

Heterogeneity in Antidepressant Treatment and Major Depressive Disorder Outcomes Among Clinicians

Sarah Rathnam, MS; Kamber L. Hart, MS; Abhishek Sharma, PhD; Pilar F. Verhaak, BS; Thomas H. McCoy, MD; Finale Doshi-Velez, PhD; Roy H. Perlis, MD, MSc

 Supplemental content

IMPORTANCE While abundant work has examined patient-level differences in antidepressant treatment outcomes, little is known about the extent of clinician-level differences. Understanding these differences may be important in the development of risk models, precision treatment strategies, and more efficient systems of care.

OBJECTIVE To characterize differences between outpatient clinicians in treatment selection and outcomes for their patients diagnosed with major depressive disorder across academic medical centers, community hospitals, and affiliated clinics.

DESIGN, SETTING, AND PARTICIPANTS This was a longitudinal cohort study using data derived from electronic health records at 2 large academic medical centers and 6 community hospitals, and their affiliated outpatient networks, in eastern Massachusetts. Participants were deidentified clinicians who billed at least 10 *International Classification of Diseases, Ninth Revision (ICD-9)* or *Tenth Revision (ICD-10)* diagnoses of major depressive disorder per year between 2008 and 2022. Data analysis occurred between September 2023 and January 2024.

MAIN OUTCOMES AND MEASURES Heterogeneity of prescribing, defined as the number of distinct antidepressants accounting for 75% of prescriptions by a given clinician; proportion of patients who did not return for follow-up after an index prescription; and proportion of patients receiving stable, ongoing antidepressant treatment.

RESULTS Among 11 934 clinicians treating major depressive disorder, unsupervised learning identified 10 distinct clusters on the basis of *ICD* codes, corresponding to outpatient psychiatry as well as oncology, obstetrics, and primary care. Between these clusters, substantial variability was identified in the proportion of selective serotonin reuptake inhibitors, selective norepinephrine reuptake inhibitors, and tricyclic antidepressants prescribed, as well as in the number of distinct antidepressants prescribed. Variability was also detected between clinician clusters in loss to follow-up and achievement of stable treatment, with the former ranging from 27% to 69% and the latter from 22% to 42%. Clinician clusters were significantly associated with treatment outcomes.

CONCLUSIONS AND RELEVANCE Groups of clinicians treating individuals diagnosed with major depressive disorder exhibit marked differences in prescribing patterns as well as longitudinal patient outcomes defined by electronic health records. Incorporating these group identifiers yielded similar prediction to more complex models incorporating individual codes, suggesting the importance of considering treatment context in efforts at risk stratification.

Author Affiliations: Harvard John A. Paulson School of Engineering and Applied Sciences, Cambridge, Massachusetts (Rathnam, Sharma, Doshi-Velez); Center for Quantitative Health, Massachusetts General Hospital, Boston (Hart, Verhaak, McCoy, Perlis); Department of Psychiatry, Harvard Medical School, Boston, Massachusetts (Hart, McCoy, Perlis).

Corresponding Author: Finale Doshi-Velez, PhD, Harvard John A. Paulson School of Engineering and Applied Sciences, 29 Oxford St, Cambridge, MA 02138 (finale@seas.harvard.edu); Roy H. Perlis, MD, MSc, Center for Quantitative Health, Massachusetts General Hospital, 185 Cambridge St, Sixth Floor, Boston, MA 02114 (rperlis@mgh.harvard.edu).

JAMA Psychiatry. 2024;81(10):1003-1009. doi:10.1001/jamapsychiatry.2024.1778
Published online July 10, 2024.

In the United States, most prescriptions for antidepressants for major depressive disorder (MDD) are written not by psychiatrists or other psychiatric prescribers, but by primary care clinicians or clinicians in other specialties.^{1,2} However, the impact of treatment setting has rarely been studied. Individuals receiving treatment for MDD in primary care settings, rather than general psychiatric care, may be less open to treatment with antidepressant medications,³ less likely to receive guideline-congruent treatment,⁴ and more likely to lack access to outpatient mental health services.⁵ Those less connected to primary care may be treated for MDD in specialty settings; for example, young and low-income women with MDD are often treated through obstetrics and gynecology (OB/GYN) practices.⁶⁻⁸ However, there are barriers to screening, diagnosis, and treatment of MDD within specialty practices, which may contribute to poorer psychiatric and medical outcomes in those settings.⁹

While prior research suggested variation in depression treatment across nonpsychiatric clinicians,¹⁰ these studies were limited to small and select cohorts of patients or comparisons across 1 or 2 clinical settings. Real-world data drawn from electronic health records (EHRs) represent an opportunity to understand variability in MDD treatment between clinicians in the context of routine clinical care, as opposed to clinical trials or survey studies (eg, a questionnaire assessing physician diagnosis and treatment practices¹⁰), and include patient-clinician heterogeneity that is often intentionally removed from prospective trials.¹¹⁻¹⁸

Most prior efforts to model outcomes in psychiatric disorders using EHR data have focused on patient-level characteristics. To date, there has been little work leveraging EHR data to examine how systemic factors, such as clinician-level characteristics like specialty and location, impact treatment outcomes. This study aimed to address this gap by analyzing EHR data to identify and differentiate prescribers who treat MDD within the outpatient networks of a large health system. Rather than relying on metadata such as clinic location and specialty, which exhibit very high levels of missingness and secular trends, lack robust ontologies, and are difficult to translate between hospitals and health systems,^{19,20} we identified clusters of clinicians based on their predominant diagnostic codes. We then sought to compare antidepressant prescribing patterns both between and within clinician clusters and to determine the extent to which different prescribing settings were associated with differential treatment outcomes.

Methods

Data and Inclusion Criteria

The study cohort was composed of all clinicians from outpatient networks affiliated with a large health system in eastern Massachusetts with 2 academic medical centers and 6 community hospitals. The data consisted of EHR data collected between March 1, 2008 (when routine electronic prescribing was standardized across the hospital systems), and April 27,

Key Points

Question To what extent do differences in clinician setting explain variability in major depression treatments and outcomes?

Findings In this cohort study derived from electronic health record data, antidepressant prescribing patterns and outcomes varied significantly between prescriber groups. Clinician clusters were significantly associated with clinical outcomes.

Meaning Studies of antidepressant prescribing in real-world settings, and efforts at risk stratification or personalization of care, should include information on treatment setting and other clinician-level factors alongside individual patient characteristics.

2022. Individual clinicians were identified with a unique anonymized identifier. All diagnostic codes (*International Classification of Diseases, Ninth Revision [ICD-9]*²¹ and *Tenth Revision [ICD-10]*²²), procedure codes, and prescriptions associated with that clinician were collected for analysis. Data analysis occurred between September 2023 and January 2024. The study was approved by the Mass General Brigham institutional review board and the Harvard University institutional review board with a waiver of informed consent because only deidentified data were used and no participant contact was required. This article was prepared in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline.

To limit the analysis to outpatient clinicians who regularly treat MDD, we restricted the data a priori to those who on average billed for at least 10 individuals with MDD per year during the period in which the clinician appeared in the dataset, where billing for MDD is defined by the ICD-9 or ICD-10 diagnosis codes in eTable 1 in Supplement 1. Analyses were limited to outpatient encounters. Patients with at least 1 MDD code during the study period were included in analyses. To ensure adequate data associated with each clinician, we restricted our analysis to clinicians with at least 100 diagnostic codes in total. (In general, a priori selection of broader thresholds and inclusion criteria was intended to maximize generalizability across health systems in which patterns of utilization and coding may vary from this one.)

To reduce the dimensionality of the diagnostic codes, they were aggregated into the 530 clinical categories defined by Clinical Classifications Software Refined (CCSR), version 2022-1 (Agency for Healthcare Research and Quality).²³ Each clinician's high-dimensional representation consisted of their counts across diagnostic CCSR codes. The dimensionality was then further reduced to 2 dimensions via uniform manifold approximation and projection (UMAP)^{24,25} and then clustered using a gaussian mixture model (GMM).²⁶ Full details are presented in the eMethods and eFigure 1 in Supplement 1. The goal of this data-driven clustering approach was to identify groups of clinicians who treat similar patients in their clinical practice. This strategy allowed us to distinguish 10 clusters of clinicians across traditional department boundaries (eg, department of psychiatry or department of internal medicine), as we anticipated that there would be ample variability in the types of patients seen within each department or clinic.

Statistical Analysis

The threshold for statistical significance was considered to be $P < .05$. Testing was 2-tailed.

Outcomes Definition

Antidepressant Prescription Rate | We first examined how the rate of antidepressant prescribing differs across clinician clusters. For each clinician, we calculated the total number of antidepressant prescriptions written as well as the total number of antidepressant prescriptions divided by the clinician's total number of diagnostic codes. The latter serves as a proxy for volume, to capture the frequency of antidepressant prescriptions relative to the services provided overall by that clinician. We also considered differences between clusters in the use of different antidepressant medications categorized by class: selective serotonin reuptake inhibitor (SSRI), selective norepinephrine reuptake inhibitor (SNRI), monoamine oxidase inhibitor, tricyclic antidepressant (TCA), or atypical or other category. After determining the proportion of antidepressant prescriptions from each medication class within each cluster, we used analysis of variance (ANOVA) for each class to determine whether there were statistically significant differences between clusters. Post hoc comparisons between clusters were then performed using Tukey honestly significant difference testing.

Antidepressant Prescription Heterogeneity | We also considered the heterogeneity of medications used by each clinician. This was calculated by determining the number of antidepressants that accounted for 75% of all prescriptions for each clinician, as a means of estimating the breadth of a given prescriber's "repertoire" of medications. This analysis was limited to clinicians with at least 10 prescriptions in the dataset. Prescription heterogeneity is reported both by individual physician and on average within each clinician cluster. We also sought to determine whether antidepressant prescribing rates were correlated with heterogeneity of prescribing using Pearson correlation.

Stable Antidepressant Treatment and Treatment Dropout | We then examined the following patient-level outcomes: (1) the rate at which individual patients remained on a stable antidepressant treatment regimen, and (2) the rate at which patients dropped out of psychiatric treatment. Stability was defined, consistent with our prior work,²⁷ as the continuation of the same antidepressant treatment in the 180-day period following the initial antidepressant prescription, with at least 1 antidepressant prescription in the first 90 days and at least 1 in the subsequent 90 days. Similarly, dropout was defined as a discontinuation of antidepressant and nonpharmacologic psychiatric treatment. In line with prior work,^{14,27} if a patient did not receive an antidepressant prescription or other psychiatric treatment in the 90 days after their initial prescription, that individual was characterized as having dropped out of treatment. We excluded patients completely lost to follow-up by excluding those with no facts in the EHR during the subsequent 90-day period.

These outcomes were calculated from the initial documented antidepressant prescription, and thus we associated the outcome with the clinician responsible for that

prescription. This conservative choice ensured that information about the future—including future clinicians—could not influence outcome predictions. Dropout and treatment stability averaged by clinician were compared across clusters using a χ^2 test, with Bonferroni correction for pairwise comparisons.

Classification Methods

In addition to analyzing the univariate association between clinician clusters and patient stability and dropout, we also assessed the utility of the clinician clusters as additional features for predicting the patient outcomes of achieving stable antidepressant treatment and treatment dropout. (This effort did not aim to derive clinically applicable prediction models per se, but rather to understand the informativeness of the clusters more generally.) We applied logistic regression and tuned L1 regularization strength and learning rate, choosing the hyperparameter combination that performed best on the area under the receiver operating characteristic curve (AUROC) metric, to predict outcomes based on all patient data as well as only based on the clusters associated with the patient's clinicians.

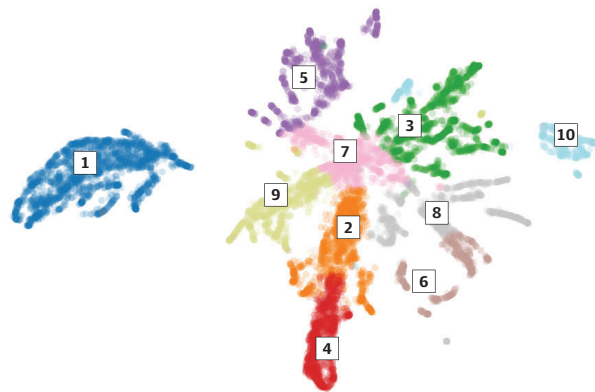
We compared the results of these regressions to a 2-phase optimization designed to allow us to determine what patient-specific demographic characteristics and codes were responsible for the outperformance of the all-patient-data model compared with the clinician cluster-only model. Phase 1 used the clinician clusters as predictors, and phase 2 added the patient demographic characteristics and codes. Additional details are included in the eMethods in [Supplement 1](#).

Results

The clinician cohort comprised 11 934 individuals associated with 381 623 unique patients. Characteristics of the patient cohort as well as flow diagrams for patient and clinician cohorts are summarized in eTable 2, eTable 3, eFigure 2, and eFigure 3 in [Supplement 1](#).

Figure 1 depicts the 2-dimensional clinician representation, referred to as an embedding, after applying dimensionality reduction via UMAP and clustering via GMM. We first sought to validate that the embedding represents coherent groupings of clinicians by assessing the distribution of selected CCSR codes across clusters. **Figure 2** illustrates that the fraction of total depressive disorder CCSR codes, uncomplicated pregnancy codes, and essential hypertension codes cluster in specific regions of the embedding (see eTable 4 in [Supplement 1](#) for associated CCSR and ICD-10 codes). The uneven distribution of such codes (as well as select others; eFigure 4 in [Supplement 1](#)) within the embedding suggests that regions of the clinician embedding are distinct from other regions in terms of the codes billed by clinicians in that region. The informativeness of the clusters was further confirmed by examining the predominant diagnostic codes billed by the clinicians in each cluster. Psychiatric disorders dominate the largest, most well-defined cluster (cluster 1), OB/GYN diagnoses dominate in cluster 10, ophthalmic diagnoses in cluster 7, and joint and connective tissue disorders

Figure 1. Two-Dimensional Embedding of Clinical Classifications Software Refined Counts by Clinician, Colored by Cluster



Clusters have been named based on our review of the predominant diagnostic codes for each cluster: 1 indicates general psychiatry; 2, primary care (low volume); 3, cancer (high volume); 4, primary care (high volume); 5, musculoskeletal pain; 6, cardiovascular disease; 7, ophthalmology; 8, kidney disease; 9, cancer (low volume); and 10, obstetrics and gynecology.

in cluster 5. (Below, clusters will be referred to by both number and a descriptive label based on the predominant diagnostic codes.) The full list of top codes for each cluster is shown in eTable 5 in Supplement 1.

Next, we described the differences between clusters in terms of prescribing patterns. Prescribers with no prescriptions during the study period were excluded from analysis. The rate of nonprescribing clinicians and the overall rate of AD prescribing (including nonprescribers) are depicted in eFigure 5, eFigure 6, and eTable 6 in Supplement 1. Clusters varied widely in overall rates of antidepressant prescription (Table) as well as in ratios between prescriptions for different classes of antidepressants (Figure 3). ANOVA indicated statistically significant differences between clusters in rates of prescribing for major antidepressant classes (SSRI: $F = 11.3$, $P < .001$; SNRI: $F = 11.8$, $P < .001$; TCA: $F = 2.4$, $P = .01$). Results for pairwise comparisons are presented in eTables 7 through 10 in Supplement 1. SSRIs were prescribed at the greatest rate in the OB/GYN cluster (cluster 10), with pairwise differences being statistically significant for 5 of the remaining 9 clusters.

We also examined the heterogeneity of prescribing by each individual in the cluster (eFigure 7 in Supplement 1 is a histogram illustrating the distribution of heterogeneity scores, eFigure 8 in Supplement 1 is the clinician embedding colored by heterogeneity score for each clinician, and eTable 11 in Supplement 1 illustrates mean heterogeneity by cluster). In general, these distributions reflect predominantly small numbers of medication classes, with the psychiatric clinicians (cluster 1) showing the most within-clinician prescribing variation. We confirmed by ANOVA that the differences in heterogeneity between clusters were statistically significant ($F = 62.3$, $P < .001$), with significant differences between clusters in post hoc comparisons (eTable 12 in Supplement 1). Clinician clusters with greater antidepressant prescribing rates did not demonstrate significantly greater

heterogeneity in prescribing ($r_8 = 0.32$, $P = .37$; eFigure 9 in Supplement 1).

Finally, we considered the variation in outcomes across the clusters. Figure 4A displays stability outcomes by clinician, with the color corresponding to the mean outcome for all patients associated with that clinician. These are summarized by cluster in eTable 13 in Supplement 1. Patient stability rates differed significantly across all clinician clusters ($\chi^2_9 = 245$, $P < .001$). Post hoc pairwise comparisons between clusters via χ^2 test with Bonferroni correction show that stability outcomes for patients associated with clinicians in high-volume primary care (cluster 4) are significantly more common than those in all other clusters except musculoskeletal pain clinicians (cluster 5). See eTable 14 in Supplement 1 for all pairwise comparisons.

Similarly, Figure 4B and eTable 15 in Supplement 1 show rates of patient dropout by cluster. As with stability, we confirmed that dropout rates differed significantly across all clinician clusters ($\chi^2_9 = 1331$, $P < .001$). Dropout was lowest among psychiatric clinicians (cluster 1) (eTable 18 in Supplement 1).

Last, eTables 16 and 17 in Supplement 1 examine the predictive validity of clinician clusters. The model using only the 10 clinician clusters as features was predictive of dropout (AUROC = 0.61; 95% CI, 0.60-0.63), albeit with less discrimination than a model using all demographic characteristics and code counts at index prescription (AUROC = 0.67; 95% CI, 0.66-0.69). The clinician clusters were also modestly predictive of achieving stability (AUROC = 0.59; 95% CI, 0.56-0.61), but less so than individual clinical features (AUROC = 0.63; 95% CI, 0.61-0.65). For both dropout and stability, the features that were significant after L1 regularization in the final model for each prediction task are summarized in eTables 18 and 19 in Supplement 1.

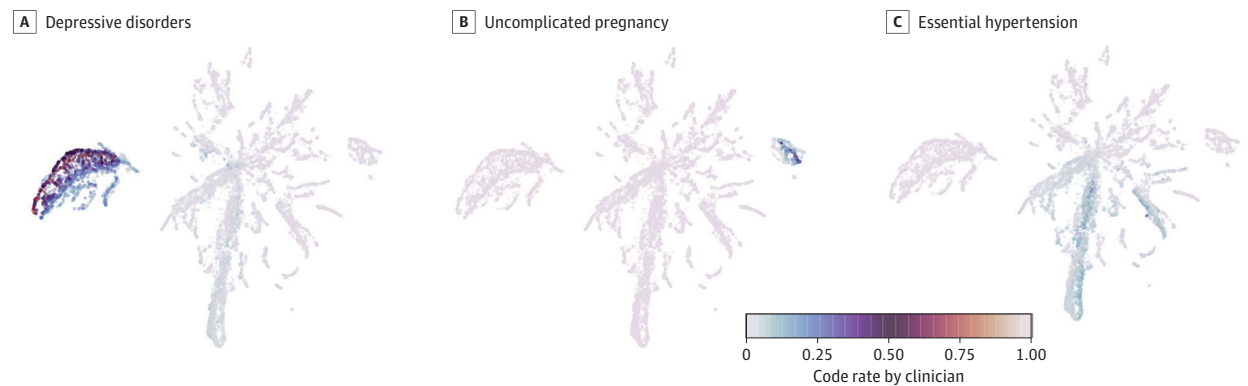
Discussion

In this study of 11 934 outpatient clinicians who treat MDD across multiple outpatient networks, we identified significant differences in antidepressant prescribing across clinician clusters and differences in antidepressant treatment outcomes in terms of the likelihood of achieving stable treatment, or dropping out of treatment, between these clusters.

While most investigations of antidepressant prescribing focus on individual clinics or general psychiatry, clinicians in the cancer and kidney disease clusters had the highest rate of antidepressant prescribing, similar to that of outpatient psychiatry after excluding nonprescribing clinicians. Rates of MDD are known to be high in these populations, and many patients in these settings may not receive adequate mental health treatment.^{28,29} Furthermore, antidepressant medications such as TCAs, SSRIs, or SNRIs may also be prescribed to treat chronic pain comorbid with depressive symptoms in this context.³⁰

As anticipated, our analyses also demonstrate that clinicians in the general psychiatry cluster exhibited the greatest

Figure 2. Depressive Disorder, Uncomplicated Pregnancy, and Essential Hypertension Codes Clustered in Specific Regions of the Embedding



Clusters are shown for depressive disorders (A), uncomplicated pregnancy (B), and essential hypertension (C).

heterogeneity of prescribing across antidepressant medication classes, with other clusters showing more limited prescribing patterns. This restriction likely reflects a tendency among primary care and other nonpsychiatric clinicians to provide standardized care in the treatment of depression that begins with SSRIs or SNRIs as recommended first-line treatment.³¹ Specialties like OB/GYN practices are also common treatment settings for MDD as they may be the only touchpoint younger women have with the medical system.⁶⁻⁸ Consistent with prior literature, our analyses demonstrate that SSRIs were prescribed at a higher rate in the OB/GYN clinician cluster than for all other clinician clusters.³²

Our findings have implications for efforts to develop precision medicine methods for the treatment of depression, highlighting the importance of considering treatment setting in such approaches in addition to patient-level features. In aggregate, the heterogeneity we identified underscores the need to consider aspects specific to the clinician alongside patient-level features in efforts to develop precision medicine strategies in psychiatry. Naively considering all depression treatment trials to be the same, as in our own prior work, risks modeling system-level features (eg, where someone receives treatment) rather than clinically relevant ones.^{13,14} Even if these sets of patient features as proxies for clinician features improve prediction, they are likely to generalize poorly across health systems and, as our results suggest, even across clinical settings in a single system—as such, empirical approaches to deriving clinician features may be valuable.^{33,34} EHRs typically include minimal metadata for clinicians, such as practice location or specialty; often, these data change over time as a health system grows, shrinks, or restructures. Further, we would not expect all clinics within a single department to be equivalent in both their practice patterns and patient populations. Finally, it may be difficult to harmonize information on practice location across health systems, limiting the generalizability of models that rely on specific locations. By choosing to cluster clinicians based on CCSR codes, we can derive more flexible and scalable means of categorizing clinician-level factors.

Table. Mean Antidepressant Prescription Rate for Clinicians in Each Cluster, Excluding Nonprescribers

Cluster	Cluster name	Antidepressant prescription rate, mean (SD) ^a
1	General psychiatry	0.05 (0.07)
2	Primary care (low volume)	0.03 (0.03)
3	Cancer (high volume)	0.02 (0.06)
4	Primary care (high volume)	0.002 (0.00)
5	Musculoskeletal pain	0.01 (0.02)
6	Cardiovascular disease	0.01 (0.02)
7	Ophthalmology	0.10 (0.10)
8	Kidney disease	0.04 (0.06)
9	Cancer (low volume)	0.08 (0.08)
10	Obstetrics and gynecology	0.02 (0.03)

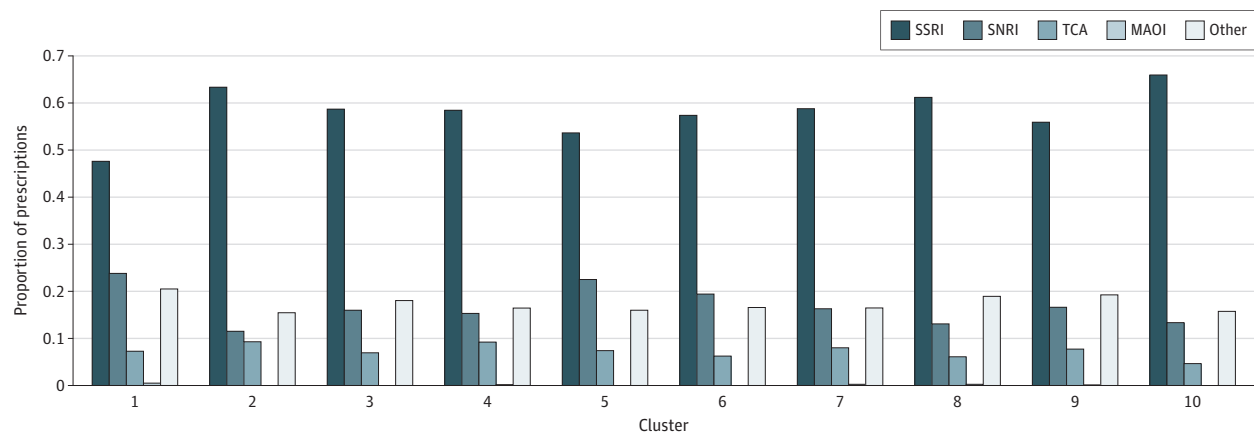
^a Antidepressant prescription rate was calculated as the number of antidepressant prescriptions divided by total diagnosis codes.

In general, health systems may be reluctant to examine individual clinician variability because of the potential medico-legal implications of identifying outliers. This concern is sometimes framed as protecting clinicians' privacy in the absence of providing informed consent—notably, a higher standard than is typically applied to large-scale EHR studies of patient characteristics. On the other hand, characterizing differences in practices and outcomes represents an opportunity to investigate the reasons for those differences and, potentially, to improve outcomes across a clinical population by adopting the most effective practices. By identifying clinicians with similar practice settings, this work lays the groundwork for efforts to understand, within a given setting, what accounts for variability among clinicians in terms of practice and outcomes.

Limitations

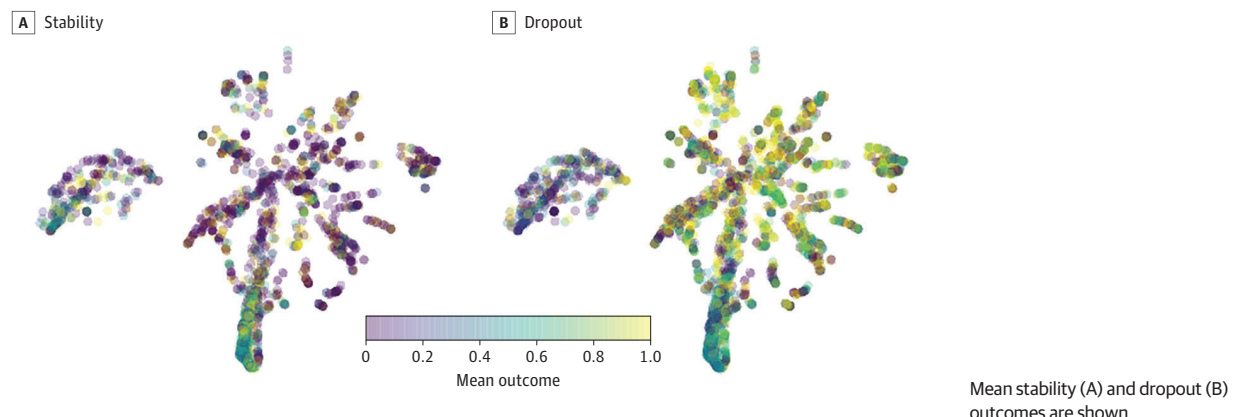
Our work has multiple limitations. First, given restrictions on the available individual-level data for each clinician, we are limited in the types of clinician-level information that can be included in our models. Second, while our cohort spans multiple academic medical centers and community hospitals, it is

Figure 3. Proportion of Prescriptions From Each Antidepressant Medication Class Stratified by Cluster



Cluster 1 indicates general psychiatry; 2, primary care (low volume); 3, cancer (high volume); 4, primary care (high volume); 5, musculoskeletal pain; 6, cardiovascular disease; 7, ophthalmology; 8, kidney disease; 9, cancer (low volume); and 10, obstetrics and gynecology. MAOI indicates monoamine oxidase inhibitor; SNRI, selective norepinephrine reuptake inhibitor; SSRI, selective serotonin reuptake inhibitor; TCA, tricyclic antidepressant.

Figure 4. Clinician Embedding Colored by Mean Stability and Dropout Outcomes of Each Clinician



limited to a single geographic region in a single country. Replication in other health systems will be valuable in characterizing variability at national and international levels. This replication will be especially important given the demonstrated differences between regions and countries in treatment practices as well as differences in the scope of practice for some clinicians, such as nurse practitioners, included in this analysis.

While our outcome measures are well suited to coded clinical data, they lack the precision of clinical trials incorporating depression rating scales. This is a standard critique of all work using large-scale clinical data, but correspondingly large-scale datasets using structured interviews and longitudinal clinical rating scales simply do not exist.^{13,15,16,18} Analyzing coded clinical data allows us to assess questions relating to the delivery of real-world clinical care, and future work should attempt to incorporate more specific symptom measures to supplement these coded data. Still, these data are particularly vulnerable to misclassification (eg, stable treatment could still reflect persistent depression) or missed diagnoses, as they

rely on artifacts of billing, not clinical assessment. Similarly, discontinuation of treatment could represent either an adequate treatment response leading to discontinuation of the medication or a decision about treatment futility. Last, we note that these data capture clinical care over a 14-year period and may reflect secular trends contributing to variations in the delivery of clinical care during the study period.

Conclusions

We found that patterns of antidepressant prescribing and treatment outcomes were markedly different across clinical contexts, but also within these contexts, even when those clinicians treated similar patient populations. Our finding that clinician-level features provided significant predictive power exposes an important analytical gap in prior work. Considering such clinician-level differences in concert with patient-level factors should facilitate the development of strategies for precision medicine and more efficient systems of care.

ARTICLE INFORMATION

Accepted for Publication: April 16, 2024.

Published Online: July 10, 2024.

doi:10.1001/jamapsychiatry.2024.1778

Author Contributions: Dr Perlis had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: McCoy, Doshi-Velez, Perlis.

Acquisition, analysis, or interpretation of data:

All authors.

Drafting of the manuscript: Rathnam, Hart, Sharma, Verhaak, McCoy, Perlis.

Critical review of the manuscript for important intellectual content: Hart, Verhaak, McCoy,

Doshi-Velez, Perlis.

Statistical analysis: Rathnam, Hart, Sharma, McCoy, Perlis.

Obtained funding: Doshi-Velez, Perlis.

Administrative, technical, or material support: Hart, Verhaak, McCoy, Doshi-Velez, Perlis.

Supervision: McCoy, Doshi-Velez, Perlis.

Conflict of Interest Disclosures: Dr McCoy reported receiving grants from the National Institute of Mental Health, National Human Genome Research Institute, Koa Health, and InterSystems and personal fees from Springer Nature outside the submitted work. Dr Doshi-Velez reported receiving grants from the National Institutes of Health during the conduct of the study. Dr Perlis reported receiving consulting fees from Vault Health, Belle Artificial Intelligence, Swan AI Studios, Mila Health, Alkermes, Genomind, Takeda, Circular Genomics, and Psy Therapeutics and holding equity in Circular Genomics, Psy Therapeutics, and Vault Health during the conduct of the study. No other disclosures were reported.

Funding/Support: This study was supported by grant RO1MH123804-03 from the National Institutes of Health (Drs Doshi-Velez and Perlis).

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

REFERENCES

- Unützer J, Park M. Strategies to improve the management of depression in primary care. *Prim Care*. 2012;39(2):415-431. doi:10.1016/j.pop.2012.03.010
- Park LT, Zarate CA Jr. Depression in the primary care setting. *N Engl J Med*. 2019;380(6):559-568. doi:10.1056/NEJMc1712493
- Van Voorhees BW, Cooper LA, Rost KM, et al. Primary care patients with depression are less accepting of treatment than those seen by mental health specialists. *J Gen Intern Med*. 2003;18(12):991-1000. doi:10.1111/j.1525-1497.2003.21060.x
- Duhoux A, Fournier L, Gauvin L, Roberge P. Quality of care for major depression and its determinants: a multilevel analysis. *BMC Psychiatry*. 2012;12(1):142. doi:10.1186/1471-244X-12-142
- Cunningham PJ. Beyond parity: primary care physicians' perspectives on access to mental health care. *Health Aff (Millwood)*. 2009;28(3)(suppl 1):w490-w501. doi:10.1377/hlthaff.28.3.w490
- Scholle SH, Haskett RF, Hanusa BH, Pincus HA, Kupfer DJ. Addressing depression in obstetrics/gynecology practice. *Gen Hosp Psychiatry*. 2003;25(2):83-90. doi:10.1016/S0163-8343(03)00006-9
- Hamel C, Lang E, Morissette K, et al. Screening for depression in women during pregnancy or the first year postpartum and in the general adult population: a protocol for two systematic reviews to update a guideline of the Canadian Task Force on Preventive Health Care. *Syst Rev*. 2019;8(1):27. doi:10.1186/s13643-018-0930-3
- Melville JL, Reed SD, Russo J, et al. Improving care for depression in obstetrics and gynecology: a randomized controlled trial. *Obstet Gynecol*. 2014;123(6):1237-1246. doi:10.1097/AOG.0000000000000231
- Deichen Hansen ME, Londoño Tobón A, Kamal Haider U, et al. The role of perinatal psychiatry access programs in advancing mental health equity. *Gen Hosp Psychiatry*. 2023;82:75-85. doi:10.1016/j.genhosppsych.2023.03.001
- Williams JW Jr, Rost K, Dietrich AJ, Ciotti MC, Zyzanski SJ, Cornell J. Primary care physicians' approach to depressive disorders: effects of physician specialty and practice structure. *Arch Fam Med*. 1999;8(1):58-67. doi:10.1001/archfam.8.1.58
- Coley RY, Boggs JM, Beck A, Simon GE. Predicting outcomes of psychotherapy for depression with electronic health record data. *J Affect Disord Rep*. 2021;6:100198. doi:10.1016/j.jadr.2021.100198
- Matarazzo BB, Brenner LA, Reger MA. Positive predictive values and potential success of suicide prediction models. *JAMA Psychiatry*. 2019;76(8):869-870. doi:10.1001/jamapsychiatry.2019.1519
- Lage I, McCoy TH Jr, Perlis RH, Doshi-Velez F. Efficiently identifying individuals at high risk for treatment resistance in major depressive disorder using electronic health records. *J Affect Disord*. 2022;306:254-259. doi:10.1016/j.jad.2022.02.046
- Pradier MF, McCoy TH Jr, Hughes M, Perlis RH, Doshi-Velez F. Predicting treatment dropout after antidepressant initiation. *Transl Psychiatry*. 2020;10(1):60. doi:10.1038/s41398-020-0716-y
- Perlis RH, Iosifescu DV, Castro VM, et al. Using electronic medical records to enable large-scale studies in psychiatry: treatment resistant depression as a model. *Psychol Med*. 2012;42(1):41-50. doi:10.1017/S0032319711000997
- McCoy TH Jr, Yu S, Hart KL, et al. High throughput phenotyping for dimensional psychopathology in electronic health records. *Biol Psychiatry*. 2018;83(12):997-1004. doi:10.1016/j.biopsych.2018.01.011
- McCoy TH Jr, Castro VM, Roberson AM, Snapper LA, Perlis RH. Improving prediction of suicide and accidental death after discharge from general hospitals with natural language processing. *JAMA Psychiatry*. 2016;73(10):1064-1071. doi:10.1001/jamapsychiatry.2016.2172
- Perlis RH, Fava M, McCoy TH Jr. Can electronic health records revive central nervous system clinical trials? *Mol Psychiatry*. 2019;24(8):1096-1098. doi:10.1038/s41380-018-0278-z
- Sweet LE, Moulaison HL. Electronic health records data and metadata: challenges for big data in the United States. *Big Data*. 2013;1(4):245-251. doi:10.1089/big.2013.0023
- Haneuse S, Arterburn D, Daniels MJ. Assessing missing data assumptions in EHR-based studies: a complex and underappreciated task. *JAMA Netw Open*. 2021;4(2):e210184-e210184. doi:10.1001/jamanetworkopen.2021.0184
- World Health Organization. *International Classification of Diseases, Ninth Revision (ICD-9)*. World Health Organization; 1977.
- World Health Organization. *International Statistical Classification of Diseases, Tenth Revision (ICD-10)*. World Health Organization; 1992.
- Agency for Healthcare Research and Quality. Clinical Classifications Software Refined (CCSR). Accessed December 4, 2023. https://hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp
- McInnes L, Healy J, Melville J. UMAP: uniform manifold approximation and projection for dimension reduction. Preprint posted September 18, 2020. doi:10.48550/arXiv.1802.03426
- van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res*. 2008;9(86):2579-2605.
- McLachlan GJ, Rathnayake S. On the number of components in a gaussian mixture model. *WIRES Data Min Knowl Discov*. 2014;4(5):341-355. doi:10.1002/widm.1135
- Hughes MC, Pradier MF, Ross AS, McCoy TH Jr, Perlis RH, Doshi-Velez F. Assessment of a prediction model for antidepressant treatment stability using supervised topic models. *JAMA Netw Open*. 2020;3(5):e205308. doi:10.1001/jamanetworkopen.2020.5308
- Zhu N, Virtanen S, Xu H, Carrero JJ, Chang Z. Association between incident depression and clinical outcomes in patients with chronic kidney disease. *Clin Kidney J*. 2023;16(11):2243-2253. doi:10.1093/ckj/sfad127
- Jansen F, Lissenberg-Witte BI, Hardillo JA, et al. Mental healthcare utilization among head and neck cancer patients: a longitudinal cohort study. *Psychooncology*. 2024;33(1):e6251. doi:10.1002/pon.6251
- Bonilla-Jaime H, Sánchez-Salcedo JA, Estevez-Cabrera MM, Molina-Jiménez T, Cortes-Altamirano JL, Alfaro-Rodríguez A. Depression and pain: use of antidepressants. *Curr Neuropharmacol*. 2022;20(2):384-402. doi:10.2174/1570159X19666210609161447
- Levkovich I, Elyoseph Z. Identifying depression and its determinants upon initiating treatment: ChatGPT versus primary care physicians. *Fam Med Community Health*. 2023;11(4):e002391. doi:10.1136/fmch-2023-002391
- Cooper WO, Willy ME, Pont SJ, Ray WA. Increasing use of antidepressants in pregnancy. *Am J Obstet Gynecol*. 2007;196(6):544.e1-544.e5. doi:10.1016/j.ajog.2007.01.033
- Burns ML, Kheterpal S. Machine learning comes of age: local impact versus national generalizability. *Anesthesiology*. 2020;132(5):939-941. doi:10.1097/ALN.0000000000003223
- Yang J, Soltan AAS, Clifton DA. Machine learning generalizability across healthcare settings: insights from multi-site COVID-19 screening. *NPJ Digit Med*. 2022;5(1):69. doi:10.1038/s41746-022-00614-9