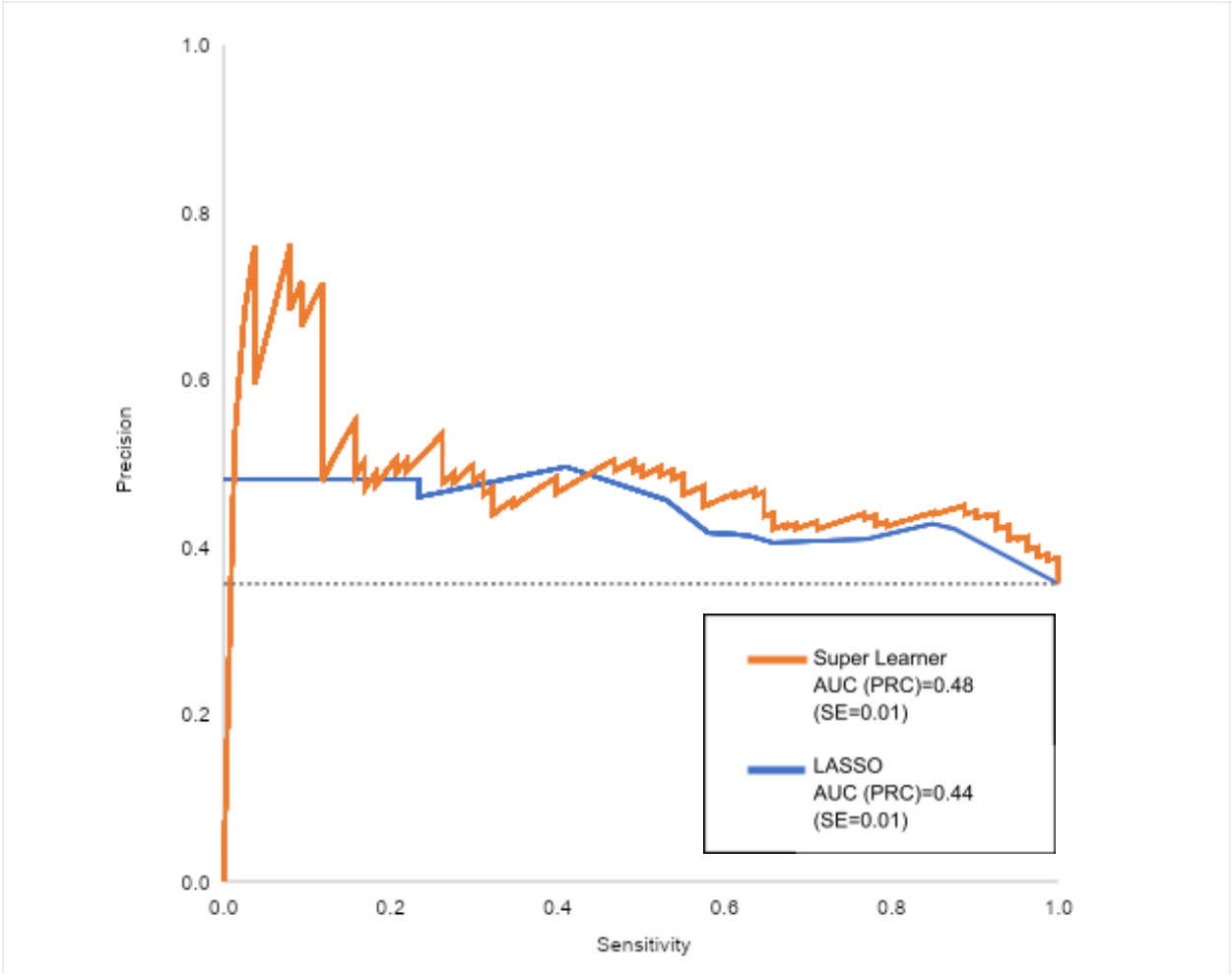
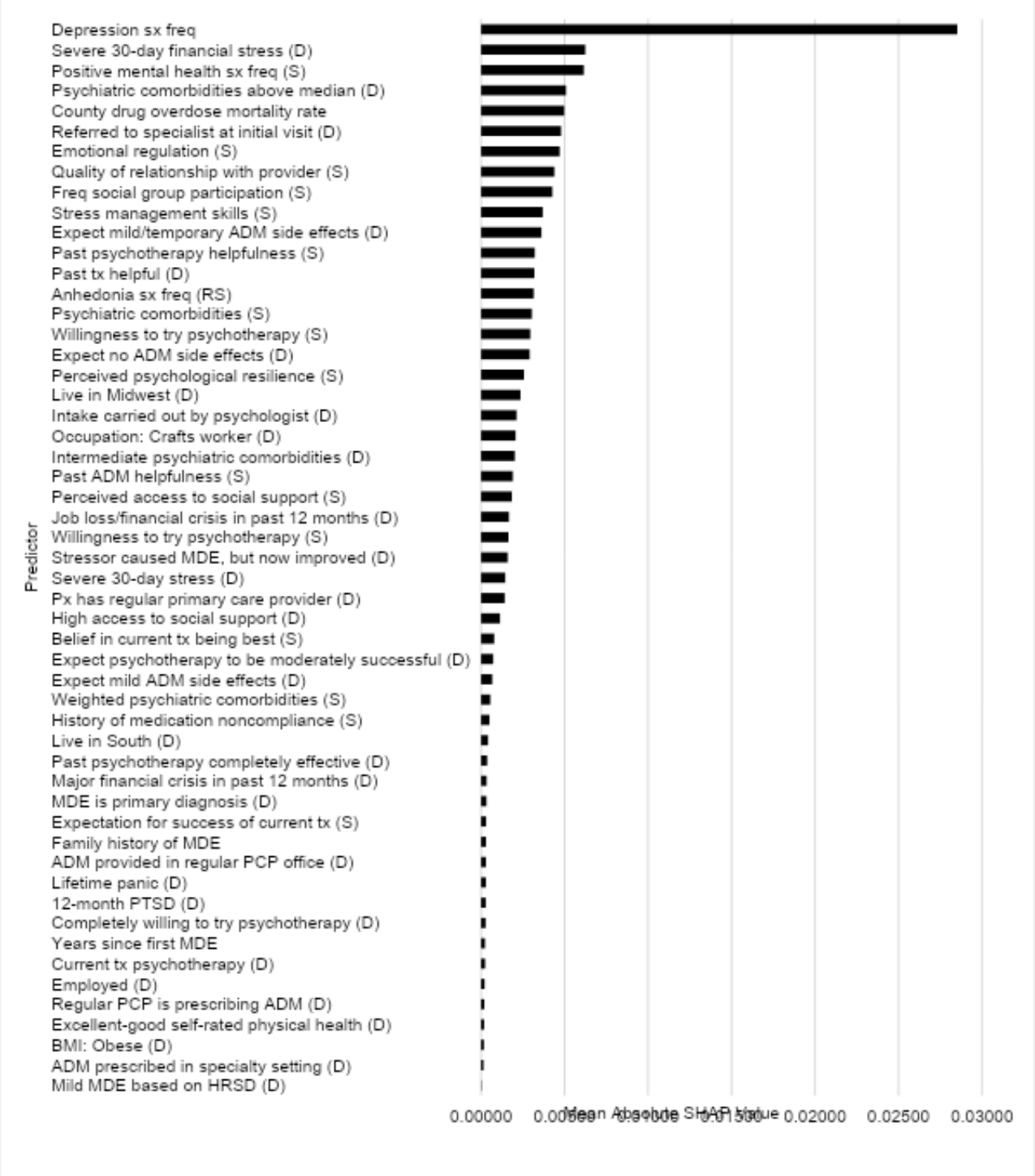


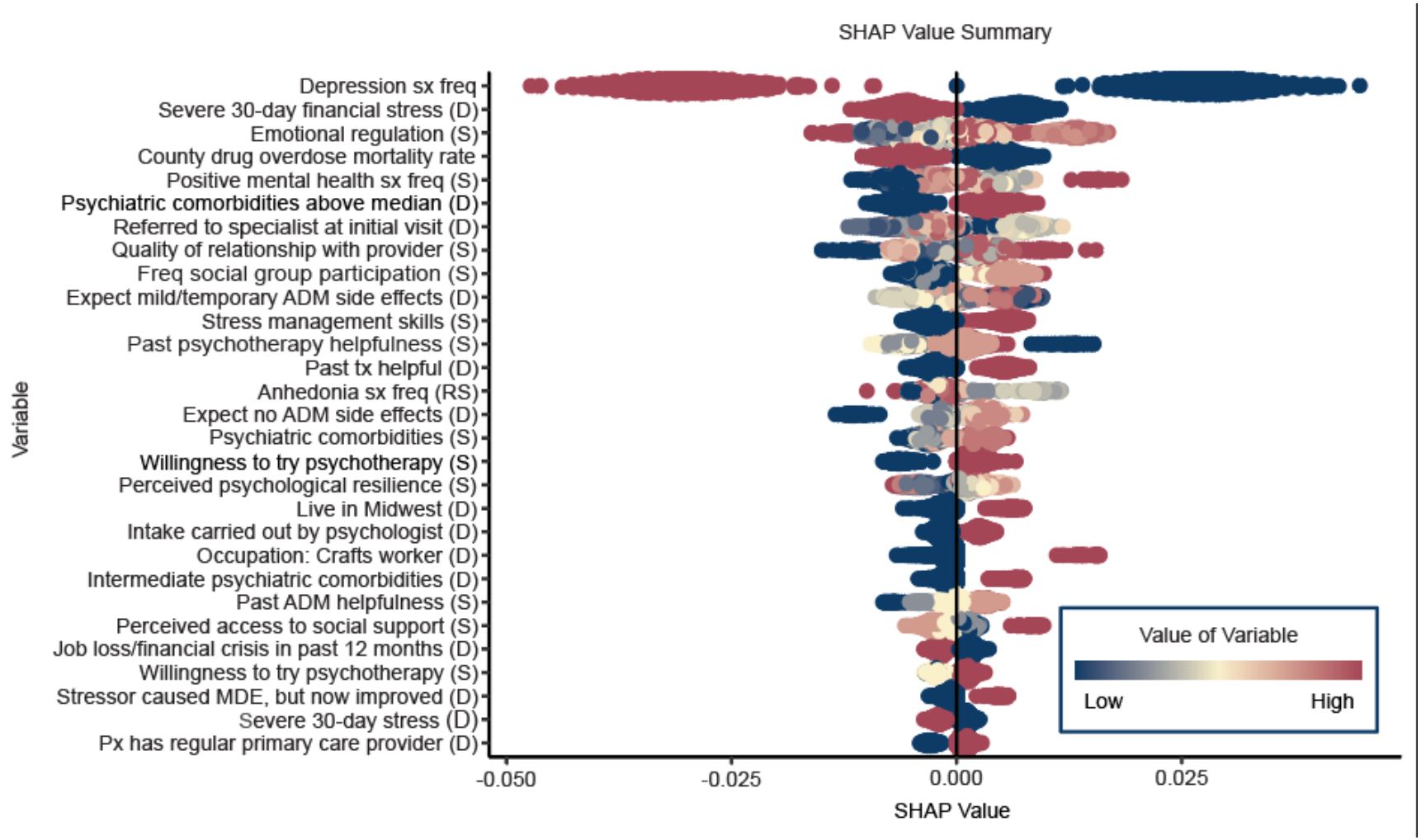
Figure 3. Predictor importance as determined by Shapley Additive Explanation (SHAP) values for the Super Learner Model in the test sample[a]



Abbreviations: sx, symptoms; freq, frequency; (D), dummy variable; (S), stabilized variable; ADM, antidepressant medication; tx, treatment; (RS), reverse stabilized; MDE, major depressive episode; px, patient; PCP, primary care provider; PTSD, post-traumatic stress disorder; BMI, body mass index; HRSD, Hamilton Rating Scale of Depression.

[a]See Supplementary Table 8 for descriptions of the predictor labels.

Figure 4. Bee swarm plot of individual-level predictor-specific SHAP values for the most important predictors in the Super Learner model[a]



Abbreviations: sx, symptoms; freq, frequency; (D), dummy variable; (S), stabilized variable; ADM, antidepressant medication; tx, treatment; (RS), reverse stabilized; MDE, major depressive episode; px, patient; PCP, primary care provider; PTSD, post-traumatic stress disorder; BMI, body mass index; HRSD, Hamilton Rating Scale of Depression.

[a]See Supplementary Table 8 for descriptions of the predictor labels.

Supplementary Table 1. ICD-9-CM and ICD-10-CM Mental and Behavioral Health Codes

Disorder	Codes
Major depression	ICD-9-CM codes: 296.2X, 296.3X, 300.4, 311 ICD-10-CM codes: F32.XX (excluding F32.81 and F32.89), F33.XX
Suicide attempt	ICD-9-CM codes: E950.X, E951.X, E952.X, E953.X, E954, E955.X, E956, E957.X ICD-10-CM codes: T14.91, T36.XX2, T37.XX2, T38.XX2, T39.XX2, T40.XX2, T41.XX2, T42.XX2, T43.XX2, T44.XX2, T45.XX, T46.XX2, T47.XX2, T48.XX2, T49.XX2, T50.XX2, T51.XX2, T52.XX2, T53.XX2, T54.XX2, T55.XX2, T56.XX2, T57.XX2, T58.XX2, T59.XX2, T60.XX2, T61.XX2, T62.XX2, T63.XX2, T64.XX2, T65.XX2, T71.1X2, T71.2X2
Bipolar disorder	ICD-9-CM codes: 296.0X, 296.1X, 296.4X, 296.5X, 296.6X, 296.7, 296.80, 296.81, 296.89, 301.13 ICD-10-CM codes: F30.X, F31.X, F34.0
Nonaffective psychosis	ICD-9-CM codes: 293.81, 293.82, 293.89, 295.XX, 297.X, 298.X, 301.22 ICD-10-CM codes: F06.0, F06.1, F20.XX, F21, F22, F23, F24, F25.X, F28, F29, F53
Dementia	ICD-9-CM codes: 290.XX, 294.1X, 294.8 ICD-10-CM codes: F01.XX, F01.XX, F03.XX
Intellectual disabilities	ICD-9-CM codes: 317, 318, 319 ICD-10-CM codes: F70, F71, F72, F73, F78, F79
Autism	ICD-9-CM code: 299.XX ICD-10-CM code: F84.0
Tourette's disorder	ICD-9-CM code: 307.23 ICD-10-CM code: F95.2
Stereotyped movement disorders	ICD-9-CM code: 307.3 ICD-10-CM code: F98.4
Borderline intellectual functioning	ICD-10-CM code: R41.83

Abbreviations. ICD-9-CM, International Classification of Diseases, Ninth Revision, Clinical Modification; ICD-10-CM, International Classification of Diseases, Tenth Revision, Clinical Modification.

Supplementary Table 2. Baseline self-report predictors[a] included in machine learning models

Risk factor domain

1. Symptom frequency scales

a. 2-week depression symptom frequency (Aish, Wasserman & Renberg, 2001; Akiskal et al., 2005; American Psychiatric Association, 2013; Kessler & Ustün, 2004; Llerena et al., 2013; Rizvi et al., 2015; Rush, Gullion, Basco Jarrett & Trivedi, 1996; Rush et al., 2003; Saffer, Lanting, Koehle, Klonsky & Iverson, 2015; Treynor, Gonzalez & Nolen-Hoeksema, 2003; Zimmerman et al., 2013): Anhedonia/loss of pleasure (sum of 5 reverse coded items), Cognitive difficulties (sum of 7 items), Positive mental health (sum of 17 items), Rumination (sum of 8 items), Dissociation (sum of 4 items), Mixed episodes (sum of 6 items), Other depression-related symptoms (sum of 16 items), Decrease in appetite (response to single item), Increase in appetite (response to single item), Decrease in weight (response to single item), Increase in weight (response to single item), Decrease/increase in appetite/weight (maximum of responses to 4 items), Sleep onset insomnia (response to single item), Mid-nocturna insomnia (response to single item), Early morning insomnia (response to single item), Hypersomnia (response to single item), Worst sleep problem (maximum of responses to 4 items), Sleep problems severity scale (sum of 4 items), Count of severe sleep problems[b]

b. Suicidality (Nock, Holmberg, Photos & Michel, 2007; Posner et al., 2009; Posner et al., 2011): 2-week suicidal ideation (Any, Number of days), Duration of suicidal ideation (response to single item), Controllability of suicidal ideation (response to single item), Onset of suicidal ideation (1 year ago, 5 years ago, more than 5 years ago), 2-week frequency of tempting fate (response to single item), Lifetime suicide attempt (Number of attempts, Ever injured, Most serious injuries were moderate/severe), 2-week suicidality scale[c]

2. Severity

a. 2-week depression symptom severity (Aish, Wasserman & Renberg, 2001; American Psychiatric Association, 2013; Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS), 2020; Kessler & Ustün, 2004; Llerena et al., 2013; Rizvi et al., 2015; Rush et al., 1996; Rush et al., 2003; Saffer et al., 2015; Zimmerman et al., 2013): Depression symptom frequency (sum of 43 standardized items), Hamilton Rating Scale for Depression severity levels (mild, moderate, severe, very severe), Quick Inventory of Depressive Symptomatology Self-Report Scale Score (sum of 16 items)

3. Subtypes

a. Endogenous depression (Kessler & Ustün, 2004; Sotsky et al., 1991; Ursano et al., 2014): Causes of lifetime depressive episodes (all/almost all caused by stressful experiences, most caused by stressful experiences, half caused by stressful experiences and half happened out of the blue, most/all happened out of the blue), Very first depressive episode was caused by stressful experiences (as opposed to happening out of the blue), Cause of current depressive episode (happened out of the blue, caused by recent stressful experiences only, caused by long-term stressful events only, caused by both recent and long-term stresses), Count of types of stresses that caused current depressive episode (sum of 4 items), Change in stresses since current depressive episode began (better or gone, somewhat better, staying the same, getting worse with no end in sight)

4. Clinical staging

a. Depression persistence (Akiskal et al., 2005; Kessler & Ustün, 2004; Sotsky et al., 1991; Ursano et al., 2014): Number of months of longest depressive episode, Number of years with severe depressive episodes, Percent of life with depression, Depression persistence-severity[d], Free of depression for at least 6 months between first and current episode

b. History of depression (Kessler & Ustün, 2004; Sotsky et al., 1991; Ursano et al., 2014): Age of first depressive episode, Number of years since onset of first depressive episode, Number of months of first depressive episode, First depressive episode in life is current episode, Number of months in current depressive episode before seeking treatment

5. Psychiatric comorbidity

a. Presenting problems (Akiskal et al., 2005; American Psychiatric Association, 2013; Anderson et al., 2018; Blevins, Weathers, Davis, Witte & Domino, 2015; Gibbons et al., 2016; Kessler & Ustün, 2004; Rush et al., 1996; Spielberger et al., 1983; Weissman et al., 2000; Zimmerman et al., 2013; Zuromski et al., 2019): Generalized anxiety (Only presenting problem, Primary problem, Primary/secondary problem, Primary/secondary/unreported

by met criteria, 4-category response)[e], Panic/phobias (Only presenting problem, Primary problem, Primary/secondary problem, Primary/secondary/unreported by met criteria, 4-category response)[f], PTSD (Only

Supplementary Table 2 (continued). Baseline self-report predictors[a] included in machine learning models

Risk factor domain

presenting problem, Primary problem, Primary/secondary problem, Primary/secondary/unreported by met criteria, 4-category response)[g], OCD (Only presenting problem, Primary problem, Primary/secondary problem, Primary/secondary/unreported by met criteria, 4-category response)[h], Alcohol/substance use problems (Only presenting problem, Primary problem, Primary/secondary problem, Primary/secondary/unreported by met criteria, 4-category response)[i], Anger control problems (Only presenting problem, Primary problem, Primary/secondary problem, Primary/secondary/unreported by met criteria, 4-category response)[j], Any other serious emotional problem (Only presenting problem, Primary problem, Primary/secondary problem, Primary/secondary/unreported by met criteria, 4-category response)[k], Depression (Only presenting problem, Primary/secondary problem, Exactly 1 other comorbid problem, Exactly 2 other comorbid problems, 2 or more other comorbid problems, 3 or more other comorbid problems), Count of primary problems, Count of secondary problems, Count of all primary/secondary problems, Count of comorbidities classified as not a problem, Count of other presenting problems

b. Comorbid presenting problems (Akiskal et al., 2005; American Psychiatric Association, 2013; Anderson et al., 2018; Blevins, Weathers, Davis, Witte & Domino, 2015; Gibbons et al., 2016; Kessler & Ustün, 2004; Rush et al., 1996; Spielberger et al., 1983; Weissman et al., 2000; Zimmerman et al., 2013; Zuromski et al., 2019): Weighted psychiatric comorbidities[l], Anxiety[m] and any externalizing[n] presenting problem, Anxiety[m] and PTSD as presenting problems, Any externalizing[n] and PTSD as presenting problems, At least 1 anxiety[m] and 1 externalizing[n] and PTSD as presenting problems, Count of anxiety presenting problems[m], Count of externalizing[n] presenting problems, Above the median of the sum of standardized panic/phobias, PTSD, and anxiety[m] comorbid presenting problem variables
c. 2-week PTSD (Blevins, Weathers, Davis, Witte & Domino, 2015; Zuromski et al., 2019): 6-Item Short-Form PCL-5 PTSD Screening Scale (sum of 6 items)[o], Number of months in life with PTSD-related symptoms
d. 30-day alcohol/substance use (Gibbons et al., 2016; Hamilton et al., 2011): Quantity-frequency of alcohol use (average number of drinks per day x frequency per week), Average number of nicotine products used per day, Frequency of drugs used per week (sum of 6 items), Marijuana used every or nearly every day, No drugs used, Heavy drinker[p], Heavy smoker[q], Heavy drug user[r], Count of alcohol/substance use related problems experienced at least once a week (sum of 7 items), Count of alcohol/substance use related problems experienced at least once a month (sum of 7 items)
e. 12-month disorders (Ursano et al., 2014): 12-month prevalence and number of months in past 12 with Generalized anxiety, Panic, Social anxiety, Specific phobia, Agoraphobia, PTSD, Obsessions/compulsions, Mania/bipolar disorder, Non-suicidal self-injurious behavior, Alcohol/substance use problems, Anger control problems, Any other serious emotional problem
f. Lifetime disorders (Ursano et al., 2014; Weissman et al., 2000): Lifetime history and number of years in life with Generalized anxiety, Panic, Social anxiety, Specific phobia, Agoraphobia, PTSD, Obsessions, Compulsions, Obsessions/compulsions, Mania/bipolar disorder, Non-suicidal self-injurious behavior, Alcohol/substance use problems, Anger control problems, Any other serious emotional problem, Lifetime history of Externalizing problems[n] (any, 2 or more), Count of lifetime externalizing problems[n], Count of lifetime anxiety problems[s], Count of all lifetime disorders

6. Functioning and quality of life

a. 2-week depression-related role impairment (Leon, Olfson, Portera, Farber & Sheehan, 1997): Full days out of role (any, number of days, percent of days), Partial days out of role (any, number of days, percent of days), Full and partial days out of role due (any, total number of days, total percent of days), Severe/very severe work impairment (response of 7-10 on 0-10 scale), Severe/very severe family life/home impairment (response of 7-10 on 0-10 scale), Severe/very severe social life impairment (response of 7-10 on 0-10 scale), Severe/very severe impairment in any area of life (response of 7-10 on 0-10 scale)

7. Early environmental exposures

a. Adverse childhood events (Dube et al., 2001; Weissman et al., 2000): Close loved one died, Close loved one attempted/died by suicide, Lived in a foster home, Sent to a juvenile detention center, Sent to a juvenile detention center or lived in a foster home, Parents/caregivers separated or divorced, Parent/caregiver was in prison for 6+ months, Parent/caregiver had a mental illness, Parent/caregiver had alcohol/substance use problems

b. Childhood trauma (Bernstein et al., 2003; Parker, Tupling & Brown, 1979): How often experienced Emotional/verbal abuse (sum of 2 items), Physical abuse (sum of 2 items), Sexual abuse (sum of 2 items),

Supplementary Table 2 (continued). Baseline self-report predictors[a] included in machine learning models

Risk factor domain

Emotional neglect (sum of 3 items), Physical neglect (sum of 3 items), How often had strict rules (response to single item), How often felt loved and cared for (sum of 3 items)

c. Parent/caregiver emotional problems (Weissman et al., 2000): Experienced depression (at least sometimes, at least often, very often), Depression and suicidality (depression at least often and suicidality at least rarely, depression very often and suicidality at least sometimes), Panic/generalized anxiety (at least sometimes, at least often), Mania/bipolar disorder (at least sometimes), Anger control problems (at least sometimes, at least often), Alcohol/substance use problems (at least sometimes, at least often, very often), Count of all emotional problems experienced[t] (sum of 8 items experienced at least often, sum of 8 items experienced very often), How often experienced a serious mental illness[u] (sum of 4 items), How often experienced psychological distress[v] (sum of 3 items)
d. Other adverse childhood experiences (Stein et al., 2018): How often family was on welfare or homeless during childhood (maximum of responses to 2 items)

8. Recent environmental stressors

a. 30-day chronic stress severity (Campbell-Sills et al., 2018b): Finances/career (maximum of responses to 2 items on 0-10 scale, severe/very severe stress=max response of 7-10, very severe stress=max response of 10), Personal health (response to single item on 0-10 scale, severe/very severe stress=response of 7-10, very severe stress=response of 10), Love life (response to single item on 0-10 scale, severe/very severe stress=response of 7-10, very severe stress=response of 10), Loved ones (maximum of responses to 2 items on 0-10 scale, severe/very severe stress=max response of 7-10, very severe stress=max response of 10), Relationships with family/others (maximum of responses to 2 items on 0-10 scale, severe/very severe stress=max response of 7-10, very severe stress=max response of 10), Life overall (response to single item on 0-10 scale, severe/very severe stress=response of 7-10, very severe stress =response of 10), Count of mild/no chronic stressors (sum of 5 items with responses of 0-3 on 0-10 scale), Count of severe/very severe chronic stressors (sum of 5 items with responses of 7-10 on 0-10 scale), Count of very severe chronic stressors (sum of 5 items with responses of 10 on 0-10 scale), How often per month currently experience physical bullying (response to single item), relational bullying (response to single item), verbal bullying (response to single item), any type of bullying (sum of 3 items)
b. 12-month stressful life events (Ursano et al., 2014): Experienced life-threatening illness/injury, Mugged/victim of armed robbery, Break-in/burglary, Mugged or break-in, Physically assaulted, Sexually assaulted or raped, Major financial crisis, Lost a job, Major financial crisis or lost a job, Serious trouble with police/arrested, Serious legal trouble/lawsuit, Serious trouble with police or legal trouble, Betrayal by someone close to you, Separation/divorce/serious romantic break-up, Break-up/falling out with close friend/relative, Betrayal or break-up Close friend/relative died, Close friend/relative had a life-threatening illness/injury, Close friend/relative experienced some other serious life crisis, Other stressful life event, Close friend/relative died or ill or other crisis, Experienced any kind of stressful life event, Experienced 2 or more stressful life events, Count of financial/legal stressful events, Count of all stressful life events

9. Personality scales

a. Personality traits (Akiskal et al., 2005; Anderson, Sellbom & Salekin, 2018; Cyders, Littlefield, Coffey & Karyadi, 2014; Gosling, Rentfrow & Swann, 2003; Zimmermann, Rossier, Meyer de Stadelhofen & Gaillard, 2005): Agreeableness scale (sum of 3 items), Alexithymia scale (sum of 4 items), Antagonism/antisocial personality traits (sum of 4 items), Detachment scale (sum of 4 items), Emotionality scale (sum of 2 items), Externally oriented thinking scale (sum of 3 items), Extraversion/openness scale (sum of 4 items), Impulsive/sensation-seeking scale (sum of 6 items), Negative urgency scale (sum of 5 items), Psychoticism personality traits (sum of 5 items)
b. Temperament (Akiskal et al., 2005; Anderson et al., 2018; Costa & McCrae, 1992; Gosling et al., 2003; Spielberger, Gorsuch, Lushene, Vagg & Jacobs, 1983): Anxious (sum of 5 items), Cyclothymic (sum of 5 items), Depressive (response to single item), Hyperthymic (sum of 3 items), Irritable (sum of 2 items)
c. Attachment style (Bartholomew & Horowitz, 1991): Dismissing-avoidant (response to single item), Fearful-avoidant (response to single item), Preoccupied/anxious-resistant (response to single item), Secure (response to single item)

Supplementary Table 2 (continued). Baseline self-report predictors[a] included in machine learning models

Risk factor domain

10. Neurocognition

a. Attentional control (Judah, Grant, Mills & Lechner, 2014): Distractibility scale (sum of 3 items), Low flexibility scale (sum of 3 items)

11. Dysfunctional cognitive schemas

a. Interpersonal needs (Van Orden, Cukrowicz, Witte & Joiner, 2012): Perceived burdensomeness (sum of 3 items), Perceived access to social support/thwarted belongingness (response to single item)
b. Cognitive distortions (Akiskal et al., 2005; Roberts, 2015): Cognitive distortions scale (sum of 6 items)

12. Protective/resilience factors

a. Social support (Kessler & Ustün, 2004; Schuster, Kessler & Aseltine, 1990): Religiosity scale (sum of 3 items), Frequency of interaction with friends/relatives (response to single item), Frequency of participation in social groups (response to single item), Seeking help with personal problems (reach out to a lot of different people, only to family/friends/closest confidants, no one), How much could rely on people for support with personal problems (response to single item), Access to social support (could rely a lot on people for support with personal problems, could rely some or more on people for support with personal problems), Number of people could rely on for support with personal problems, Number of confidants, Negative social networks (response to how often people make too many demands = response to how often people argue with you)
b. Emotional regulation (Garnefski & Kraaij, 2007; Gross & John, 2003; Medrano & Trogolo, 2016; Schlotz, Yim, Zoccola, Jansen & Schulz, 2011): Cognitive reappraisal scale (sum of 6 items), Difficulties in regulation of emotional response scale (sum of 4 items), Difficulties in processing emotions scale (sum of 4 items), Putting things into perspective/stress management skills scale (sum of 3 items), Refocus on planning scale (sum of 3 items), Self-blame scale (sum of 3 items), Perceived stress reactivity scale (sum of 7 items), Reactivity/regulation scale (Sum of Perceived stress reactivity scale, Difficulties in regulation of emotional response scale, Difficulties in processing emotions scale)
c. Resilience (Campbell-Sills et al., 2018a; Campbell-Sills & Stein, 2007): Perceived psychological resilience scale (sum of 12 items)

13. Comorbid physical disorders

a. Medications (Kessler & Ustün, 2004): Number of medications taken per day for ongoing physical problems
b. Health care visits (Kessler & Ustün, 2004): Number of health care visits for physical problems in the past 12 months
c. Health indicators: BMI (underweight, normal weight, normal weight or more, overweight, overweight or more, obese, obese or more, morbidly obese), Rating of overall physical health (excellent, very good, good, fair, poor, 5-category response, very good or better, good or better, fair or better)
d. Continuity of care (Safran et al., 1998): Provider of routine physical health care (has a regular PCP, has a regular place, no regular PCP or regular place), Number of years going to the regular PCP/place for routine physical health care
e. TBIs (Ursano et al., 2014): Lifetime prevalence, Number of lifetime TBIs (1, 2 or more, total), Age of first TBI, 12-month prevalence
f. Somatic symptoms (Axelsson, Andersson, Ljótsson, Wallhed Finn & Hedman, 2016; Toussaint et al., 2016): Severity of Distressing/bothersome symptoms in past 30 days (frequency of symptoms per week x length per day x severity of symptoms), Duration of symptoms (none, 1-3 months, 4-6 months, 7-12 months, 1-2 years, more than 2 years, 6-category response, 1+ months, 4+ months, 7+ months, 1+ years), Perception of symptom severity (sum of 8 items), Anxiety about symptoms (sum of 4 items[w]), Somatic symptom disorder scale (sum of 15 items)

14. Family history of psychopathology

a. Family history of depression (Weissman et al., 2000): Number of parents/relatives with a history of depression

15. Socio-demographics

a. Age: Age at baseline survey (19-34, 35-49, 50-59, 60+, 35 or more, 50 or more, 60 or more)

b. Children: Any biological or stepchildren, Number of children (0, 1, 2, 3+), Age of oldest child (under 6, under 13, under 18), Age of youngest child (under 6, under 13, under 18), Currently pregnant/partner currently pregnant

Supplementary Table 2 (continued). Baseline self-report predictors[a] included in machine learning models

Risk factor domain

c. Education: High school or less, Some college, College graduate, Graduate school or more, 4-category response, Some college or less, College graduate or less
d. Employment status: Employed, Retired, Student, Disabled, Unemployed, Other
e. Nativity: Born in the US
f. Occupational category: Executive/administrator/senior manager, Professional, Technical support, Sales, Clerical/administrative support, Service, Crafts worker/precision production, Operator/laborer, Other
g. Race/ethnicity: Non-Hispanic white, Non-Hispanic black, Hispanic, Other
h. Marital/relationship status: Relationship status (married/cohabitating, engaged, steadily dating, dating but not in a steady relationship, not currently dating, 5-category response), Number of years married/steadily dating, Quality of marriage/relationship (response to single item on 0-10 scale)
i. Sex: Male/Female

16. Treatment characteristics

a. Treatment preferences (Kessler & Ustün, 2004; Milosevic, Levy, Alcolado & Radomsky, 2015; Steidtmann et al., 2012): Likelihood to participate in a RCT of new antidepressant (response to single item), Willingness to try psychotherapy (response to single item on 0-10 scale, unwilling=response of 0-2, moderately willing=response of 3-7, completely willing=response of 8-10, 3-category response), Willingness to try ADM (response to single item on 0-10 scale, unwilling=response of 0-2, moderately willing=response of 3-7, completely willing=response of 8-10, 3-category response), Willingness to try other kinds of treatment (response to single item on 0-10 scale), ADM preference (preferred specific type of antidepressant, preferred specific class)
b. Treatment expectations (Curry et al., 2006; McHorney, Victor Spain, Alexander & Simmons, 2009; Unni, 2008; Unni, Olson & Farris, 2014; Vik, Maxwell & Hogan, 2004; Wisniewski, Rush, Balasubramani, Trivedi & Nierenberg, 2006): Expectation for psychotherapy to successfully treat depression (response to single item on 0-10 scale, failure=response of 0-2, moderate success=response of 3-7, complete success=response of 8-10, 3-category response), Expectation for ADM to successfully treat depression (response to single item on 0-10 scale, failure=response of 0-2, moderate success=response of 3-7, complete success=response of 8-10, 3-category response), Expectation for combination of psychotherapy and ADM to successfully treat depression (response to single item on 0-10 scale, failure=response of 0-2, moderate success=response of 3-7, complete success=response of 8-10, 3-category response), Expectation for ECT to successfully treat depression (response to single item on 0-10 scale, failure=response of 0-2, moderate success=response of 3-7, complete success=response of 8-10, 3-category response), Expectation for ketamine therapy to successfully treat depression (response to single item on 0-10 scale, failure=response of 0-2, moderate success=response of 3-7, complete success=response of 8-10, 3-category response), Expectation for ADM side effects (No side effects, Mild but temporary, Mild and long-term, Moderate but temporary, Moderate and long-term, Severe but temporary, Severe and long-term, Any, Any mild, Any moderate, Any severe, Any temporary, Any long-term), Concerns about ADM (sum of 4 items), Expectation for success of current treatment (response to single item), Belief in current treatment being the best (response to single item)
c. Health literacy (Chew et al., 2008; Haun, Valerio, McCormack, Sørensen & Paasche-Orlow, 2014): Number of hours spent discussing depression treatment options before initial visit, Number of hours spent researching depression treatment options before initial visit, Total number of hours spent discussing/researching depression treatment options before initial visit, Inadequate health literacy (sum of 3 items)
d. Treatment provider (Baumann, Baumann, Le Bihan & Chau, 2008; Bieber, Müller, Nicolai, Hartmann & Eich, 2010; Safran et al., 1998): Type of provider currently treating depression (PCP/NP, psychiatrist, psychologist, PCP/NP and psychiatrist, PCP/NP and psychologist, psychiatrist and psychologist), Previously received mental health treatment from same provider currently treating depression, Quality of relationship with provider (sum of 8 items)
e. Current assigned treatment (Kessler & Ustün, 2004): Individual/group counseling/psychotherapy, Medication
f. Treatment history (Kessler & Ustün, 2004; Sotsky et al., 1991; Ursano et al., 2014): Received any treatment in the past, Age of first receiving treatment, Percent of years in treatment, Number of times hospitalized overnight for depression, Current episode is first episode/first time receiving treatment, Current treatment is the same as previous self-reported treatment, Types of treatment received in the past (Individual counseling/psychotherapy, Group counseling/psychotherapy, ADM, Internet guided self-help therapy/self-help support

group/ECT/ketamine/other, Combinations of different types), Count of different types of treatment ever received (sum of 8 items), Helpfulness of past treatment (was helpful, was not helpful), Effectiveness of past

Supplementary Table 2 (continued). Baseline self-report predictors[a] included in machine learning models

Risk factor domain

psychotherapy treatment (at least very effective, completely effective, maximum of responses to 2 items), Effectiveness of past ADM treatment (never received, not effective, not very, somewhat effective, at least somewhat effective, very effective, at least very effective, completely effective, response to single item)

g. Treatment compliance (Morisky, Ang, Krousel-Wood & Ward, 2008): Medication non-compliance scale (sum of 5 items)

Abbreviations: PTSD, post-traumatic stress disorder; OCD, obsessive-compulsive disorder; BMI, body mass index; PCP, primary care provider; TBI, traumatic brain injury; RCT, randomized controlled trial; ADM, antidepressant medication; ECT, electroconvulsive therapy; NP, nurse practitioner.

[a]We included dichotomous variables, nested dichotomous variables, categorical and ordinal variables, continuous variables, standardized and stabilized continuous variables as potential predictors in the feature selection for additive algorithms.

[b]Severe sleep problems included taking 30 minutes or more to fall asleep, waking up throughout the night a few times or more, waking up too early more than half the time or more, and sleeping longer than 9 hours a day or more.

[c]Standardized sum of 4 or more days with suicide ideation in the past 2 weeks; suicidal thoughts in past 2 weeks were somewhat difficult, very difficult, or impossible to control; and tempted fate rarely, sometimes, often, or very often in the past 2 weeks.

[d]This variable was created by summing the standardized percent of life with depression variable, standardized number of months of longest depressive episode variable, and standardized depressive temperament item "People tell me I am often unable to see the lighter side of things". Higher scores indicate higher depressive persistence/severity/temperament.

[e]Generalized anxiety as either a primary or secondary presenting problem as indicated by the patient or if the patient reported anxiety 7 or more months of the past 12 and reported symptoms of DSM-5 Generalized Anxiety Disorder Criteria A and C.

[f]Panic/phobias as either a primary or secondary presenting problem as indicated by the patient or if the patient reported panic/phobias 7 or more months of the past 12.

[g]PTSD as either a primary or secondary presenting problem as indicated by the patient or a score in the clinical range of a PCL-5 screening scale calibrated to the full PCL-5.

[h]OCD as either a primary or secondary presenting problem as indicated by the patient or if the patient reported having obsessions or compulsions every month for the past 12 months.

[i]Alcohol/substance use problems as either a primary or secondary presenting problem as indicated by the patient or if the patient met threshold scoring rules on the Patient-Reported Outcomes Measurement Information System (PROMIS) 30-day Alcohol/Substance Use Short Form-7a.

[j]Anger control problems as either a primary or secondary presenting problem as indicated by the patient or if the patient reported anger attacks 7 or more months of the past 12 and reported that one or more of three statements described them either a lot or exactly (“Sometimes I get so furious that I could hurt someone”; “I snap at people when I get angry”; “Sometimes I get so mad that I trash everything”).

[k]Any other serious emotional problem as either a primary or secondary presenting problem as indicated by the patient or if the patient reported any other serious emotional problem 7 or more months of the past 12.

[l]Count of possible self-reported comorbidities classified as primary, secondary, unreported by met criteria, or not a presenting problem.

[m]Anxiety included panic/phobias, OCD, and generalized anxiety.

[n]Externalizing problems included anger control problems, alcohol/substance use problems, and any other serious emotional problem.

[o]Defined as a score of 38+ on the PTSD Checklist for DSM-5 based on responses to the 6-item short-form version of the PCL-5 calibrated to the full PCD-5.

[p]30-day heavy drinking was defined as alcohol quantity*frequency >= 15.

[q]30-day heavy smoking was defined as using 26 or more nicotine products per day.

[r]30-day heavy drug use was defined as self-reported use of prescription opioids (either without a doctor’s prescription or more than prescribed to get high, buzzed, or numbed out) or heroin/street fentanyl at least once a month or using prescription stimulants, tranquilizers, muscle relaxers, or sedatives (either without a doctor’s prescription or more than prescribed to get high, buzzed, or numbed out) at least 1 day a week or using marijuana every or nearly every day or using any other illegal nonprescription drug at least 1 day a week in the past 30 days.

[s]Anxiety problems included generalized anxiety, panic, social anxiety, specific phobia, agoraphobia, PTSD, obsessions, and compulsions.

[t]Emotional problems included depression, panic, generalized anxiety, mania/bipolar disorder, anger control problems, alcohol/substance use problems, suicidality, and any other serious emotional problem.
[u]Serious mental illnesses included mania/bipolar disorder, anger control problems, alcohol/substance use problems, and any other serious emotional problem.
[v]Psychological distress included depression, panic, and generalized anxiety.
[w]Sum of 30-day frequency of anxiety/worry/distress about symptoms, severity of anxiety/worry/distress, time spent thinking/focusing on symptoms, severity of interference in life due to symptoms or anxiety/worry/distress about symptoms.

Supplementary Table 3. Administrative predictors[a] included in machine learning models

Risk factor domain

1. Psychiatric comorbidity
   a. Health care visits: Number of visits with any ICD code for Suicidal Ideation[b], 2 or more visits with ICD code for diagnosis of any substance use problems in past 12 months and at least 1 was face-to-face
   b. Diagnoses (Lew et al., 2009): ICD code for diagnosis of PTSD in past 12 months, Lifetime history of previously diagnosed mental disorders (Any adjustment disorders, anxiety disorders, depressive disorders, PTSD, substance use disorders, other reactions to stress, other mental disorders, total number), Mental disorders in EHR at initial visit (Any adjustment disorders, anxiety disorders, depression disorders, PTSD, substance use disorders, other reactions to stress, other mental disorders, total number), Primary mental disorder in EHR at initial visit (Any adjustment disorders, anxiety disorders, depressive disorders, PTSD, substance use disorders, other reactions to stress, other mental disorders), Primary diagnosis at initial visit (depression, a mental disorder other than depression and not a physical disorder, a physical disorder and not depression), Any visit with ICD code for diagnosis of Polytrauma Clinical Triad - Depression[b], Any visit with ICD code for diagnosis of Polytrauma Clinical Triad – PTSD[b]

2. Recent environmental stressors
   a. Accidents: Number of visits with any ICD code for Accidents (past 6 months, past 12 months, past 2 years, past 5 years)

3. Comorbid physical disorders
   a. Medications: Prescribed any Antimigraine Medications[b], Prescribed any Non-Opioid Analgesic Medications[b], Prescribed any Opioid Analgesic Medications[b], Prescribed any Pain Medications[b], Prescribed any Medications
   b. Health care visits (Chronic Pain Research Alliance, 2015; Lew et al., 2009; Mayhew et al., 2019): Number of visits with CPT code for any Sleep apnea related procedures[b], Number of visits with CPT code for any Pain related procedures[b], Number of visits with ICD code for diagnosis of any of the 10 Chronic Overlapping Pain Conditions[b], Number of visits with ICD code for diagnosis of any of the 13 Pain Condition Crosswalk Clusters[b], Any visit with ICD code for diagnosis of Polytrauma Clinical Triad - Chronic Pain[b], Any visit with ICD code for diagnosis of Polytrauma Clinical Triad - TBI[b], Any visit with ICD code for diagnosis of Polytrauma Clinical Triad - Chronic Pain and TBI in the past 3 months
   c. Health indicators (Charlson, Pompei, Ales & MacKenzie, 1987): Charlson Comorbidity Index Score (0, 1, 2 or more, 3-category response), Worst pain on 0-10 NRS pain scale[b]

4. Socio-demographics
   a. Marital status: Currently married, divorced, separated, widowed, never married, previously married, 3-category response (currently married, previously married, never married)
   b. Housing stability: Ever homeless[b], Number of visits with any ICD code for Problems with housing, material resources, and social isolation[b]
   c. Neighborhood characteristics: Census region (Northeast, Midwest, South, West)

5. Treatment characteristics
   a. Current treatment characteristics: Received a referral to specialist during initial visit, Received psychotherapy at a primary care facility, Received psychotherapy at a specialty mental health facility, Number of psychotherapy visits in the 84 days after the initial visit
   b. Treatment provider: Treatment provider at initial visit was a primary care provider (as opposed to a mental health specialist)
   c. Facility characteristics: Driving time in minutes to the nearest VHA primary care facility, Ratio of medical/social positions lost/onboards at the facility where the patient visited in the year prior to the initial visit, Setting of initial visit (community based outpatient clinic, specialty mental health, primary care), Facility of initial visit had at least one full-time integrated mental health specialist on staff, Facility of initial visit did not have a full-time integrated mental health specialist on staff
   d. Treatment history: Number of visits with any ICD code for Noncompliance with treatment in the past 5 years

---

Abbreviations: ICD, International Classification of Diseases; PTSD, post-traumatic stress disorder; EHR, electronic health record; CPT, Current Procedural Terminology; TBI, traumatic brain injury; NRS, numeric rating scale; VHA, Veterans Health Administration.

[a]We included dichotomous variables, nested dichotomous variables, categorical and ordinal variables, continuous variables, standardized and stabilized continuous variables as potential predictors in the feature selection for additive algorithms.
[b]In the past month, past 2 months, past 3 months, past 6 months, past 12 months, past 2 years, and past 5 years before initial visit.

Supplementary Table 4. Geospatial predictors[a] included in machine learning models

Risk factor domain

1. Socio-demographics

a. GDP (Bureau of Economic Analysis, 2021): Per capita county GDP in thousands of chain-linked 2012 dollars[b], Per capita nominal county GDP in real-time dollars, Difference in per capita county GDP (chain-linked 2012 dollars) from previous year
b. Poverty (United States Census Bureau, 2020): Percent of population below 150% of the poverty line at the Census Block Group
c. Income (United States Department of Commerce & Bureau of Economic Analysis, 2018): Per capita personal income (by residence) in thousands of dollars in the county
d. Homelessness (United States Department of Housing and Urban Development, 2014): Annual rate of homelessness per 1,000 Census Track population on a single given night in January
e. Quality of health (University of Wisconsin Population Health Institute & Robert Wood Johnson Foundation, 2021): Composite health outcomes measure based on length of life and quality of life in the county[c], Overall health outcome summary score in the county (higher = worse)[c], Years of potential life lost before age 75 (age-adjusted) per 100,000 in the county population
f. Healthcare coverage (Centers for Medicare and Medicaid Services (CMS), 2018): Medicaid eligible rate per capita in the county
g. Mortality rates (National Center for Health Statistics (NCHS), 2019): Infant mortality rate (<1 year old) from all causes per 100,000 infants in the county, Mortality rate due to alcohol, drugs, external causes, HIV/AIDS, homicide, liver disease per 100,000 in the county population, Suicide rate (by any method) per 100,000 in the county population
h. Urbanicity (United States Department of Agriculture: Economic Research Service, 2020): Major metro area, urban area, rural area

---

Abbreviation: GDP, Gross domestic product; HIV/AIDS, human immunodeficiency virus, acquired immunodeficiency syndrome.

[a]We included dichotomous variables, nested dichotomous variables, categorical and ordinal variables, continuous variables, standardized and stabilized continuous variables as potential predictors in the feature selection for additive algorithms.

[b]Inflation-adjusted measure of area's gross product, based on national prices for the goods and serviced produced within the area. The real estimates of GDP are measured in chained (2012) dollars.

[c]The county-level composite health outcomes measure is a sum of the following standardized variables: Years of potential life lost before age 75 (age-adjusted) per 100,000 people, Percent of adults reporting fair or poor health (age-adjusted), Average number of days in a month with poor physical health (age-adjusted), Average number of days in a month with poor mental health (age-adjusted), Percent of very low weight live births (<2,500 grams); the overall health outcomes summary score is a sum of the variables (non-standardized forms) used in the Composite health outcomes measure, with higher scores = worse county-level health outcomes.

Supplementary Table 5. Algorithms used in the Super Learner ensemble machine learning analysis[a]

Algorithm	Description
I. Super Learner	Super Learner is an ensemble machine learning approach that uses cross-validation (CV) to select a weighted combination of predicted outcome scores across a collection of candidate algorithms (learners) to yield an optimal combination according to a pre-specified criterion that performs at least as well as the best component algorithm. R package: *SuperLearner* (Polley, LeDell, Kennedy, Lendle & van der Laan, 2018; van der Laan, Polley & Hubbard, 2007)
II. Learners in the Super Learner library	
A. Logistic regression	Maximum likelihood estimation with logistic link function. R package: s*tats* (Nelder & Wedderburn, 1972)
B. Elastic Net	Elastic net is a regularization method that minimizes the problem of overlap among predictors by explicitly penalizing over-fitting with a composite penalty λ{MPP x Plasso + (1- MPP) X Pridge}, where MPP is a mixing parameter penalty with values between 0 and 1 that controls relative weighting between the lasso penalty (Plasso) and the ridge penalty (Pridge). The parameter λ controls the total amount of penalization. The ridge penalty handles multicollinearity by shrinking all coefficients smoothly towards 0 but retains all variables in the model. The lasso penalty allows simultaneous coefficient shrinkage and variable selection, tending to select at most one predictor in each strongly correlated set, but at the expense of giving unstable estimates in the presence of high multicollinearity. The elastic net approach of combining the ridge and lasso penalties has the advantage of yielding more stable and accurate estimates than either ridge or lasso alone while maintaining model parsimony. R package: *glmnet* (Friedman, Hastie & Tibshirani, 2010) Hyperparameters[a]: α=(0.0, 0.2, 0.4, 0.6, 0.8, 1.0).
C. Splines	
C1. Adaptive splines	Adaptive spline regression flexibly captures both linear and piecewise non-linear associations as well as interactions among these associations by connecting linear segments (splines) of varying slopes and smooths to create piece-wise curves (basis functions). Final fit is built using a stepwise procedure that selects the optimal combination of basis functions. R package: *earth* (Milborrow, 2016) Hyperparameters[a]: degree = (1, 3, 5), penalty = (2, 4, 6).
C2. Adaptive polynomial splines	Adaptive polynomial splines are like adaptive splines but differ in the order in which basis functions (e.g., linear versus nonlinear) are added to build the final model. R package: *polspline* (Kooperberg, 2015)
D. Decision trees – bagging	Random Forest. Independent variables are partitioned (based on contiguous values) and stacked to build decision trees that are combined (ensemble) to create an aggregate "forest". Random forest builds numerous trees in bootstrapped samples and generates an aggregate prediction by averaging across trees, thereby reducing over-fitting. R package: *ranger* (Wright & Ziegler, 2017) Hyperparameters[a]: max.depth = (2, 3, 4, 5), min.node.size = (16, 64, 256, 400), num.trees = (3000), mtry = ($\lfloor\sqrt{\#\ variables}\rfloor$,$\lfloor\sqrt{\#\ variables}/2\rfloor$)

Supplementary Table 5 (continued). Algorithms used in the Super Learner ensemble machine learning analysis[a]

Algorithm	Description
E. Support vector machines	Support vector machines treat independent variables as dimensions in high dimensional space and attempt to identify the best hyperplane (linear, polynomial, radial, or sigmoid kernel) to separate the sample into classes (e.g., cases and non-cases) with maximum distance between classes. R package: *WeightSVM* (Xu, 2020) Hyperparameters[a]: kernel = (radial), cost = (0.1, 1, 10, 50), gamma (0.0001, 0.001, 0.01, 0.1, 1)
F. Decision trees – boosting	
F1. Gradient Boosting Machine	GBMs build a sequential ensemble of shallow successive decision trees that iteratively learn the residuals from prior trees. This is a flexible method, where the number of trees, interaction depth, and shrinkage are leveraged to build flexible models. R package: *CatBoost* (Prokhorenkova, Gusev, Vorobev, Dorogush & Gulin, 2019) Hyperparameters[a,b]: Iterations = (500, 1000, 2000), learning rate = (0.05, 0.03, 0.01), depth = (2, 3, 4, 5), min data in leaf = (4, 16, 64, 256), max_leaves = (2, 4, 8, 16, 32)
F2. Extreme Gradient Boosting	A fast and efficient implementation of gradient boosting. R package: *XGBoost* (Chen & Guestrin, 2016) Hyperparameters[a,b]: ntrees = (5000), max_depth = (2, 3, 4, 5), shrinkage = (0.10, 0.05, 0.01) gamma = (0, 4, 16, 64), minobspernode = (3, 9, 27, 81), colsample_bytree = (1.0, 0.8, 0.6), subsample = (1, 0.9, 0.8), colsample_bynode = (1, 0.9, 0.8)
G. Discrete Bayesian Additive Regression Trees Sampler	Bayesian trees are based on an underlying probability model (priors) for the structure and likelihood for data in terminal nodes. The aggregate tree is generated by averaging across tree posteriors (reducing overfit). R package: *dbarts* (Dorie et al., 2021) Hyperparameters[a]: ntree=100
H. Mean	Arithmetic mean
I. Stratified outcome prevalence	Stratified outcome prevalence

[a]Hyperparameters: Default values were used unless otherwise noted. Algorithms included for all combinations of hyperparameters.
[b]Random selection of 100 hyperparameter combinations included in analysis for gradient boosting algorithms. Learners were selected into the SuperLearner in two stages. In the first stage, cross-validated predicted probabilities and AUCs for each learner were estimated by fitting the learners in the training sample using 10-fold cross-validation. Resubstitution predicted probabilities and AUCs were also estimated by fitting the learners in the full training sample and estimating predicted probabilities in the same sample. Learners with a high cross-validated AUC, that were not overfit/underfit by resubstitution AUC, and that had a small drop-off in performance between resubstitution and cross-validation were selected for the next stage. This selection was done by algorithm and at least one learner from each algorithm was carried forward. In the second stage, learners selected from the first stage were entered into a SuperLearner, which was then fit to the training sample using 10-fold cross-validation. Learners that were assigned weights in at least 2 folds when solving for the non-negative least squares were selected for the final SuperLearner model.

Supplementary Table 6. Distribution of socio-demographic characteristics, baseline depression severity, and treatment response among the full baseline sample, analytic sample, and patients lost to follow-up

	Weighted for baseline non-response						Also weighted for loss to follow-up		Difference between analytic sample and patients lost to follow-up	
	Baseline sample (n=809)		Analytic sample (n=660)		Patients lost to follow-up (n=149)		Analytic sample (n=660)			
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	X^2	Df
Age									11.65[a]	3
19-34	24.2	(1.6)	22.5	(1.8)	31.6	(4.0)	23.3	(1.9)		
35-49	30.7	(1.8)	31.0	(2.0)	29.4	(3.9)	30.7	(2.1)		
50-59	19.3	(1.5)	21.6	(1.7)	9.9	(2.6)	20.8	(1.7)		
60+	25.7	(1.6)	24.9	(1.8)	29.1	(3.9)	25.2	(1.8)		
Sex									0.90	1
Female	26.3	(1.7)	27.1	(1.9)	23.0	(3.7)	26.1	(1.9)		
Male	73.7	(1.7)	72.9	(1.9)	77.0	(3.7)	73.9	(1.9)		
Race/ethnicity									1.62	3
White	62.3	(1.9)	62.8	(2.1)	60.1	(4.3)	67.9	(2.1)		
Black	18.2	(1.5)	17.3	(1.7)	22.0	(3.8)	13.6	(1.5)		
Hispanic	12.4	(1.3)	12.5	(1.4)	11.8	(2.9)	11.7	(1.5)		
Other	7.2	(1.0)	7.4	(1.1)	6.0	(2.1)	6.9	(1.1)		
Marital status									2.14	4
Currently married	55.6	(1.9)	54.4	(2.1)	60.5	(4.3)	54.9	(2.2)		
Divorced	20.9	(1.5)	21.4	(1.7)	18.6	(3.5)	21.3	(1.7)		
Separated	5.9	(0.9)	5.8	(1.0)	6.4	(2.1)	5.9	(1.1)		
Widowed	2.3	(0.5)	2.5	(0.6)	1.5	(1.1)	2.7	(0.7)		
Never married	15.4	(1.4)	15.9	(1.6)	13.0	(2.9)	15.2	(1.6)		
Census region									5.86	3
Northeast	8.4	(1.1)	7.3	(1.1)	13.0	(3.1)	8.0	(1.2)		
Midwest	17.8	(1.4)	17.4	(1.6)	19.1	(3.3)	19.6	(1.8)		
South	53.8	(1.9)	54.1	(2.1)	52.6	(4.3)	51.5	(2.2)		
West	20.0	(1.5)	21.2	(1.7)	15.3	(3.0)	20.9	(1.8)		
Urbanicity									2.55	2
Major metro	84.6	(1.3)	83.5	(1.5)	88.9	(2.6)	79.5	(1.8)		
Urban	14.2	(1.3)	15.2	(1.5)	10.0	(2.5)	18.8	(1.7)		
Rural	1.2	(0.4)	1.2	(0.4)	1.1	(.8)	1.7	(0.6)		
% of population below 1.5x of poverty line									2.75	3
Least low income	18.7	(1.5)	19.7	(1.7)	14.5	(3.0)	21.5	(1.9)		
2nd quartile	26.2	(1.6)	25.3	(1.8)	29.8	(3.9)	28.2	(2.0)		
3rd quartile	29.0	(1.7)	28.6	(1.9)	31.0	(4.0)	27.0	(1.9)		
Most low income	26.1	(1.6)	26.4	(1.8)	24.7	(3.7)	23.2	(1.7)		
Type of ADM									2.25	6
TeCA (Mirtazapine)	5.8	(0.9)	5.9	(1.0)	5.1	(1.9)	4.7	(0.8)		
NDRI (Bupropion)	15.7	(1.3)	15.3	(1.5)	17.2	(3.2)	15.7	(1.5)		
SARI (Trazodone)	15.0	(1.4)	15.1	(1.5)	14.2	(3.0)	15.8	(1.7)		
SNRIs	16.8	(1.4)	16.2	(1.5)	19.1	(3.4)	15.6	(1.6)		
Duloxetine	11.1	(1.2)	10.2	(1.2)	15.2	(3.1)	9.2	(1.2)		
Venlafaxine	5.6	(0.8)	6.1	(1.0)	3.8	(1.6)	6.4	(1.1)		

Supplementary Table 6 (continued). Distribution of socio-demographic characteristics, baseline depression severity, and treatment response among the full baseline sample, analytic sample, and patients lost to follow-up

	Weighted for baseline non-response						Also weighted for loss to follow-up			
	Baseline sample (n=809)		Analytic sample (n=660)		Patients lost to follow-up (n=149)		Analytic sample (n=660)		Difference betwe analytic sample a patients lost to follow-up	
	%	(SE)	%	(SE)	%	(SE)	%	(SE)	X^2	df
SSRI	57.0	(1.9)	56.9	(2.1)	57.7	(4.3)	59.7	(2.1)		
Citalopram	5.4	(0.8)	5.6	(0.9)	4.5	(1.8)	6.7	(1.2)		
Escitalopram	14.4	(1.3)	13.7	(1.4)	17.7	(3.3)	14.2	(1.6)		
Fluoxetine	8.5	(1.0)	8.5	(1.2)	8.4	(2.3)	8.2	(1.1)		
Fluvoxamine	0.2	(0.1)	0.2	(0.2)	0.0	--	0.1	(0.1)		
Paroxetine	2.8	(0.6)	2.8	(0.7)	2.9	(1.5)	3.1	(0.7)		
Sertraline	25.8	(1.6)	26.1	(1.8)	24.2	(3.7)	27.4	(2.0)		
SMS	0.3	(0.2)	0.4	(0.3)	0.0	--	0.6	(0.5)		
Vilazodone	0.1	(0.1)	0.2	(0.2)	0.0	--	0.5	(0.5)		
Vortioxetine	0.1	(0.1)	0.2	(0.2)	0.0	--	0.1	(0.1)		
TCA	2.9	(0.6)	3.1	(0.7)	2.0	(1.2)	3.0	(0.6)		
Amitriptyline	1.1	(0.4)	1.3	(0.4)	0.6	(0.6)	1.4	(0.4)		
Doxepin	0.7	(0.3)	0.6	(0.3)	0.7	(0.7)	0.5	(0.3)		
Nortriptyline	1.1	(0.4)	1.2	(0.5)	0.7	(0.7)	1.1	(0.4)		
Baseline Depression Severity[b]									6.26	3
Mild	30.1	(1.7)	31.2	(1.9)	25.5	(3.7)	33.4	(2.1)		
Moderate	35.6	(1.8)	34.3	(2.0)	41.1	(4.3)	33.7	(2.1)		
Severe	21.4	(1.6)	20.6	(1.7)	24.9	(3.8)	20.0	(1.8)		
Very Severe	12.9	(1.3)	13.9	(1.5)	8.6	(2.2)	12.9	(1.4)		
Treatment response	--		36.5	(2.0)	--		35.7	(2.1)		
(n)	(809)		(660)		(149)		(660)			

Abbreviations: SE, standard error; df, degrees of freedom; ADM, antidepressant medication; TeCA, tetracyclic antidepressants; NDRI, norepinephrine-dopamine reuptake inhibitors; SARI, serotonin antagonist reuptake inhibitors; SNRIs, serotonin-norepinephrine reuptake inhibitors; SSRI, selective serotonin reuptake inhibitor; SMS, serotonin modulator and stimulator; TCA, tricyclic antidepressants.

[a]Significant at the .05 level, two-sided test.

[b]Based on a transformation of baseline QIDS-SR scores to Hamilton Rating Scale of Depression categories using published transformation rules (Table 3 in Rush et al., 2003).

Supplementary Table 7. Relative risk of fairness in Super Learner top tertile full sample using robust standard errors (n=660)

	Age				Sex				Race/ethnicity				
	Main effects		Interaction		Main effects		Interaction		Main effects		Interaction		
	RR	(95% CI)	RR	(95% CI)	RR	(95% CI)	RR	(95% CI)	RR	(95% CI)	RR	(95% CI)	
SL top tertile	3.3[a]	(2.6-4.3)	3.2[a]	(2.3-4.4)	3.3[a]	(2.6-4.2)	4.1[a]	(2.4-6.9)	3.3[a]	(2.6-4.2)	3.2[a]	(2.0-5.1)	3
51 years old or older	0.9	(0.7-1.1)	0.8	(0.5-1.3)	--	--	--	--	--	--	--	--	
Male	--	--	--	--	1.1	(0.8-1.3)	1.3	(0.8-2.3)	--	--	--	--	
Non-Hispanic White	--	--	--	--	--	--	--	--	1.0	(0.8-1.3)	1.0	(0.6-1.6)	
College grad or more	--	--	--	--	--	--	--	--	--	--	--	--	0
Interaction	--	--	1.0	(0.7-1.9)	--	--	0.8	(0.4-1.4)	--	--	1.0	(0.6-1.8)	

Abbreviations: RR, relative risk; CI, confidence interval.

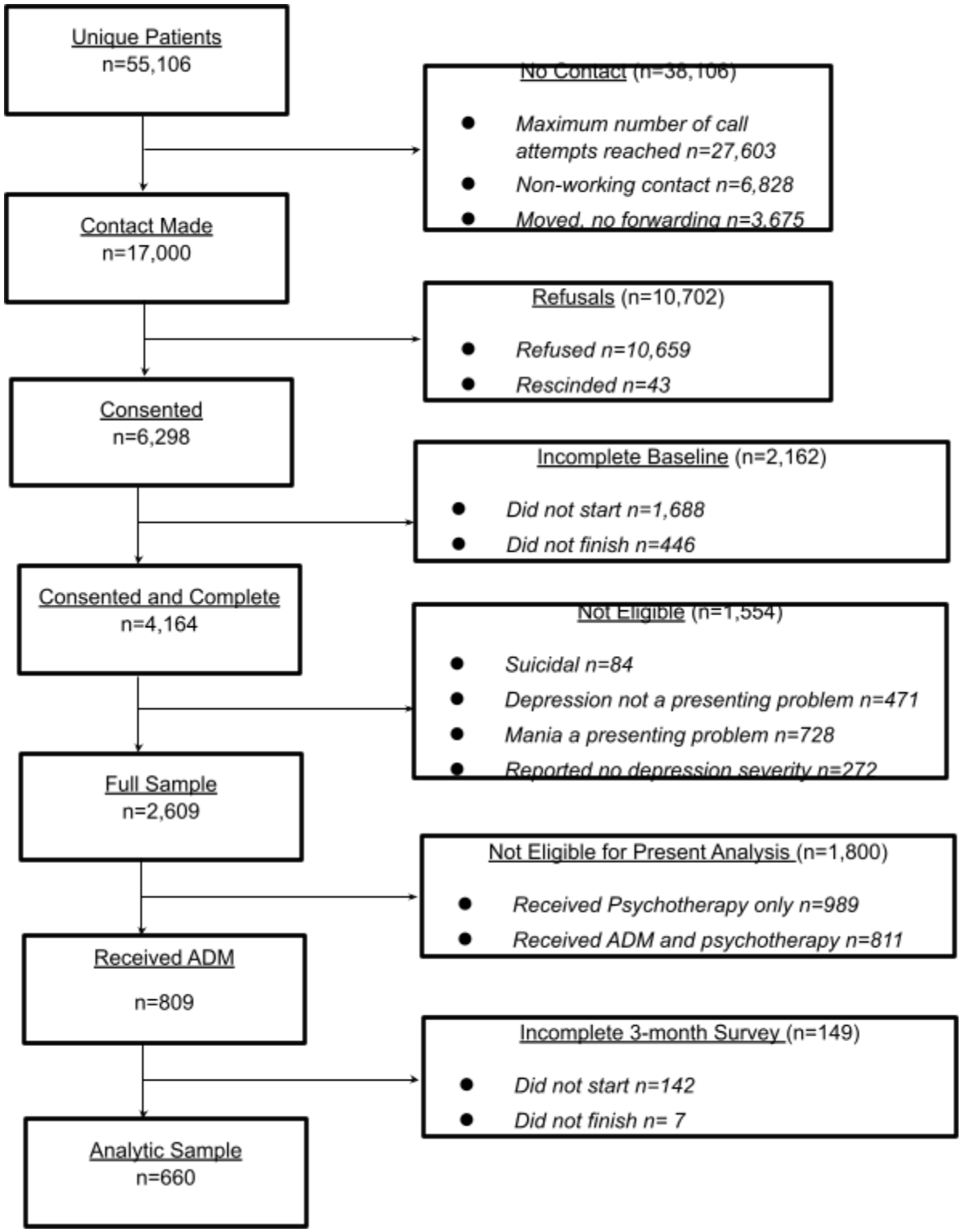
[a]Significant at the .05 level, two-sided test.

Supplementary Table 8. Description of predictors selected by the Super Learner model

Predictor	Description[a]
Depression sx freq	Sum of 43 standardized items
Severe 30-day financial stress (D)	--
Positive mental health sx freq (S)	Stabilized sum of 17 standardized items
Psychiatric comorbidities above median (D)	--
County drug overdose mortality rate	Annual rate per 100,000 population
Referred to specialist at initial visit (D)	--
Emotional regulation (S)	Stabilized sum of 6 standardized items
Quality of relationship with provider (S)	Stabilized sum of 8 standardized items
Freq social group participation (S)	Stabilized response to single item
Stress management skills (S)	Stabilized sum of 3 standardized items
Expect mild/temporary ADM side effects (D)	--
Past psychotherapy helpfulness (S)	Stabilized maximum of responses to 2 items
Past tx helpful (D)	--
Anhedonia sx freq (RS)	Reversed stabilized sum of 5 standardized items
Psychiatric comorbidities (S)	Stabilized count of possible comorbidities classified as not being presenting problems
Willingness to try psychotherapy (S)	Stabilized response to single item
Expect no ADM side effects (D)	--
Perceived psychological resilience (S)	Stabilized sum of 12 standardized items
Live in Midwest (D)	--
Intake carried out by psychologist (D)	--
Occupation: Crafts worker (D)	--
Intermediate psychiatric comorbidities (D)	--
Past ADM helpfulness (S)	Stabilized response to single item
Perceived access to social support (S)	Stabilized response to single item
Job loss/financial crisis in past 12 months (D)	--
Willingness to try psychotherapy (S)	Truncated 3-category stabilized response to single item
Stressor caused MDE, but now improved (D)	--
Severe 30-day stress (D)	--
Px has regular primary care provider (D)	--
High access to social support (D)	--
Belief in current tx being best (S)	Stabilized response to single item
Expect psychotherapy to be moderately successful (D)	--
Expect mild ADM side effects (D)	--
Weighted psychiatric comorbidities (S)	Stabilized count of possible self-reported comorbidities classified as primary (2 points), secondary (1 point), or not a presenting problem. Multiple presenting problems of each type could be reported
History of medication noncompliance (S)	Stabilized response to single item
Live in South (D)	--
Past psychotherapy completely effective (D)	--
Major financial crisis in past 12 months (D)	--
MDE is primary diagnosis (D)	--
Expectation for success of current tx (S)	Stabilized response to single item.
Family history of MDE	Count
ADM provided in regular PCP office (D)	--
Lifetime panic (D)	--
12-month PTSD (D)	--
Completely willing to try psychotherapy (D)	--
Years since first MDE	Count
Current tx psychotherapy (D)	--
Employed (D)	--
Regular PCP is prescribing ADM (D)	--
Excellent-good self-rated physical health (D)	--
BMI: Obese (D)	--
ADM prescribed in specialty setting (D)	--
Mild MDE based on HRSD (D)	--

Abbreviations: sx, symptoms; freq, frequency; (D), dummy variable; (S), stabilized variable; ADM, antidepressant medication; tx, treatment; (RS), reverse stabilized; MDE, major depressive episode; px, patient; PCP, primary care provider; PTSD, post-traumatic stress disorder; BMI, body mass index; HRSD, Hamilton Rating Scale of Depression.
[a]Descriptions are provided only for predictors that are not dichotomies.

Supplementary Figure 1. Flow diagram of patients recruited into the study among those seen for incident depression as reported in the Veterans Health Administration electronic medical records from 12/2018 - 6/2020



REFERENCES

Aish, A. M., Wasserman, D. & Renberg, E. S. (2001) Does Beck's Hopelessness Scale really measure several components? *Psychological Medicine, 31*(2), 367-372. doi:10.1017/s0033291701003300

Akiskal, H. S., Mendlowicz, M. V., Jean-Louis, G., Rapaport, M. H., Kelsoe, J. R., Gillin, J. C. & Smith, T. L. (2005) TEMPS-A: validation of a short version of a self-rated instrument designed to measure variations in temperament. *Journal of Affective Disorders, 85*(1-2), 45-52. doi:10.1016/j.jad.2003.10.012

American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders: DSM-5. *Arlington, VA*.

Anderson, J. L., Sellbom, M. & Salekin, R. T. (2018) Utility of the Personality Inventory for DSM-5-Brief Form (PID-5-BF) in the Measurement of Maladaptive Personality and Psychopathology. *Assessment, 25*(5), 596-607. doi:10.1177/1073191116676889

Axelsson, E., Andersson, E., Ljótsson, B., Wallhed Finn, D. & Hedman, E. (2016) The health preoccupation diagnostic interview: inter-rater reliability of a structured interview for diagnostic assessment of DSM-5 somatic symptom disorder and illness anxiety disorder. *Cognitive Behaviour Therapy, 45*(4), 259-269. doi:10.1080/16506073.2016.1161663

Bartholomew, K. & Horowitz, L. M. (1991) Attachment styles among young adults: a test of a four-category model. *Journal of Personality and Social Psychology, 61*(2), 226-244. doi:10.1037//0022-3514.61.2.226

Baumann, M., Baumann, C., Le Bihan, E. & Chau, N. (2008) How patients perceive the therapeutic communications skills of their general practitioners, and how that perception

affects adherence: use of the TCom-skill GP scale in a specific geographical area. *BMC Health Services Research, 8*, 244. doi:10.1186/1472-6963-8-244

Bernstein, D. P., Stein, J. A., Newcomb, M. D., Walker, E., Pogge, D., Ahluvalia, T., . . . Zule, W. (2003) Development and validation of a brief screening version of the Childhood Trauma Questionnaire. *Child Abuse & Neglect, 27*(2), 169-190. doi:10.1016/s0145-2134(02)00541-0

Bieber, C., Müller, K. G., Nicolai, J., Hartmann, M. & Eich, W. (2010) How does your doctor talk with you? Preliminary validation of a brief patient self-report questionnaire on the quality of physician-patient interaction. *Journal of Clinical Psychology in Medical Settings, 17*(2), 125-136. doi:10.1007/s10880-010-9189-0

Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K. & Domino, J. L. (2015) The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation. *Journal of Traumatic Stress, 28*(6), 489-498. doi:10.1002/jts.22059

Bureau of Economic Analysis. (2021). Gross Domestic Product CAGDP1: GDP Summary by County and MSA Dataset. Retrieved from https://apps.bea.gov/regional/downloadzip.cfm.

Campbell-Sills, L., Kessler, R. C., Ursano, R. J., Sun, X., Taylor, C. T., Heeringa, S. G., . . . Stein, M. B. (2018a) Predictive validity and correlates of self-assessed resilience among U.S. Army soldiers. *Depression and Anxiety, 35*(2), 122-131. doi:10.1002/da.22694

Campbell-Sills, L. & Stein, M. B. (2007) Psychometric analysis and refinement of the Connor-davidson Resilience Scale (CD-RISC): Validation of a 10-item measure of resilience. *Journal of Traumatic Stress, 20*(6), 1019-1028. doi:10.1002/jts.20271

Campbell-Sills, L., Ursano, R. J., Kessler, R. C., Sun, X., Heeringa, S. G., Nock, M. K., . . . Stein, M. B. (2018b) Prospective risk factors for post-deployment heavy drinking and alcohol or substance use disorder among US Army soldiers. *Psychological Medicine, 48*(10), 1624-1633. doi:10.1017/s0033291717003105

Centers for Medicare and Medicaid Services (CMS). (2018). CMS Medicaid Analytic eXtract (MAX) Person Summary File: Area Health Resources Files. Retrieved from https://data.hrsa.gov/data/download.

Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases, 40*(5), 373-383. doi:10.1016/0021-9681(87)90171-8

Chen, T. & Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Chew, L. D., Griffin, J. M., Partin, M. R., Noorbaloochi, S., Grill, J. P., Snyder, A., . . . Vanryn, M. (2008) Validation of screening questions for limited health literacy in a large VA outpatient population. *Journal of General Internal Medicine, 23*(5), 561-566. doi:10.1007/s11606-008-0520-5

Chronic Pain Research Alliance. (2015). Impact of chronic overlapping pain conditions on public health and the urgent need for safe and effective treatment. Retrieved from http://www.chronicpainresearch.org/public/CPRA_WhitePaper_2015-FINAL-Digital.pdf.

Costa, P. T. & McCrae, R. R. (1992) *Neo personality inventory-revised (NEO PI-R)*. Psychological Assessment Resources: Odessa, FL.

Curry, J., Rohde, P., Simons, A., Silva, S., Vitiello, B., Kratochvil, C., . . . March, J. (2006) Predictors and moderators of acute outcome in the Treatment for Adolescents with Depression Study (TADS). *Journal of the American Academy of Child and Adolescent Psychiatry, 45*(12), 1427-1439. doi:10.1097/01.chi.0000240838.78984.e2

Cyders, M. A., Littlefield, A. K., Coffey, S. & Karyadi, K. A. (2014) Examination of a short English version of the UPPS-P Impulsive Behavior Scale. *Addictive Behaviors, 39*(9), 1372-1376. doi:10.1016/j.addbeh.2014.02.013

Dorie, V., Chipman, H., McCulloch, R., Dadgar, A., Team, R. C., Draheim, G. U., . . . Richard G, D. (2021). Discrete Bayesian Additive Regression Trees Sampler. Retrieved from https://cran.r-project.org/web/packages/dbarts/dbarts.pdf.

Dube, S. R., Anda, R. F., Felitti, V. J., Chapman, D. P., Williamson, D. F., & Giles, W. H. (2001) Childhood abuse, household dysfunction, and the risk of attempted suicide throughout the life span: findings from the Adverse Childhood Experiences Study. JAMA, 286(24), 3089–3096. doi:10.1001/jama.286.24.3089

Friedman, J., Hastie, T. & Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statisitcal Software, 33*(1), 1-22.

Garnefski, N. & Kraaij, V. (2007) The Cognitive Emotion Regulation Questionnaire. *European Journal of Psychological Assessment, 23*(3), 141-149. doi:10.1027/1015-5759.23.3.141

Gibbons, L. E., Fredericksen, R., Merrill, J. O., McCaul, M. E., Chander, G., Hutton, H., . . . Crane, P. K. (2016) Suitability of the PROMIS alcohol use short form for screening in a HIV clinical care setting. *Drug and Alcohol Dependence, 164*, 113-119. doi:10.1016/j.drugalcdep.2016.04.038

Gosling, S. D., Rentfrow, P. J. & Swann, W. B. (2003) A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504-528. doi:https://doi.org/10.1016/S0092-6566(03)00046-1

Gross, J. J. & John, O. P. (2003) Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology, 85*(2), 348-362. doi:10.1037/0022-3514.85.2.348

Hamilton, C. M., Strader, L. C., Pratt, J. G., Maiese, D., Hendershot, T., Kwok, R. K., . . . Haines, J. (2011) The PhenX Toolkit: get the most from your measures. *American Journal of Epidemiology, 174*(3), 253-260. doi:10.1093/aje/kwr193

Haun, J. N., Valerio, M. A., McCormack, L. A., Sørensen, K. & Paasche-Orlow, M. K. (2014) Health Literacy Measurement: An Inventory and Descriptive Summary of 51 Instruments. *Journal of Health Communication, 19*(sup2), 302-333. doi:10.1080/10810730.2014.936571

Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS). (2020). Interpretation. Retrieved from http://www.ids-qids.org/interpretation.html.

Judah, M. R., Grant, D. M., Mills, A. C. & Lechner, W. V. (2014) Factor structure and validation of the Attentional Control Scale. *Cognition & Emotion, 28*(3), 433-451. doi:10.1080/02699931.2013.835254

Kessler, R. C. & Ustün, T. B. (2004) The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research, 13*(2), 93-121. doi:10.1002/mpr.168

Kooperberg, C. (2015). polspline: Polynomial Spline Routines. Retrieved from https://mran.microsoft.com/snapshot/2015-11-04/web/packages/polspline/polspline.pdf.

Leon, A. C., Olfson, M., Portera, L., Farber, L. & Sheehan, D. V. (1997) Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *International Journal of Psychiatry in Medicine, 27*(2), 93-105. doi:10.2190/t8em-c8yh-373n-1uwd

Lew, H. L., Otis, J. D., Tun, C., Kerns, R. D., Clark, M. E. & Cifu, D. X. (2009) Prevalence of chronic pain, posttraumatic stress disorder, and persistent postconcussive symptoms in OIF/OEF veterans: polytrauma clinical triad. *Journal of Rehabilitation Research and Development, 46*(6), 697-702. doi:10.1682/jrrd.2009.01.0006

Llerena, K., Park, S. G., McCarthy, J. M., Couture, S. M., Bennett, M. E. & Blanchard, J. J. (2013) The Motivation and Pleasure Scale-Self-Report (MAP-SR): reliability and validity of a self-report measure of negative symptoms. *Comprehensive Psychiatry, 54*(5), 568-574. doi:10.1016/j.comppsych.2012.12.001

Mayhew, M., DeBar, L. L., Deyo, R. A., Kerns, R. D., Goulet, J. L., Brandt, C. A. & Von Korff, M. (2019) Development and Assessment of a Crosswalk Between ICD-9-CM and ICD-10-CM to Identify Patients with Common Pain Conditions. *The Journal of Pain, 20*(12), 1429-1445. doi:10.1016/j.jpain.2019.05.006

McHorney, C. A., Victor Spain, C., Alexander, C. M. & Simmons, J. (2009) Validity of the adherence estimator in the prediction of 9-month persistence with medications prescribed for chronic diseases: a prospective analysis of data from pharmacy claims. *Clinical therapeutics, 31*(11), 2584-2607. doi:10.1016/j.clinthera.2009.11.030

Medrano, L. & Trogolo, M. (2016) Construct Validity of the Difficulties in Emotion Regulation Scale: Further Evidence Using Confirmatory Factor Analytic Approach. *Abnormal and Behavioural Psychology, 2*(2). doi:10.4172/2472-0496.1000117

Milborrow, S. (2016). earth: Multivariate Adaptive Regression Splines. Retrieved from https://cran.r-project.org/web/packages/earth/index.html.

Milosevic, I., Levy, H. C., Alcolado, G. M. & Radomsky, A. S. (2015) The Treatment Acceptability/Adherence Scale: Moving Beyond the Assessment of Treatment Effectiveness. *Cognitive Behaviour Therapy, 44*(6), 456-469. doi:10.1080/16506073.2015.1053407

Morisky, D. E., Ang, A., Krousel-Wood, M. & Ward, H. J. (2008) Predictive validity of a medication adherence measure in an outpatient setting. *Journal of Clinical Hypertension (Greenwich, Conn.), 10*(5), 348-354. doi:10.1111/j.1751-7176.2008.07572.x

National Center for Health Statistics (NCHS). (2019). NCHS Mortality Data Files: The County Health Rankings & Roadmaps National Data. Retrieved from https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation.

Nelder, J. A. & Wedderburn, R. W. M. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society: Series A (General), 135*(3), 370-384. doi:https://doi.org/10.2307/2344614

Nock, M. K., Holmberg, E. B., Photos, V. I. , & Michel, B. D. (2007) The Self-Injurious Thoughts and Behaviors Interview: Development, reliability, and validity in an adolescent sample. Psychological Assessment, 19(3), 309-317. doi: 10.1037/1040-3590.19.3.309

Parker, G., Tupling, H. & Brown, L. B. (1979) A Parental Bonding Instrument. *British Journal of Medical Psychology, 52*(1), 1-10. doi:https://doi.org/10.1111/j.2044-8341.1979.tb02487.x

Polley, E., LeDell, E., Kennedy, C., Lendle, S. & van der Laan, M. (2018). SuperLearner: Super Learner Prediction. Retrieved from https://cran.r-project.org/web/packages/SuperLearner/index.html.

Posner, K., Brent, D., Lucas, C., Gould, M., Stanley, B., Brown, G., . . . Mann, J. (2009) Columbia-suicide severity rating scale (C-SSRS). *New York, NY: Columbia University Medical Center*.

Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., . . . Mann, J. J. (2011) The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *The American Journal of Psychiatry, 168*(12), 1266-1277. doi:10.1176/appi.ajp.2011.10111704

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulin, A. (2019). CatBoost: unbiased boosting with categorical features. Retrieved from https://arxiv.org/abs/1706.09516.

Rizvi, S. J., Quilty, L. C., Sproule, B. A., Cyriac, A., Michael Bagby, R. & Kennedy, S. H. (2015) Development and validation of the Dimensional Anhedonia Rating Scale (DARS) in a community sample and individuals with major depression. *Psychiatry Research, 229*(1-2), 109-119. doi:10.1016/j.psychres.2015.07.062

Roberts, M. B. (2015) Inventory of Cognitive Distortions: Validation of a Measure of Cognitive Distortions Using a Community Sample. *PCOM Psychology Dissertations*, 325.

Rush, A. J., Gullion, C. M., Basco, M. R., Jarrett, R. B. & Trivedi, M. H. (1996) The Inventory of Depressive Symptomatology (IDS): psychometric properties. *Psychological Medicine, 26*(3), 477-486. doi:10.1017/s0033291700035558

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., . . . Keller, M. B. (2003) The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*(5), 573-583. doi:10.1016/s0006-3223(02)01866-8

Saffer, B. Y., Lanting, S. C., Koehle, M. S., Klonsky, E. D. & Iverson, G. L. (2015) Assessing cognitive impairment using PROMIS(®) applied cognition-abilities scales in a medical outpatient sample. *Psychiatry Research, 226*(1), 169-172. doi:10.1016/j.psychres.2014.12.043

Safran, D. G., Kosinski, M., Tarlov, A. R., Rogers, W. H., Taira, D. H., Lieberman, N. & Ware, J. E. (1998) The Primary Care Assessment Survey: tests of data quality and measurement performance. *Medical Care, 36*(5), 728-739. doi:10.1097/00005650-199805000-00012

Schlotz, W., Yim, I. S., Zoccola, P. M., Jansen, L. & Schulz, P. (2011) The Perceived Stress Reactivity Scale: measurement invariance, stability, and validity in three countries. *Psychological Assessment, 23*(1), 80-94. doi:10.1037/a0021148

Schuster, T. L., Kessler, R. C. & Aseltine, R. H., Jr. (1990) Supportive interactions, negative interactions, and depressed mood. *American Journal of Community Psychology, 18*(3), 423-438. doi:10.1007/bf00938116

Sotsky, S. M., Glass, D. R., Shea, M. T., Pilkonis, P. A., Collins, J. F., Elkin, I., . . . et al. (1991) Patient predictors of response to psychotherapy and pharmacotherapy: findings in the

NIMH Treatment of Depression Collaborative Research Program. *The American Journal of Psychiatry, 148*(8), 997-1008. doi:10.1176/ajp.148.8.997

Spielberger, C., Gorsuch, R., Lushene, R., Vagg, P. R. & Jacobs, G. (1983) *Manual for the State-Trait Anxiety Inventory (Form Y1 – Y2)*.

Steidtmann, D., Manber, R., Arnow, B. A., Klein, D. N., Markowitz, J. C., Rothbaum, B. O., . . . Kocsis, J. H. (2012) Patient treatment preference as a predictor of response and attrition in treatment for chronic depression. *Depression and Anxiety, 29*(10), 896-905. doi:10.1002/da.21977

Stein, M. B., Campbell-Sills, L., Ursano, R. J., Rosellini, A. J., Colpe, L. J., He, F., . . . Kessler, R. C. (2018) Childhood Maltreatment and Lifetime Suicidal Behaviors Among New Soldiers in the US Army: Results From the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *The Journal of Clinical Psychiatry, 79*(2). doi:10.4088/JCP.16m10900

Toussaint, A., Murray, A. M., Voigt, K., Herzog, A., Gierk, B., Kroenke, K., . . . Löwe, B. (2016) Development and Validation of the Somatic Symptom Disorder-B Criteria Scale (SSD-12). *Psychosomatic Medicine, 78*(1), 5-12. doi:10.1097/psy.0000000000000240

Treynor, W., Gonzalez, R. & Nolen-Hoeksema, S. (2003) Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research, 27*(3), 247-259.

United States Census Bureau. (2020). 2010-2018 American Community Survey Dataset: 1-Year and 5-Year Estimates. Retrieved from https://www2.census.gov/.

United States Department of Agriculture: Economic Research Service. (2020). Rural-Urban Continuum Codes Data Sets. Retrieved from https://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx.

United States Department of Commerce & Bureau of Economic Analysis. (2018). Personal Income Summary - Personal Income, Population, Per Capita Personal Income File: Area Health Resources Files. Retrieved from https://data.hrsa.gov/data/download.

United States Department of Housing and Urban Development. (2014). HUD Exchange Annual Homeless Assessment Report: Point-in-Time (PIT) Counts by CoC Data Files. Retrieved from https://www.hudexchange.info/homelessness-assistance/ahar/#2014-and-prior-reports.

University of Wisconsin Population Health Institute & Robert Wood Johnson Foundation. (2021). The County Health Rankings & Roadmaps. County Health Rankings National Data, CHR SAS Analytic Data. Retrieved from https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation.

Unni, E. J. (2008). Development of models to predict medication non-adherence based on a new typology. Retrieved from https://www.proquest.com/openview/6ff64a8168fb86e704b5ab650bece61a/1?pq-origsite=gscholar&cbl=18750.

Unni, E. J., Olson, J. L. & Farris, K. B. (2014) Revision and validation of Medication Adherence Reasons Scale (MAR-Scale). *Current Medical Research and Opinion, 30*(2), 211-221. doi:10.1185/03007995.2013.851075

Ursano, R. J., Colpe, L. J., Heeringa, S. G., Kessler, R. C., Schoenbaum, M. & Stein, M. B. (2014) The Army study to assess risk and resilience in servicemembers (Army STARRS). *Psychiatry, 77*(2), 107-119. doi:10.1521/psyc.2014.77.2.107

van der Laan, M. J., Polley, E. C. & Hubbard, A. E. (2007) Super learner. *Statistical Applications in Genetics and Molecular Biology, 6*, Article 25. doi:10.2202/1544-6115.1309

Van Orden, K. A., Cukrowicz, K. C., Witte, T. K. & Joiner, T. E. (2012) Thwarted belongingness and perceived burdensomeness: construct validity and psychometric properties of the Interpersonal Needs Questionnaire. *Psychological Assessment, 24*(1), 197-215. doi:10.1037/a0025358

Vik, S. A., Maxwell, C. J. & Hogan, D. B. (2004) Measurement, correlates, and health outcomes of medication adherence among seniors. *The Annals of Pharmacotherapy, 38*(2), 303-312. doi:10.1345/aph.1D252

Weissman, M. M., Wickramaratne, P., Adams, P., Wolk, S., Verdeli, H. & Olfson, M. (2000) Brief screening for family psychiatric history: the family history screen. *Archives of General Psychiatry, 57*(7), 675-682. doi:10.1001/archpsyc.57.7.675

Wisniewski, S. R., Rush, A. J., Balasubramani, G. K., Trivedi, M. H. & Nierenberg, A. A. (2006) Self-rated global measure of the frequency, intensity, and burden of side effects. *Journal of Psychiatric Practice, 12*(2), 71-79. doi:10.1097/00131746-200603000-00002

Wright, M. N. & Ziegler, A. (2017) ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *Journal of Statisitcal Software, 77*(1), 1 - 17. doi:10.18637/jss.v077.i01

Xu, T. (2020). WeightSVM: Subject Weighted Support Vector Machines. Retrieved from https://cran.r-project.org/web/packages/WeightSVM/index.html.

Zimmerman, M., Martinez, J. H., Attiullah, N., Friedman, M., Toba, C., Boerescu, D. A. & Ragheb, M. (2013) A new type of scale for determining remission from depression: the

Remission from Depression Questionnaire. *Journal of Psychiatric Research, 47*(1), 78-82. doi:10.1016/j.jpsychires.2012.09.006

Zimmermann, G., Rossier, J., Meyer de Stadelhofen, F. & Gaillard, F. (2005) Alexithymia Assessment and Relations with Dimensions of Personality. *European Journal of Psychological Assessment, 21*(1), 23-33. doi:10.1027/1015-5759.21.1.23

Zuromski, K. L., Ustun, B., Hwang, I., Keane, T. M., Marx, B. P., Stein, M. B., . . . Kessler, R. C. (2019) Developing an optimal short-form of the PTSD Checklist for DSM-5 (PCL-5). *Depression and Anxiety, 36*(9), 790-800. doi:10.1002/da.22942