

## ARTICLE



# Developing an individualized treatment rule for Veterans with major depressive disorder using electronic health records

Nur Hani Zainal<sup>1</sup>, Robert M. Bossarte<sup>2</sup>, Sarah M. Gildea<sup>1</sup>, Irving Hwang<sup>1</sup>, Chris J. Kennedy<sup>3</sup>, Howard Liu<sup>1,4</sup>, Alex Luedtke<sup>5,6</sup>, Brian P. Marx<sup>7,8</sup>, Maria V. Petukhova<sup>1</sup>, Edward P. Post<sup>9,10</sup>, Eric L. Ross<sup>11</sup>, Nancy A. Sampson<sup>1</sup>, Erik Sverdrup<sup>12</sup>, Brett Turner<sup>13</sup>, Stefan Wager<sup>12</sup> and Ronald C. Kessler<sup>1</sup>✉

© The Author(s), under exclusive licence to Springer Nature Limited 2024

Efforts to develop an individualized treatment rule (ITR) to optimize major depressive disorder (MDD) treatment with antidepressant medication (ADM), psychotherapy, or combined ADM-psychotherapy have been hampered by small samples, small predictor sets, and suboptimal analysis methods. Analyses of large administrative databases designed to approximate experiments followed iteratively by pragmatic trials hold promise for resolving these problems. The current report presents a proof-of-concept study using electronic health records (EHR) of  $n = 43,470$  outpatients beginning MDD treatment in Veterans Health Administration Primary Care Mental Health Integration (PC-MHI) clinics, which offer access not only to ADMs but also psychotherapy and combined ADM-psychotherapy. EHR and geospatial databases were used to generate an extensive baseline predictor set (5,865 variables). The outcome was a composite measure of at least one serious negative event (suicide attempt, psychiatric emergency department visit, psychiatric hospitalization, suicide death) over the next 12 months. Best-practices methods were used to adjust for nonrandom treatment assignment and to estimate a preliminary ITR in a 70% training sample and to evaluate the ITR in the 30% test sample. Statistically significant aggregate variation was found in overall probability of the outcome related to baseline predictors (AU-ROC = 0.68, S.E. = 0.01), with test sample outcome prevalence of 32.6% among the 5% of patients having highest predicted risk compared to 7.1% in the remainder of the test sample. The ITR found that psychotherapy-only was the optimal treatment for 56.0% of patients (roughly 20% lower risk of the outcome than if receiving one of the other treatments) and that treatment type was unrelated to outcome risk among other patients. Change in aggregate treatment costs of implementing this ITR would be negligible, as 16.1% fewer patients would be prescribed ADMs and 2.9% more would receive psychotherapy. A pragmatic trial would be needed to confirm the accuracy of the ITR.

*Molecular Psychiatry* (2024) 29:2335–2345; <https://doi.org/10.1038/s41380-024-02500-0>

## INTRODUCTION

Major depressive disorder (MDD) is a highly burdensome disorder [1, 2]. The most common first-line treatments are antidepressant medication (ADM), psychotherapy, and combined ADM-psychotherapy [3]. Meta-analyses of randomized controlled trials find psychotherapies either alone or in combination with ADM have better aggregate MDD symptom response and remission than ADM-only over 6- to 12-month follow-up periods [4]. A recent meta-analysis found that psychotherapy-only is also associated with significantly lower risk of a composite serious negative outcome (either suicide attempt, ED visit, psychiatric hospitalization, and/or suicide death in the 12 months after initiating treatment) than either ADM-only or combined treatment [5]. Most patients also prefer psychotherapy with or without ADM to ADM-only [6]. Yet the great majority of MDD patients are

treated with ADM-only [3] due to its greater availability [7] and lower costs [8].

The above observations suggest that aggregate treatment response would improve if the number of MDD patients receiving psychotherapy increased. This is unrealistic in the near term, though, given shortages in trained psychotherapists. However, if the benefits of psychotherapy are much greater for some patients than others, it might be possible to increase access to psychotherapy for the patients who need it most if a method existed to determine these patients. Numerous predictors of differential MDD treatment response have been suggested for this purpose [9–15]. However, none of these predictors is sufficiently strong to be of practical value by itself. Recent studies consequently have attempted to develop an individualized treatment rule (ITR) that combines information across a range

<sup>1</sup>Department of Health Care Policy, Harvard Medical School, Boston, MA, USA. <sup>2</sup>Department of Psychiatry and Behavioral Neurosciences, University of South Florida, Tampa, FL, USA. <sup>3</sup>Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. <sup>4</sup>Center of Excellence for Suicide Prevention, Canandaigua VA Medical Center, Canandaigua, NY, USA. <sup>5</sup>Department of Statistics, University of Washington, Seattle, WA, USA. <sup>6</sup>Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. <sup>7</sup>National Center for PTSD, VA Boston Healthcare System, Boston, MA, USA. <sup>8</sup>Department of Psychiatry, Boston University School of Medicine, Boston, MA, USA. <sup>9</sup>Center for Clinical Management Research, VA Ann Arbor Health Care System, Ann Arbor, MI, USA. <sup>10</sup>Department of Medicine, University of Michigan Medical School, Ann Arbor, MI, USA. <sup>11</sup>Department of Psychiatry, Larner College of Medicine, University of Vermont, Burlington, VT, USA. <sup>12</sup>Graduate School of Business, Stanford University, Stanford, CA, USA. <sup>13</sup>Harvard T.H. Chan School of Public Health, Boston, MA, USA. ✉email: [kessler@hcp.med.harvard.edu](mailto:kessler@hcp.med.harvard.edu)

Received: 16 August 2023 Revised: 23 February 2024 Accepted: 27 February 2024  
Published online: 14 March 2024

of prescriptive predictors to optimize the selection of either (i) psychotherapy vs. ADM [16–20], (ii) combined treatment vs. ADM [21, 22], (iii) combined treatment vs. psychotherapy [23], or (iv) combined treatment vs. either of the two monotherapies [24–29]. These studies have been limited, though, by small samples, limited predictor sets, and suboptimal analysis methods [30, 31].

We present here an attempt to move beyond prior ITR development studies by determining if a preliminary ITR can be developed from data in a large electronic health record (EHR) administrative data system. We focus on patients treated in Veterans Health Administration (VHA) Primary Care Mental Health Integration (PC-MHI) clinics, as such clinics offer not only ADMs but also psychotherapy and combined ADM-psychotherapy. The outcome is a composite measure of experiencing one or more serious negative events in the 365 after initiating treatment: suicide attempts, psychiatric emergency department visits, psychiatric urgent care visits, psychiatric hospitalizations, or suicide-related deaths. Prior research showed that all these negative outcomes are significantly elevated among MDD patients in PC-MHI clinics [32, 33].

As this is an observational study, not a randomized controlled trial, results will be biased to the extent that type of treatment is nonrandom with respect to confounders. Great care was consequently taken to use best-practices methods described below and in the Supplementary Methods section to adjust for nonrandom treatment assignment and improve both aggregate model accuracy and fairness [34]. Previous research shows that observational analyses based on such methods can be useful in generating preliminary ITRs so long as they are followed by pragmatic trials in which the ITRs are experimentally evaluated and refined [35].

## METHODS

### Sample

Our sample consisted of all  $n = 43,470$  PC-MHI patients beginning MDD outpatient treatment with either an ADM and/or psychotherapy between October 1, 2015 and December 31, 2016. Exclusion criteria included: (i) any previous MDD treatment in the past 365 days; (ii) any lifetime diagnosis of autism, bipolar disorder, borderline intellectual functioning, dementia, intellectual disability, nonaffective psychosis, stereotyped movement disorder, or Tourette's disorder; (iii) any lifetime VHA treatment with antimanic or antipsychotic medication; or (iv) an administratively recorded suicide attempt in the prior 365 days. ADM-only treatment was defined as receiving an ADM prescription but not a referral for psychotherapy during the first PC-MHI visit. Psychotherapy-only was defined as either seeing a psychotherapist or receiving an appointment for such a visit but not receiving an ADM prescription on the day of the initial PC-MHI visit. Combined treatment was defined as receiving an ADM prescription and either seeing a psychotherapist or receiving an appointment for such a visit the day of the first PC-MHI visit. The human subjects committees of both Harvard Medical School and the Canandaigua VA Boston Medical Center approved these procedures. HIPAA and VHA waivers of consent were obtained to conduct secondary analyses with this dataset.

### Outcome

The serious negative events making up our outcome were selected because information about these events, unlike the symptom severity measures conventionally used as outcomes in depression treatment trials, is available for all PC-MHI patients in the VHA Corporate Data Warehouse [36], VHA Suicide Prevention Applications Network (SPAN; [37]), or National Death Index (NDI; [38]). As detailed in the Supplementary Literature Review, a recent meta-analysis found 34 published randomized trials that assessed comparative treatment effects of ADM-only, psychotherapy-only, and combined treatment on these outcomes [5], but none of those studies attempted to develop an ITR to optimize treatment assignment. It is important to recognize that an ITR to minimize risk of this composite negative outcome might not be optimal for other outcomes such as symptom response or remission.

## Predictors

Potential predictors were extracted from the VHA Corporate Data Warehouse [36] (Supplementary Table 1) and a geospatial social determinants of health (SDoH) database on characteristics of patient residential neighborhoods (Census Tracts, Block Groups), Counties, and States (e.g., economic deprivation, social cohesion) that have been linked in previous research to time-space variation in indicators of mental disorders [39] (Supplementary Table 2). The 5865 variables in this combined database operationalized four broad classes of such prescriptive predictors (see Supplementary Literature Review): psychopathological risk factors (i.e., history of treated mental disorders, treatment types, suicidality), comorbid physical disorders and treatments (including medications suspected to increase self-harm risk; Supplementary Table 3), SDoH (at both the individual and geospatial levels), and facility-level quality indicators [9, 12–15]. Missingness, which occurred only for geospatial variables, was addressed by using nearest-neighbor imputations to fill in missing scores by assigning non-missing values from contiguous areas.

## Analysis methods

*Estimating aggregate negative outcomes.* The sample was split into a 70% training sample and a 30% test sample. We began by estimating an aggregate risk model [40] to predict whether each patient would experience the negative outcome using information from all pretreatment predictors but ignoring type of treatment received. This model was estimated using the Super Learner (SL) stacked generalization machine learning (ML) method [41], an ensemble ML method that pools results across a user-specified range of algorithms (Supplementary Table 4). This was done with the *SuperLearner* R package [42]. See Supplementary Methods for further details. Model fit was evaluated by calculating area under the receiver operating characteristic curve (AU-ROC). We also inspected observed within-ventile and cumulative sensitivity and positive predictive value in the test sample based on predictions from the training sample.

*Estimating average treatment effects.* We then estimated average treatment effects by adjusting for significant differences in the extensive battery of baseline covariates across treatment types. Best-practices methods were used to do this by combining two types of adjustments [43]. (i) The first was a propensity score weighting method to adjust for nonrandom treatment assignment [44] based on the Random Forests (RF) ML method [45]. (ii) The second was an outcome modeling method that predicted probability of the outcome for each patient separately within each treatment arm from baseline covariates, again using RF, and adjusted for differences in multivariate distributions of significant predictors across arms by assigning patients who received each treatment the average scores on these predictors. The predicted outcome scores in the treatment groups generated in these outcome models were then compared to estimate average treatment effect (ATE). As detailed in the Supplementary Methods section, results based on these two methods were combined using *doubly robust methods* to yield consistent results if either method was correct and to reduce finite-sample bias. The specific doubly robust method used was targeted minimum loss-based estimation [46], implemented in the *tmle3* R package [47]. See Supplementary Methods for further details.

*Estimating heterogeneity of treatment effects.* We then estimated between-patient differences in comparative treatment effects, referred to below as Heterogeneity of Treatment Effects (HTE), using Generalized Random Forest (GRF), a doubly robust ML approach that expands on RF with a focus on individual differences in comparative treatment effects with respect to observed baseline variables while adjusting for confounding due to these variables [48, 49]. Analyses were implemented in the *grf* R package [50]. See Supplementary Methods for further details. In broad outline, the analysis entailed developing a model using a two-step doubly robust method. The first step estimated an expected outcome for each patient for each of the three treatment types by estimating treatment-specific models in the subsamples of patients who received each treatment and then imputing predicted treatment-specific outcomes based on those models to all patients regardless of type of treatment received. The second step used these first-stage within-patient estimates to create within-patient scores for differences in expected outcomes across treatments and then used those difference scores as outcomes in a second set of RF models. Importantly, the latter models directly fit interaction

terms (i.e., differences in expected outcomes across treatments within patients) without requiring the correct specification of main effects, simplifying the task of estimating HTE. The ITR for a specific patient was defined as the treatment associated with the lowest predicted probability of the negative outcome for that patient. See Supplementary Methods for further details.

The ITR was then evaluated in the test sample by: (i) dividing the test sample into three subgroups depending on which treatment type was estimated by the ITR in the training sample to have the lowest probability of the negative outcome; and then (ii) using *tmle3* to estimate ATE within each of these subgroups in the test sample. If the ITR improves on a non-individualized treatment strategy, we would expect predicted probability of the negative outcome to be significantly lower for the treatment type estimated to be optimal. Importantly, only information in the test sample was used to evaluate the ITR, whereas only information in the training sample was used to estimate the ITR.

**Predictor importance.** Predictor importance was examined using the kernel Shapley Additive Explanations (SHAP) method [51] implemented in the *fastshap R* package [52]. This method estimates the effect of changing a predictor from its observed value to the sample mean separately averaged across all logically possible permutations of other predictors. More important predictors are associated with higher mean absolute SHAP values. Proportional mean absolute SHAP values ( $SHAP_p$ ) were calculated by dividing mean absolute SHAP values of classes and important predictors within classes by the mean absolute SHAP value for the entire model. Bee swarm plots were used to identify dominant directions and distributions of associations. See Supplementary Methods for further details.

### Reporting

We followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines [53] to report our analyses intended to build predictive models.

## RESULTS

### Sample attributes

Most patients received either ADM-only (46.9%) or psychotherapy-only (39.7%). The remaining 13.4% received combined ADM-psychotherapy. 25.6–33.1% of patients across treatment types were 60+ years of age, 17.0–20.6% were 50–59 years old, 23.1–27.9% were 35–49 years old, and 23.1–29.2% were younger than 35 years old. The great majority (82.0–83.5%) of patients were male. 44.3–46.7% were married. Over half (50.5–55.1%) lived in the South and the great majority (87.5–88.4%) lived in Major Metropolitan Areas (Table 1).

### Aggregate prediction of negative outcomes

8.6% (S.E. = 0.2%) of patients experienced the negative outcome. The aggregate SL risk model estimated in the training sample predicted this outcome in the test sample with AU-ROC = 0.68 (S.E. = 0.01) using an optimal weighting scheme across classifiers (Supplementary Table 5). Observed SN was significantly higher than the 5.0% expected by chance in the top 3 ventiles (18.9–7.0%) in the test sample, with 35.2% of all patients having the negative outcomes falling into this top 15% of the sample (Table 2), PPV in the top ventile of 32.6% (S.E. = 1.8), and cumulative PPV across the top 3 ventiles of 20.2% (S.E. = 0.9) compared to 6.6% (S.E. = 0.5) across the remaining 17 ventiles. Observed sensitivities were significantly lower than chance, in comparison, in the bottom 9 ventiles of the test sample (3.4–1.2%), with only 23.3% of all patients with the negative outcome falling into this 45% of the sample. Cumulative PPV in this bottom 45% of the sample was 4.5%.

### Important predictors of aggregate negative outcomes

The most important classes of predictors of aggregate negative outcome risk were physical disorders ( $SHAP_p = 53.3\%$ ), patient-level SDoH ( $SHAP_p = 35.6\%$ ), and number of visits for mental and/or substance disorders ( $SHAP_p = 22.9\%$ ) (Fig. 1). Strikingly, only substance disorders featured among the most important individual predictors in the last of these categories. Predictors involving

**Table 1.** Selected baseline characteristics of patients by treatment group in the total sample.

	Treatment group					
	ADM-only		Psychotherapy-only		Combined	
	%	(S.E.)	%	(S.E.)	%	(S.E.)
<b>Age</b>						
19–34	29.2	(0.3)	23.1	(0.3)	28.9	(0.6)
35–49	25.7	(0.3)	23.1	(0.3)	27.9	(0.6)
50–59	17.0	(0.3)	20.6	(0.3)	17.6	(0.5)
60+	28.1	(0.3)	33.1	(0.4)	25.6	(0.6)
<b>Sex</b>						
Female	18.0	(0.5)	16.5	(0.3)	18.0	(0.3)
Male	82.0	(0.5)	83.5	(0.3)	82.0	(0.3)
<b>Race/ethnicity</b>						
Hispanic	11.1	(0.2)	9.6	(0.2)	10.4	(0.4)
Non-Hispanic Black	25.4	(0.3)	29.4	(0.3)	27.1	(0.6)
Non-Hispanic White	55.6	(0.3)	53.3	(0.4)	54.7	(0.7)
Other	7.9	(0.2)	7.7	(0.2)	7.9	(0.4)
<b>Marital status</b>						
Married	46.7	(0.3)	46.3	(0.4)	44.3	(0.7)
Divorced	24.3	(0.3)	24.9	(0.3)	25.9	(0.6)
Separated	5.4	(0.2)	5.0	(0.2)	4.9	(0.3)
Widowed	2.5	(0.1)	2.8	(0.1)	2.4	(0.2)
Never married	21.1	(0.3)	20.9	(0.3)	22.5	(0.5)
<b>Census region</b>						
Northeast	9.7	(0.2)	10.6	(0.2)	9.8	(0.4)
Midwest	17.6	(0.3)	16.7	(0.3)	15.8	(0.5)
South	50.4	(0.4)	50.0	(0.4)	55.1	(0.7)
West	22.2	(0.3)	22.6	(0.3)	19.3	(0.5)
<b>Urbanicity</b>						
Major metro	88.0	(0.2)	88.4	(0.2)	87.5	(0.4)
Urban	4.9	(0.2)	5.0	(0.2)	5.5	(0.3)
Rural	7.1	(0.2)	6.6	(0.2)	7.0	(0.3)
(n)	(20,392)		(17,264)		(5814)	

ADM antidepressant medication, Combined, both ADM and psychotherapy, S.E. standard error.

treatments of mental and/or substance use disorders were notable for their low importance ( $SHAP_p = 5.6\%$ ). There were only four individual predictors with  $SHAP_p$  values of at least 5%, all associated with increased probability of the negative outcome. Three of these involved housing problems—administrative evidence of having been homeless in the past 5 years ( $SHAP_p = 6.6\%$ ) and having had housing/economic problems in the past 5 years ( $SHAP_p = 6.1\%$ ) and count of number of VA housing problem visits in the past 5 years ( $SHAP_p = 5.3\%$ ). The fourth was a measure of high physical comorbidity ( $SHAP_p = 6.0\%$ ).

### Average treatment effects

Average treatment-specific probability of the outcome adjusting for observed differences in baseline variables did not differ

**Table 2.** Observed prevalence of the negative outcome in the test sample ( $n = 12,986$  by ventiles of predicted probability of the negative outcome based on the SL analysis carried out in the training sample ( $n = 30,484$ )<sup>a</sup>.

Predicted risk distribution <sup>b</sup>	Sensitivity		Cumulative sensitivity		Positive predictive value		Cumulative positive predictive value	
	%	(S.E.)	%	(S.E.)	%	(S.E.)	%	(S.E.)
20 (Highest risk)	18.9	(1.2)	18.9	(1.2)	32.6	(1.8)	32.6	(1.8)
19	9.3	(0.9)	28.2	(1.3)	16.0	(1.4)	24.3	(1.2)
18	7.0	(0.8)	35.2	(1.4)	12.0	(1.3)	20.2	(0.9)
17	5.8	(0.7)	41.0	(1.5)	10.0	(1.2)	17.7	(0.7)
16	6.2	(0.7)	47.1	(1.5)	10.6	(1.2)	16.3	(0.6)
15	6.2	(0.7)	53.3	(1.5)	10.6	(1.2)	15.3	(0.6)
14	5.3	(0.7)	58.6	(1.5)	9.1	(1.1)	14.4	(0.5)
13	4.6	(0.6)	63.2	(1.4)	8.0	(1.1)	13.6	(0.5)
12	4.7	(0.6)	67.9	(1.4)	8.2	(1.1)	13.0	(0.4)
11	4.6	(0.6)	72.6	(1.3)	8.0	(1.1)	12.5	(0.4)
10	4.1	(0.6)	76.7	(1.3)	7.1	(1.0)	12.0	(0.4)
9	3.4	(0.5)	80.1	(1.2)	5.9	(0.9)	11.5	(0.4)
8	2.9	(0.5)	83.0	(1.1)	5.1	(0.9)	11.0	(0.3)
7	3.6	(0.6)	86.6	(1.0)	6.2	(0.9)	10.7	(0.3)
6	3.7	(0.6)	90.3	(0.9)	6.3	(1.0)	10.4	(0.3)
5	2.1	(0.4)	92.4	(0.8)	3.7	(0.7)	10.0	(0.3)
4	2.1	(0.4)	94.5	(0.7)	3.5	(0.7)	9.6	(0.3)
3	2.3	(0.5)	96.8	(0.5)	4.0	(0.8)	9.3	(0.3)
2	2.0	(0.4)	98.8	(0.3)	3.4	(0.7)	9.0	(0.3)
1 (Lowest risk)	1.2	(0.3)	100.0	(0.0)	2.2	(0.6)	8.6	(0.2)

Sensitivity the proportion of patients with a negative outcome who were in the ventile indicated by the row heading, Cumulative sensitivity the cumulation of sensitivities from the ventile with highest risk through the ventile indicated by the row heading Positive predictive value, the proportion of patients in the ventile indicated by the row heading who had a negative outcome; Cumulative positive predictive value, the average positive predictive value from the ventile with highest risk through the ventile indicated by the row heading; %, observed values of sensitivity and positive predictive value in the test sample; S.E., standard error of %.

SL Super Learner.

<sup>a</sup>Ventiles are subgroups, each made up of 5% of patients in the test sample, ordered by their predicted probabilities of a negative outcome based on the model estimated in the training sample and applied to the test sample. The area under the receiver operating characteristic curve (AU-ROC) in the test sample of the prediction model is 0.68 (S.E. = 0.01). The negative outcome was defined as a composite measure of any of the following serious negative events in the 365 days after initiating treatment: psychiatric emergency department visit, psychiatric urgent care visit, psychiatric hospitalization, administratively recorded suicide attempt, or suicide death.

<sup>b</sup>Defined by ventiles of predicted probability of the negative outcome in the test sample.

significantly across the three treatment types: 9.1% (S.E. = 0.3) for ADM-only; 8.5% (S.E. = 0.3) for psychotherapy-only; and 8.8% (S.E. = 0.4) for combined ADM-psychotherapy ( $\chi^2_2 = 2.0$ ,  $p = 0.36$ ) (Table 3).

### Heterogeneity of treatment effects

ADM-only was estimated by the ITR to be optimal with respect to low probability of the negative outcome for 8.5% of patients, psychotherapy-only for another 55.0%, and combined treatment for the remaining 36.5%. Patients estimated by the ITR to be optimized by a given treatment were only slightly more likely than others to receive that treatment (Supplementary Table 6). However, estimated ATE in the test sample differed significantly by treatment type received in the subgroup predicted by the ITR to be optimized with psychotherapy-only ( $\chi^2_2 = 7.2$ ,  $p = 0.028$ ). As predicted by the ITR, probability of the negative outcome in this subgroup was significantly lower among the patients treated with psychotherapy-only (6.9%, S.E. = 0.5) than those treated with either ADM-only (8.4% S.E. = 0.5,  $\chi^2_1 = 5.2$ ,  $p = 0.023$ ) or combined ADM-psychotherapy (9.0%, S.E. = 0.9,  $\chi^2_1 = 4.3$ ,  $p = 0.038$ ), suggesting that the proportion of patients with this outcome is proportionally about 20% lower when treated with

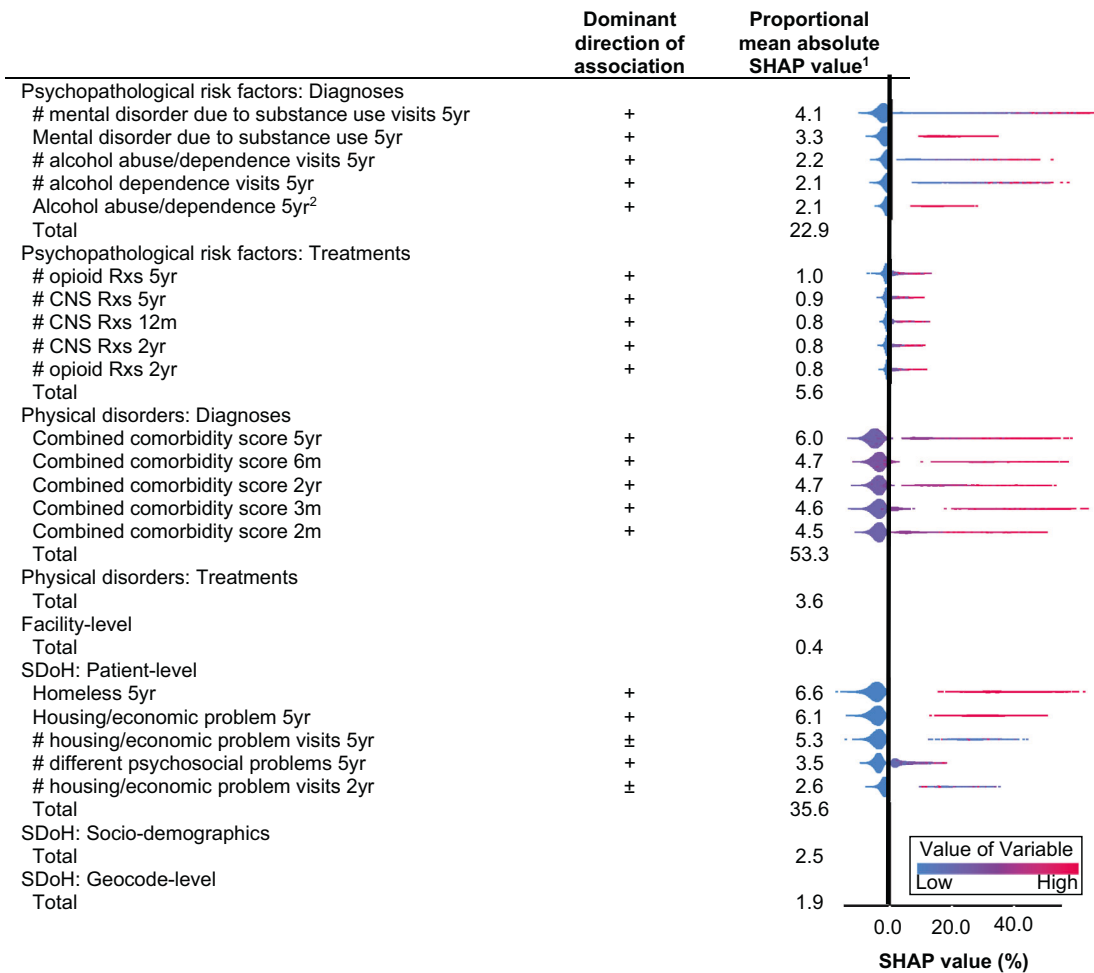
psychotherapy-only than one of the other treatments. In comparison, estimated ATE in the test sample did not differ significantly by treatment received either in the subgroup predicted by the ITR to be optimized with ADM-only ( $\chi^2_2 = 0.9$ ,  $p = 0.63$ ) or with combined ADM-psychotherapy ( $\chi^2_2 = 0.4$ ,  $p = 0.81$ ) (Table 4).

### Important predictors of heterogeneity of treatment effects

Given that the only significant HTE involved optimization with psychotherapy-only, we focused on that outcome in considering predictor importance (Fig. 2). By far the most important class of predictors of being optimized with psychotherapy-only was geocode-level SDoH (SHAP<sub>p</sub> = 93.5%). Physical disorders were the only other class of predictors with a SHAP<sub>p</sub> in double digits (13.7%). All individual predictors with SHAP<sub>p</sub> values of at least 5% were geocode-level SDoH variables, the most important of them involving low county-level access to medical treatment.

### The implications of using the ITR for treatment assignment

The test sample results suggest that all the 56.0% of patients who were predicted by the ITR to be optimized by psychotherapy-only should receive psychotherapy-only if the goal of treatment is to



**Fig. 1 Predictors of high negative outcome risk in the test sample.** SHAP Shapley Additive exPlanations, # Number of, 5 yr 5 years before outpatient treatment visit, Rx prescription, CNS central nervous system, 12 m 12 months before outpatient treatment visit, 2 yr 2 years before outpatient treatment visit, 6 m 6 months before outpatient treatment visit, 3 m 3 months before outpatient treatment visit, 2 m 2 months before outpatient treatment visit, SDoH social determinants of health. <sup>1</sup>See Supplementary Methods for discussion of how to interpret SHAP. <sup>2</sup>Visits with alcohol abuse uncomplicated, with intoxication or alcohol-induced disorders, any dependence, or unspecified use with intoxication or alcohol-induced disorders as the primary diagnosis as indicated by a selected set of International Classification of Diseases, Ninth/Tenth Revision, Clinical Modification (ICD-9/10-CM) codes.

minimize risk of the negative outcome considered here, as estimated prevalence of that outcome was proportionally about 20% lower (6.9%) when these patients were treated optimally than otherwise (8.6–9.0%). However, the test sample results suggest that the remaining 44.0% of patients would have equivalent prevalence of the negative outcome no matter which treatment they received. If treatment assignment was made according to these results, 56% of patients would be treated with psychotherapy-only and 44% with ADM-only (i.e., the least expensive and most readily available treatment). This compares with 53.1% of patients who were observed to receive psychotherapy (39.7% psychotherapy-only and 13.4% combined ADM-psychotherapy) and 60.3% observed to receive ADM (46.9% ADM-only and 13.4% combined ADM-psychotherapy), for a total increase of 2.9% in the number of patients who would receive psychotherapy and a total decrease of 16.1% in the number of patients who would receive ADM under the ITR compared to current practice. If the test sample conditional ATE results reflect causal effects, these changes in treatment assignment would be expected to result in a 7.7% proportion decrease in prevalence of the negative outcome across all patients (from 8.6% to 8.0%).

**DISCUSSION**

**Prevalence and prediction of aggregate negative outcomes**

We are aware of no previous research that attempted to use administrative data to predict serious negative events of the sort considered here among patients in outpatient MDD treatment. However, previous research has documented that administrative data can predict two components in our outcome, suicide attempts and suicide deaths after outpatient visits in both civilian (e.g., [54]) and military (e.g., [55]) samples. The PPV of the negative outcome in the highest risk ventiles for our aggregate model was sufficiently high (32.6% predicted probability of the negative outcome among the 5% of patients with highest predicted risk) that some clinicians might think of these patients as likely treatment-resistant. If so, both practice guidelines and health-economic analyses would support starting with more intensive ADM/psychotherapy than is standard, such as aggressive medication dosing, more frequent psychotherapy scheduling, or more invasive treatments (e.g., antipsychotic augmentation, electroconvulsive therapy, transcranial magnetic stimulation, or ketamine [56–59]). It might be that, relative to current practice, one of these approaches would reduce futile trial and error efforts to find less intensive effective treatments for these cases.

Which of these or other more intensive courses of action would be preferable among patients at high risk of a serious negative event of the sort considered here cannot be determined from our results, as our models were not designed to identify optimal alternative treatments beyond the three considered here. There would certainly be barriers to implementing any more intensive courses of action as first-line treatments, not least of which in the civilian healthcare sector would be a payment structure that often

**Table 3.** Estimated average treatment effect (ATE) for each treatment type in total training sample ( $n = 30,484$ )<sup>a</sup>.

	Total sample	
	%	(S.E.)
I. Treatment-specific probability of the outcome		
ADM-only	9.1	(0.3)
Psychotherapy-only	8.5	(0.3)
Combined	8.8	(0.4)
II. Average treatment effect (ATE)		
ADM-only vs. Psychotherapy-only	0.5	(0.4)
ADM-only vs. Combined	0.2	(0.5)
Psychotherapy-only vs. Combined	-0.3	(0.5)
$\chi^2_2$	2.0	

% estimated probability of experiencing a the negative outcome among patients receiving each type of treatment or differences in these probabilities after adjusting for differences in observed baseline characteristics by treatment type, *S.E.* standard error of %, *ADM* antidepressant medication, Combined, both ADM and psychotherapy.

<sup>a</sup>The negative outcome was defined as a composite measure of any of the following serious negative events in the 365 days after initiating treatment: psychiatric emergency department visit, psychiatric urgent care visit, psychiatric hospitalization, administratively recorded suicide attempt, or suicide death. The individualized treatment rule (ITR) was developed in the training sample using the causal forests algorithm with the *grf* R package. The ITR was evaluated in the test sample using targeted minimum loss-based estimation with the *tmle3* R package. See Supplementary Methods for further details.

requires several failed trials of first-line treatments before more intensive interventions are reimbursed [60]. But there is some indication that going directly to more intensive treatments could be acceptable to payers as an extension of existing stepped-care approaches if this was demonstrated to be cost-effective [61].

### Aggregate treatment effects

Our finding of nonsignificant aggregate differences in prevalence of the composite negative outcome across treatment types is consistent with previous MDD randomized controlled trials that evaluated comparative effectiveness of ADM-only, psychotherapy-only, and combined ADM-psychotherapy in reducing psychiatric hospitalizations [62, 63], psychiatric emergency department visits [64], and suicide deaths [65, 66]. However, as noted in the introduction, a recent meta-analysis of these studies found that psychotherapy-only is associated with significantly lower risk than either ADM-only or combined ADM-psychotherapy of a composite serious negative outcome of the sort we considered here [5]. Our sample size was large enough to detect differences of the magnitude found in the meta-analysis, but the estimated ATEs in our sample, although consistent in sign with those of the meta-analysis (i.e., psychotherapy-only associated with lowest risk followed by combined treatment and ADM-only), were nonsignificant. This suggests either that ATEs are weaker in the VHA than in the more general population samples included in the meta-analysis and/or that residual bias existed in our nonexperimental analysis that under-estimated the benefit of psychotherapy-only. A pragmatic trial would be required to adjudicate between these two contending possibilities.

### Heterogeneity of treatment effects

We found significant HTE for the 56% of the sample that was estimated to be optimized by psychotherapy-only, with the estimated benefit of optimal versus suboptimal treatment assignment associated with a roughly 20% proportional reduction in risk of the negative outcome. An effect of this magnitude would be considered clinically significant but small, raising the question of whether this level of HTE, even if it could be confirmed in a pragmatic trial, would be sufficiently large to warrant carrying out such a trial. The alternative, given the results of the recent

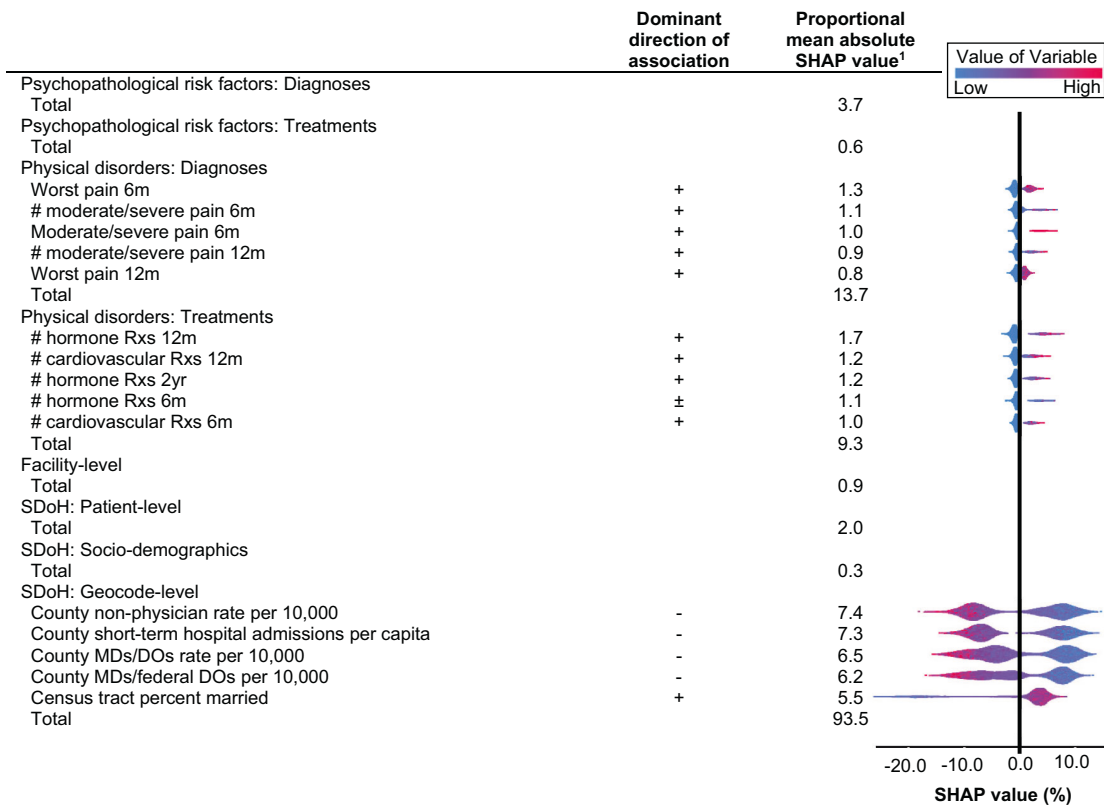
**Table 4.** Estimated average treatment effect (ATE) for each treatment type in subsamples of the test sample estimated by the individualized treatment rule (ITR) be optimized by the different treatment types ( $n = 12,986$ )<sup>a</sup>.

I. Among patients predicted to be optimized by...	ADM-only		Psychotherapy-only		Combined ADM-psychotherapy	
	%	(S.E.)	%	(S.E.)	%	(S.E.)
A. Treatment-specific probability of the outcome						
ADM-only	7.1	(1.1)	8.4	(0.5)	9.9	(0.7)
Psychotherapy-only	8.7	(1.4)	6.9	(0.5)	9.9	(0.7)
Combined	8.8	(2.3)	9.0	(0.9)	10.7	(1.2)
B. Average treatment effect (ATE)						
ADM-only vs. Psychotherapy-only	-1.5	(1.8)	1.6	(0.7)	0.1	(1.0)
ADM-only vs. Combined	-1.7	(2.6)	-0.6	(1.0)	-0.8	(1.4)
Psychotherapy-only vs. Combined	-0.1	(2.7)	-2.2	(1.0)	-0.9	(1.4)
$\chi^2_2$		0.9			7.2 <sup>b</sup>	0.4

% estimated probability of experiencing a negative outcome among patients receiving each type of treatment or differences in these probabilities after adjusting for differences in observed baseline characteristics by treatment type, *S.E.* standard error of %, *ADM* antidepressant medication, Combined, both ADM and psychotherapy.

<sup>a</sup>The negative outcome was defined as a composite measure of any of the following serious negative events in the 365 days after initiating treatment: psychiatric emergency department visit, psychiatric urgent care visit, psychiatric hospitalization, administratively recorded suicide attempt, or suicide death. The ITR was developed in the training sample using the causal forests algorithm with the *grf* R package. The ITR was evaluated in the test sample using targeted minimum loss-based estimation with the *tmle3* R package. See Supplementary Methods for further details.

<sup>b</sup>Significant at the 0.05 level, two-sided test.



**Fig. 2 Predictors of being optimized by psychotherapy only.** SHAP Shapley Additive exPlanations, 6 m, 6 months before outpatient treatment visit; # Number of, 12 m 12 months before outpatient treatment visit, Rx prescription, 2 yr 2 years before outpatient treatment visit, SDoH social determinants of health, MD doctor of medicine, DO doctor of osteopathic medicine. <sup>1</sup>See Supplementary Methods for discussion of how to interpret SHAP.

meta-analysis cited above, might be to recommend psychotherapy-only in VHA PC-MHI clinics whenever concerns exist about any of the serious negative events considered here. However, even in PC-MHI clinics MDD patients often wait up to 8 weeks before being seen by a psychotherapist. Such patients are routinely prescribed ADMs at the initial PC-MHI visit in which they are referred to psychotherapy. PC-MHI psychotherapy-only patients, in comparison, are those who are either able to see a psychotherapist at the time of their initial visit or who get a near-term psychotherapy appointment at the time of this initial visit. Given that this quick access to psychotherapy is not always possible, our ITR might be useful in PC-MHI settings when triage decisions are needed about which patients to prioritize for near-term psychotherapy.

Our ITR might be more important in VHA primary care settings that do not qualify as PC-MHI sites (i.e., do not have at least one full-time psychotherapist on staff), where a lower proportion of MDD patients have access to psychotherapy, and even more so in clinics not associated with VHA, where only a small minority of MDD patients receive psychotherapy [3]. In both cases, though, new ITRs should be developed and evaluated before considering a pragmatic trial. Given the stronger estimates of ATE in the recent meta-analysis of controlled trials than in our observational analysis, it might well be that HTE would be stronger in these other settings than in VHA PC-MHI clinics.

It is also important to repeat a point made earlier: that HTE with respect to more conventional measures of MDD treatment response and remission might be quite different than for the serious negative events considered in the current analysis. Meta-analysis of controlled trials found stronger HTE for psychotherapy-only, ADM-only, and combined treatment with respect to these more typical outcome measures [4] than for the serious negative events considered in our study [5]. The possibility that the same

might be true for HTE adds to the argument in favor of carrying out such a trial.

**Predictor importance**

It is hazardous to place too much emphasis on predictor importance either in modeling aggregate risk of negative outcomes or HTE, as the ML methods used to train these models are designed to optimize overall model prediction accuracy at the expense of the accuracy of individual predictors [67]. This means that the predictors highlighted as important are really markers of predicting associations involving the (sometimes many) other baseline variables that are correlated significantly with the predictors designated important. Furthermore, important predictors cannot be assumed to be important causes, as markers of unmeasured causal factors often emerge as important predictors. Within the context of these cautions, it is noteworthy that physical disorders and patient-level indicators of SDoH emerged as more important than psychopathological factors as predictors of aggregate risk, whereas geocode-level indicators of SDoH were by far the most important predictors of HTE involving the comparative benefits of psychotherapy-only. It is unclear whether the same predictors would be important in predicting MDD symptom response or remission in VHA, or in predicting the serious negative outcomes considered here in other treatment settings.

**Strengths and limitations**

This study had several noteworthy strengths, including a large representative sample of patients, an extensive array set of baseline variables related to constructs found by prior studies to predict MDD HTE, and the use of a best-practices approach to estimate the ITR. All of these have been identified as important weaknesses in prior ITR studies of psychiatric disorders [68].

But the study also had some noteworthy limitations. First, and most importantly, the data were observational rather than experimental. This meant that unmeasured confounding variables could have introduced bias into estimates even though we adjusted for an exhaustive set of baseline covariates and our aggregate results had a similar, albeit weaker, pattern than in prior controlled trials that evaluated the comparative effectiveness of ADM-only, psychotherapy-only, and combined treatment in preventing the negative events that made up our composite outcome.

Second, despite our use of an extensive battery of baseline covariates, the predictor set did not include information about some variables that have been the focus of prior research, including biomarkers [69] and several other potentially important prescriptive predictors [70, 71]. Although the absence of biomarker information might not be seen as a major limitation given that biomarkers have not as yet shown great promise as predictors of MDD HTE, other variables we did not include, most notably patient treatment preferences, are known both to influence the type of treatment received and to predict treatment outcomes [72, 73]. We attempted to develop proxy measures for patient treatment preferences by abstracting information from EHRs that might be indicators of past MDD treatments perceived by the patient as successful (as indicated by adherence over a period of time long enough to be considered a full course of treatment) vs. not successful (as indicated either by a short course of treatment consistent with treatment dropout or by a serious negative outcome, such as a psychiatric emergency department visit or hospitalization, in the course of treatment, suggesting that the treatment failed). However, it would have been better to have direct measures of patient preferences than these proxies.

Third, although we excluded patients previously treated with antipsychotic medications to make sure there were no cases of bipolar disorder or nonaffective psychosis in the sample, this had the effect also of excluding refractive MDD cases treated with novel antipsychotics. It might be that retaining such patients in future extensions and using information about their history of antipsychotic treatment might improve prediction accuracy, increase external validity, and possibly help define a subgroup of patients who are optimized with combined ADM-psychotherapy.

Fourth, information on the types of ADMs and psychotherapies received were not used in the analysis. This was by design given that previous research has failed to document substantial evidence of HTE across ADM types [74]. Information on the effectiveness of psychotherapy is strongest for Cognitive Behavioral Therapy [4] but psychotherapy type was not recorded in the VHA EHR. In addition, information on treatment adherence was not used in the analysis even though VHA records contain information on prescription refills and psychotherapy sessions attended and missed. We did not include these measures, as they are endogenous (i.e., could be a consequence rather than a cause of treatment nonresponse). Information on ADM dosage was not extracted from the EHR.

Fifth, our use of a composite outcome of serious negative events limited generalizability. We know from previous randomized controlled trials that the comparative treatment effects of ADM-only, psychotherapy-only, and combined ADM-psychotherapy are stronger for symptom response-remission [4] and functioning [75] than for the serious negative events that made up our composite outcome measure [5]. It might be that something along the same lines holds for HTE; that is, that HTE is stronger with respect to symptom response-remission and functioning than with respect to the serious negative events we considered. If so, our ITR for psychotherapy-only might be a lower-bound estimate.

Sixth, caution should be used in extrapolating results to all potential VHA patients, as MDD treatment is less available in VHA

clinics that do not qualify for a designation as PC-MHI and in rural areas. In addition, some depressed Veterans are more reluctant than others to seek treatment. Even greater caution should also be used in extrapolating results outside of VHA given the unique socio-demographic characteristics of Veterans and the fact that the VHA system provides relatively high access to affordable and quality care, including much higher access to psychotherapy, than in the civilian healthcare system.

### Future directions

As noted in the introduction, the next logical step in evaluating the ITR developed here would be to implement a pragmatic trial in which the ITR is used to help provide clinical decision support for treatment selection to some, but not all, VHA patients seeking treatment for MDD. Replication and expansion of the analysis carried out here in that presumably large experimental sample could then be used to refine the ITR by including easily collected baseline and follow-up measures of patient self-reported symptom response. It is noteworthy in this regard that such a pragmatic trial could be carried out at low cost merely by using administrative data of the sort we used here to define the subset of patients hypothesized to be optimized by psychotherapy-only and to randomize access to this information as a clinical decision support tool.

If the results reported here are not judged to be sufficiently strong to warrant such a trial, then they might be strong enough to justify more intensive observational analyses that address the limitations noted above in our study. The most plausible way to do this would be by using instrumental variable methods. The key to this approach would be to identify one or more baseline variables (instrumental variables; IVs) that could plausibly be thought of as influencing treatment assignment but not influencing treatment outcomes other than through treatment assignment [76]. This approach is becoming increasingly common in nonexperimental studies of treatment effects using EHRs based on two general classes of plausible IVs. One potentially useful class of IVs involves geographic information [77]. These are typically used in studies of the effects of innovative treatments that are at first available only in limited geographic areas (so long as the patients in the trial are restricted to those residing close to the treatment setting in which they obtain treatment to avoid bias due to selection into innovative treatments by early adopters from other areas). This type of instrument might be expected to be of little value in VHA because all VHA medical centers should adopt the PC-MHI model over the next few years [78]. However, large variations continue to exist across VHA centers in PC-MHI implementation, resulting in the proportion of VHA patients seeking MDD treatment obtaining psychotherapy to vary widely across centers in ways that could be treated as a valid IV [71].

A second class of IVs involves clinician preference [79]. These IVs can be constructed from administrative databases to describe the past treatment patterns of the provider treating each patient. These past treatment patterns are typically strong predictors of patient treatment assignment and might plausibly be thought to influence outcomes only through treatment assignment. This type of IV is most useful when substantial variation exists across providers in the types of preferred treatments. Previous research has shown that this variation exists across prescribers in types of ADMs prescribed and that these differences are strong predictors of subsequent prescribing patterns [80]. Our preliminary analysis of VHA data shows that the same is true for past psychotherapy referral patterns of primary care physicians.

It is important to note that IV methods are highly sensitive to misspecification. Formal tests consequently should be used to evaluate the validity of IVs before using them [81, 82]. It is also noteworthy that statistical methods exist to estimate HTE in the context of IV models [83]. However, measures of prescriptive predictors remain important for developing strong ITRs even



when valid IVs are present. The section on limitations noted several important prescriptive predictors that we could not measure in our study, including baseline MDD symptom severity (which, as noted above, could also be used as an additional outcome), patient treatment preferences, and psychotherapy type. As measurement-based care becomes more common in large health systems like VHA, it will become more and more possible to carry out studies like the current one with patient-reported symptom measures used both as prescriptive predictors and outcomes. Even before this time, though, natural language processing (NLP) methods have shown considerable promise in extracting information from clinical notes that allow PHQ-9 scores to be approximated [84, 85]. The same methods have been used to extract information from clinical notes about SDoH as prescriptive predictors [86]. Although we are aware of no comparable studies designed to elicit information about patient treatment preferences, psychotherapy type, or other potentially important prescriptive predictors of MDD HTE, dramatic recent successes in large language models on various NLP tasks suggest that it might be useful to attempt to develop such models [87]. VHA would be an ideal site for such studies given that the VA Informatics and Computing and Infrastructure system has created a consolidated free test database to facilitate large-scale NLP analyses [88].

## CONCLUSIONS

We were able to define a subgroup of patients with high risk of the composite negative outcome considered here based on information available in administrative records at the time of beginning treatment. We also found evidence for substantial MDD HTE with respect to the benefit of psychotherapy-only. Finally, we found no evidence for the benefit of combined ADM-psychotherapy in any subgroup of patients, although it is noteworthy that we under-represented patients who were previously refractive by excluding from analysis those with a history of treatment with novel antipsychotics. This information could be useful to target more intensive treatments to high-risk patients and to increase the match between the type of first-line treatment received and the treatment likely to be optimal for remaining patients. Replication would be needed before implementing such changes, though, both to confirm the stability of the aggregate risk model and to determine if the same predictors can be used to predict aggregate variation in risk of more common outcomes involving MDD symptom response-remission. NLP methods applied to clinical notes could also be used both to generate a proxy MDD symptom response-remission outcome measure and to obtain estimates of baseline prescriptive predictors that could refine ITR development in conjunction with the use of IV methods to address the problem of nonrandom treatment assignment. Supplementary information is available on MP's website.

## DATA AVAILABILITY

No. Data were obtained through a nontransferable data use agreement with the Veterans Health Administration.

## REFERENCES

1. GBD 2019 Diseases and Injuries Collaborators. 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. *Lancet*. 2020;396:P1204–22.
2. Herrman H, Patel V, Kieling C, Berk M, Buchweitz C, Cuijpers P, et al. Time for united action on depression: a lancet-world psychiatric association commission. *Lancet*. 2022;399:957–1022.
3. Olfson M, Blanco C, Marcus SC. Treatment of adult depression in the United States. *JAMA Intern Med*. 2016;176:1482–91.
4. Cuijpers P, Miguel C, Harrer M, Plessen CY, Ciharova M, Ebert D, et al. Cognitive behavior therapy vs. control conditions, other psychotherapies, pharmacotherapies and combined treatment for depression: A comprehensive meta-

analysis including 409 trials with 52,702 patients. *World Psychiatry*. 2023;22:105–15.

5. Zainal NH. Is combined antidepressant medication (ADM) and psychotherapy better than either monotherapy at preventing suicide attempts and other psychiatric serious adverse events for depressed patients? A rare events meta-analysis. *Psychol Med*. Online ahead of print 15 November 2023. <https://doi.org/10.1017/s0033291723003306>.
6. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. *J Clin Psychiatry*. 2013;74:595–602.
7. Goodwin RD, Dierker LC, Wu M, Galea S, Hoven CW, Weinberger AH. Trends in U.S. depression prevalence from 2015 to 2020: the widening treatment gap. *Am J Prev Med*. 2022;63:726–33.
8. Ross EL, Vijan S, Miller EM, Valenstein M, Zivin K. The cost-effectiveness of cognitive behavioral therapy versus second-generation antidepressants for initial treatment of major depressive disorder in the United States: a decision analytic model. *Ann Intern Med*. 2019;171:785–95.
9. Cohen ZD, DeRubeis RJ. Treatment selection in depression. *Annu Rev Clin Psychol*. 2018;14:209–36.
10. Driessen E, Dekker JJM, Peen J, Van HL, Maina G, Rosso G, et al. The efficacy of adding short-term psychodynamic psychotherapy to antidepressants in the treatment of depression: a systematic review and meta-analysis of individual participant data. *Clin Psychol Rev*. 2020;80:101886.
11. Driessen E, Fokkema M, Dekker JJM, Peen J, Van HL, Maina G, et al. Which patients benefit from adding short-term psychodynamic psychotherapy to antidepressants in the treatment of depression? A systematic review and meta-analysis of individual participant data. *Psychol Med*. 2023;53:6090–101.
12. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Ebert DD, et al. Using patient self-reports to study heterogeneity of treatment effects in major depressive disorder. *Epidemiol Psychiatr Sci*. 2017;26:22–36.
13. Kraus C, Kadriu B, Lanzenberger R, Zarate CA Jr, Kasper S. Prognosis and improved outcomes in major depression: a review. *Transl Psychiatry*. 2019;9:127.
14. Maj M, Stein DJ, Parker G, Zimmerman M, Fava GA, De Hert M, et al. The clinical characterization of the adult patient with depression aimed at personalization of management. *World Psychiatry*. 2020;19:269–93.
15. Perna G, Alciati A, Daccò S, Grassi M, Caldirola D. Personalized psychiatry and depression: the role of sociodemographic and clinical variables. *Psychiatry Investig*. 2020;17:193–206.
16. Cohen ZD, DeRubeis RJ, Hayes R, Watkins ER, Lewis G, Byng R, et al. The development and internal evaluation of a predictive model to identify for whom Mindfulness-Based Cognitive Therapy (MBCT) offers superior relapse prevention for recurrent depression versus maintenance antidepressant medication. *Clin Psychol Sci*. 2023;11:59–76.
17. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The personalized advantage index: translating research on prediction into individualized treatment recommendations. A demonstration. *PLoS One*. 2014;9:e83875.
18. Elkin I, Shea MT, Watkins JT, Imber SD, Sotsky SM, Collins JF, et al. National Institute of Mental Health treatment of depression collaborative research program. General effectiveness of treatments. *Arch Gen Psychiatry*. 1989;46:971–82.
19. Vittengl JR, Clark AL, Thase ME, Jarrett RB. Initial Steps to inform selection of continuation cognitive therapy or fluoxetine for higher risk responders to cognitive therapy for recurrent major depressive disorder. *Psychiatry Res*. 2017;253:174–81.
20. Wallace ML, Frank E, Kraemer HC. A novel approach for developing and interpreting treatment moderator profiles in randomized clinical trials. *JAMA Psychiatry*. 2013;70:1241–7.
21. Huibers MJ, van Breukelen G, Roelofs J, Hollon SD, Markowitz JC, van Os J, et al. Predicting response to cognitive therapy and interpersonal therapy, with or without antidepressant medication, for major depression: a pragmatic trial in routine practice. *J Affect Disord*. 2014;152:154:146–54.
22. Qiu X, Wang Y. Composite interaction tree for simultaneous learning of optimal individualized treatment rules and subgroups. *Stat Med*. 2019;38:2632–51.
23. Gunter L, Zhu J, Murphy SA. Variable selection for qualitative interactions. *Stat Methodol*. 2011;1:42–55.
24. Laber EB, Zhao YQ. Tree-based methods for individualized treatment regimes. *Biometrika*. 2015;102:501–14.
25. Lorenzo-Luaces L, DeRubeis RJ, van Straten A, Tiemens B. A prognostic index (PI) as a moderator of outcomes in the treatment of depression: a proof of concept combining multiple variables to inform risk-stratified stepped care models. *J Affect Disord*. 2017;213:78–85.
26. Lorenzo-Luaces L, Rodriguez-Quintana N, Riley TN, Weisz JR. A placebo prognostic index (PI) as a moderator of outcomes in the treatment of adolescent depression: could it inform risk-stratification in treatment with cognitive-behavioral therapy, fluoxetine, or their combination? *Psychother Res*. 2021;31:5–18.

27. Nemeroff CB, Heim CM, Thase ME, Klein DN, Rush AJ, Schatzberg AF, et al. Differential responses to psychotherapy versus pharmacotherapy in patients with chronic forms of major depression and childhood trauma. *Proc Natl Acad Sci USA*. 2003;100:14293–6.
28. Song R, Kosorok M, Zeng D, Zhao Y, Laber E, Yuan M. On sparse representation for optimal individualized treatment selection with penalized outcome weighted learning. *Stat*. 2015;4:59–68.
29. Zhao Y, Zeng D, Rush AJ, Kosorok MR. Estimating individualized treatment rules using outcome weighted learning. *J Am Stat Assoc*. 2012;107:1106–18.
30. Luedtke A, Sadikova E, Kessler RC. Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. *Clin Psychol Sci*. 2019;7:445–61.
31. VanderWeele TJ, Luedtke AR, van der Laan MJ, Kessler RC. Selecting optimal subgroups for treatment using many covariates. *J Epidemiol*. 2019;30:334–41.
32. Ashrafioun L, Pigeon WR, Conner KR, Leong SH, Oslin DW. Prevalence and correlates of suicidal ideation and suicide attempts among veterans in primary care referred for a mental health evaluation. *J Affect Disord*. 2016;189:344–50.
33. Trivedi RB, Post EP, Sun H, Pomerantz A, Saxon AJ, Piette JD, et al. Prevalence, comorbidity, and prognosis of mental health among US Veterans. *Am J Public Health*. 2015;105:2564–9.
34. Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, et al. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat Biomed Eng*. 2023;7:719–42.
35. Kessler RC, Luedtke A. Pragmatic precision psychiatry: a new direction for optimizing treatment selection. *JAMA Psychiatry*. 2021;78:1384–90.
36. U.S. Department of Veterans Affairs. Corporate Data Warehouse (CDW): U.S. Department of Veterans Affairs; 2023 [updated 2023 Jan 11; cited 2023 Jun 26]. Available from: [https://www.hsrd.research.va.gov/for\\_researchers/cdw.cfm](https://www.hsrd.research.va.gov/for_researchers/cdw.cfm).
37. Hoffmire C, Stephens B, Morley S, Thompson C, Kemp J, Bossarte RM. VA suicide prevention applications network: a national healthcare system-based suicide event tracking system. *Public Health Rep*. 2016;131:816–21.
38. U.S. Centers for Disease Control and Prevention. National Death Index: U.S. Centers for Disease Control and Prevention; 2022 [updated 2022 Jan 10; cited 2023 Jun 26]. Available from: <https://www.cdc.gov/nchs/ndi/index.htm>.
39. Du S, Yao J, Shen GC, Lin B, Udo T, Hastings J, et al. Social drivers of mental health: A U.S. study using machine learning. *Am J Prev Med*. 2023;65:827–34.
40. Kent DM. Overall average treatment effects from clinical trials, one-variable-at-a-time subgroup analyses and predictive approaches to heterogeneous treatment effects: toward a more patient-centered evidence-based medicine. *Clin Trials* 2023;20:328–37.
41. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol*. 2007;6:Article25.
42. Polley E, LeDell E, Kennedy C, Lendle S, van der Laan M. SuperLearner: Super Learner Prediction: The Comprehensive R Archive Network; 2021 [updated 2021 May 10; cited 2023 Jun 26]. Available from: <https://cran.r-project.org/web/packages/SuperLearner/index.html>.
43. Funk MJ, Westreich D, Wiesen C, Stürmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *Am J Epidemiol*. 2011;173:761–7.
44. Desai RJ, Franklin JM. Alternative approaches for confounding adjustment in observational studies using weighting based on the propensity score: a primer for practitioners. *BMJ*. 2019;367:l5657.
45. Breiman L. Random forests. *Mach Learn*. 2001;45:5–32.
46. van der Laan M, Gruber S. Working Paper 290: Targeted minimum loss based estimation of an intervention specific mean outcome: U.C. Berkeley Division of biostatistics working paper series; 2011 [updated 2011 Aug; cited 2023 Jun 26]. Available from: <https://biostats.bepress.com/ucbbiostat/paper290/>.
47. Coyle J tmlc3: The Extensible TMLE framework: tmlc3; 2021 [cited 2023 Jun 26]. Available from: <https://tlverse.org/tmlc3/>.
48. Athey S, Tibshirani R, Wager S. Generalized random forests. *Ann Stat*. 2019;47:1179–203.
49. Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests. *J Am Stat Assoc*. 2018;113:1228–42.
50. Tibshirani J, Athey S, Friedberg R, Hadad V, Hirshberg D, Miner L, et al. Package 'grf': Generalized Random Forests 2022 [updated 2022 Dec 15; cited 2023 Jun 26]. Available from: <https://cran.r-project.org/web/packages/grf/grf.pdf>.
51. Lundberg S, Lee SL. A unified approach to interpreting model predictions. 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA; December 4–9, 2017.
52. Greenwell B. fastshap: Fast Approximate Shapley Values version 0.0.7 2021 [updated 2021 Dec 6; cited 2023 Jun 26]. Available from: <https://cran.r-project.org/web/packages/fastshap/index.html>.
53. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD statement. *Ann Intern Med*. 2015;162:55–63.
54. Penfold RB, Johnson E, Shortreed SM, Ziebell RA, Lynch FL, Clarke GN, et al. Predicting suicide attempts and suicide deaths among adolescents following outpatient visits. *J Affect Disord*. 2021;294:39–47.
55. Kessler RC, Stein MB, Petukhova MV, Bliese P, Bossarte RM, Bromet EJ, et al. Predicting suicides after outpatient mental health visits in the Army Study to Assess Risk and Resilience in Servicemembers (Army STARRS). *Mol Psychiatry*. 2017;22:544–51.
56. Malhi GS, Bell E, Boyce P, Bassett D, Berk M, Bryant R, et al. The 2020 Royal Australian and New Zealand College of psychiatrists clinical practice guidelines for mood disorders: bipolar disorder summary. *Bipolar Disord*. 2020;22:805–21.
57. McIntyre RS, Rosenblat JD, Nemeroff CB, Sanacora G, Murrrough JW, Berk M, et al. Synthesizing the evidence for ketamine and esketamine in treatment-resistant depression: an international expert opinion on the available evidence and implementation. *Am J Psychiatry*. 2021;178:383–99.
58. McIntyre RS, Suppes T, Tandon R, Ostacher M. Florida best practice psychotherapeutic medication guidelines for adults with major depressive disorder. *J Clin Psychiatry*. 2017;78:703–13.
59. Ross EL, Zivin K, Maixner DF. Cost-effectiveness of electroconvulsive therapy vs. pharmacotherapy/psychotherapy for treatment-resistant depression in the United States. *JAMA Psychiatry*. 2018;75:713–22.
60. Reti IM. A rational insurance coverage policy for repetitive transcranial magnetic stimulation for major depression. *J Ect*. 2013;29:e27–e28.
61. Delgado J, Ali S, Fleck K, Agnew C, Southgate A, Parkhouse L, et al. Stratified care vs. stepped care for depression: a cluster randomized clinical trial. *JAMA Psychiatry*. 2022;79:101–8.
62. Browne G, Steiner M, Roberts J, Gafni A, Byrne C, Dunn E, et al. Sertraline and/or interpersonal psychotherapy for patients with dysthymic disorder in primary care: 6-month comparison with longitudinal 2-year follow-up of effectiveness and costs. *J Affect Disord*. 2002;68:317–30.
63. Schramm E, van Calker D, Dykierok P, Lieb K, Kech S, Zobel I, et al. An intensive treatment program of interpersonal psychotherapy plus pharmacotherapy for depressed inpatients: acute and long-term results. *Am J Psychiatry*. 2007;164:768–77.
64. Riggs PD, Mikulich-Gilbertson SK, Davies RD, Lohman M, Klein C, Stover SK. A randomized controlled trial of fluoxetine and cognitive behavioral therapy in adolescents with major depression, behavior problems, and substance use disorders. *Arch Pediatr Adolesc Med*. 2007;161:1026–34.
65. Hollon SD, DeRubeis RJ, Evans MD, Wiemer MJ, Garvey MJ, Grove WM, et al. Cognitive therapy and pharmacotherapy for depression. Singly and in combination. *Arch Gen Psychiatry*. 1992;49:774–81.
66. Vitiello B, Silva SG, Rohde P, Kratochvil CJ, Kennard BD, Reinecke MA, et al. Suicidal events in the treatment for adolescents with depression study (TADS). *J Clin Psychiatry*. 2009;70:741–7.
67. Molnar C. Interpretable machine learning: a guide for making black box models explainable. Christoph Molnar: Munich, Germany, 2022.
68. Meehan AJ, Lewis SJ, Fazel S, Fusar-Poli P, Steyerberg EW, Stahl D, et al. Clinical prediction models in psychiatry: A systematic review of two decades of progress and challenges. *Mol Psychiatry*. 2022;27:2700–8.
69. Furman JL, Trivedi MH. Chapter 29 - Biomarker-based treatment selection: A precision medicine approach for depression. In: Quevedo J, Carvalho AF, Zarate CA, editors. *Neurobiology of Depression*. Academic Press: Cambridge, MA, 2019, pp 331–40.
70. Glasgow RE, Kwan BM, Matlock DD. Realizing the full potential of precision health: the need to include patient-reported health behavior, mental health, social determinants, and patient preferences data. *J Clin Transl Sci*. 2018;2:183–5.
71. Leung LB, Ziobrowski HN, Puac-Polanco V, Bossarte RM, Bryant C, Keusch J, et al. Are veterans getting their preferred depression treatment? A national observational study in the Veterans Health Administration. *J Gen Intern Med*. 2022;37:3235–41.
72. Delevry D, Le QA. Effect of treatment preference in randomized controlled trials: Systematic review of the literature and meta-analysis. *Patient*. 2019;12:593–609.
73. Windle E, Tee H, Sabitova A, Jovanovic N, Priebe S, Carr C. Association of patient treatment preference with dropout and clinical outcomes in adult psychosocial mental health interventions: a systematic review and meta-analysis. *JAMA Psychiatry*. 2020;77:294–302.
74. Maslej MM, Furukawa TA, Cipriani A, Andrews PW, Sanches M, Tomlinson A, et al. Individual differences in response to antidepressants: a meta-analysis of placebo-controlled randomized clinical trials. *JAMA Psychiatry*. 2021;78:490–7.
75. Kamenov K, Twomey C, Cabello M, Prina AM, Ayuso-Mateos JL. The efficacy of psychotherapy, pharmacotherapy and their combination on functioning and quality of life in depression: a meta-analysis. *Psychol Med*. 2017;47:414–25.
76. Guo Z, Cheng J, Lorch SA, Small DS. Using an instrumental variable to test for unmeasured confounding. *Stat Med*. 2014;33:3528–46.
77. Vertosick EA, Assel M, Vickers AJ. A systematic review of instrumental variable analyses using geographic region as an instrument. *Cancer Epidemiol*. 2017;51:49–55.

78. U.S. Department of Veterans Affairs. Primary Care-Mental Health Integration (PC-MHI) 2022 [updated 2022 Sep 19. Available from: <https://www.patientcare.va.gov/primarycare/PCMHI.asp>.
79. Brookhart MA, Schneeweiss S. Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results. *Int J Biostat.* 2007;3:Article 14.
80. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *J Clin Epidemiol.* 2013;66:1386–96.
81. Ertefaie A, Small DS, Flory JH, Hennessy S. A tutorial on the use of instrumental variables in pharmacoepidemiology. *Pharmacoepidemiol Drug Saf.* 2017;26:357–67.
82. Swanson SA, Miller M, Robins JM, Hernán MA. Definition and evaluation of the monotonicity condition for preference-based instruments. *J Epidemiol.* 2015;26:414–20.
83. Qiu H, Carone M, Sadikova E, Petukhova M, Kessler RC, Luedtke A. Optimal individualized decision rules using instrumental variable methods. *J Am Stat Assoc.* 2021;116:174–91.
84. Adekkanattu P, Sholle ET, DeFerio J, Pathak J, Johnson SB, Campion TR Jr. Ascertaining depression severity by extracting Patient Health Questionnaire-9 (PHQ-9) scores from clinical notes. *AMIA Annu Symp Proc.* 2018;2018:147–56.
85. Xu Z, Vekaria V, Wang F, Cukor J, Su C, Adekkanattu P, et al. Using machine learning to predict antidepressant treatment outcome from electronic health records. *Psychiatr Res Clin Pract.* 2023;5:118–25.
86. Han S, Zhang RF, Shi L, Richie R, Liu H, Tseng A, et al. Classifying social determinants of health from unstructured electronic health records using deep learning-based natural language processing. *J Biomed Inform.* 2022;127:103984.
87. Wang G, Yang G, Du Z, Fan L, Li X ClinicalGPT: Large language models finetuned with diverse medical data and comprehensive evaluation. *arXiv.* 2023; e-pub ahead of print 16 June 2023; <https://doi.org/10.48550/arXiv.2306.0996>.
88. U.S. Department of Veterans Affairs. VA Informatics and Computing Infrastructure (VINCI): U.S. Department of Veterans Affairs; 2022 [updated 2022 March 16; cited 2023 Jun 27]. Available from: <https://www.research.va.gov/programs/vinci/default.cfm>.

## ACKNOWLEDGEMENTS

**Disclaimer** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Veterans Health Administration.

## AUTHOR CONTRIBUTIONS

RCK and RMB conceptualized and designed the study. RCK and NHZ drafted the manuscript. HL and ES led the statistical analysis. RMB obtained funding. CJK, RCK, AL, ES, and SW provided administrative, technical, and material support. RCK, AL, NAS, ES, and SW supervised the study. All authors were involved in the acquisition, analysis, or interpretation of data. All critically revised the final manuscript draft and approved the submission.

## FUNDING

The research reported here was funded by the U.S. National Institute of Mental Health, Grant number R01MH121478, and by the VA Center of Excellence for Suicide Prevention, Canandaigua VA Medical Center.

## COMPETING INTERESTS

In the past 3 years, RCK was a consultant for Cambridge Health Alliance, Canandaigua VA Medical Center, Holmusk, Partners Healthcare, Inc., RallyPoint Networks, Inc., and Sage Therapeutics. He has stock options in Cerebral Inc., Mirah, PYM, Roga Sciences and Verisense Health. The remaining authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41380-024-02500-0>.

**Correspondence** and requests for materials should be addressed to Ronald C. Kessler.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.