Clinical Relevance of a Machine Learning Model for Automated Analyses of Depression Severity: the ePHQ-9 in Treatment-Resistant Depression

Pedro Alves, Carl D. Marci, MD, Joseph Zabinski PhD, MEM, Michael Batech, Costas Boussios, PhD | OM1, Inc., Boston, MA, USA

Background

Lack of widespread use of the Patient Health Questionnaire 9-item (PHQ-9) in clinical practice limits the number of patients with adequate outcome measurements to assess treatment follow-up and address other questions for patients with major depressive disorder (MDD) using real-world data (RWD). This study leveraged a previously developed and validated machine learning model of PHQ-9 scores and applied it to patients with a clinical variant of MDD (i.e., treatment resistant depression) using relevant notes from electronic medical records (EMR) from a large MDD real-world dataset (OM1, Inc.).

Objective

The PHQ-9 is a validated measure for depressive symptom severity but it is inconsistently documented in real-world data (RWD). This limits RWD studies addressing disease severity, treatment response and the patient journey in MDD. A previous AI and ML effort successfully estimated PHQ-9 scores creating the ePHQ-9 from clinical notes with strong analytic performance.[1] This study assessed the association between observed PHQ-9 and ePHQ-9 scores and physician-attested treatment-resistant depression (TRD), a known phenomenon in the literature without a diagnostic code, to further validate the ePHQ-9's utility in RWD.[2]

Methods

A large, curated U.S. based real-world dataset of MDD patients including claims and electronic medical records (EMR) with structured and unstructured clinical notes was used to identify a study cohort of patients with psychiatrist-attested TRD identified via natural language processing of text from the unstructured EMR notes. All patients had an observed PHQ-9 or ePHQ-9 score, or both, within 30 days of TRD attestation. The association between PHQ-9/ePHQ-9 disease severity (with five categorical divisions) and TRD status was evaluated by quantifying the proportion of TRD patients in each severity category. To facilitate visualization and interpretation, subsamples with a fixed ratio (1:4) of TRD to non-TRD patients were evaluated.

Results

- The dataset included 77,871 patients (29,608 with observed PHQ-9, 61,794 with ePHQ-9).
- Physician-attested TRD was present in 1,927 (observed PHQ-9 group) and 1,424 (ePHQ-9 group).
- TRD proportion increased monotonically with severity for both measures.
- Figure 1: For observed PHQ-9, the TRD proportion in the fixed-ratio subsamples ranged from 8.9% (none-minimal) to 36.5% (severe) (r = 0.31).
- Figure 2: The ePHQ-9 showed a stronger relationship (r = 0.42), with the TRD proportion ranging from 4.1% (none-minimal) to 52.7% (severe).

Figure 1. Physician-Attested TRD Proportion by PHQ-9 Severity

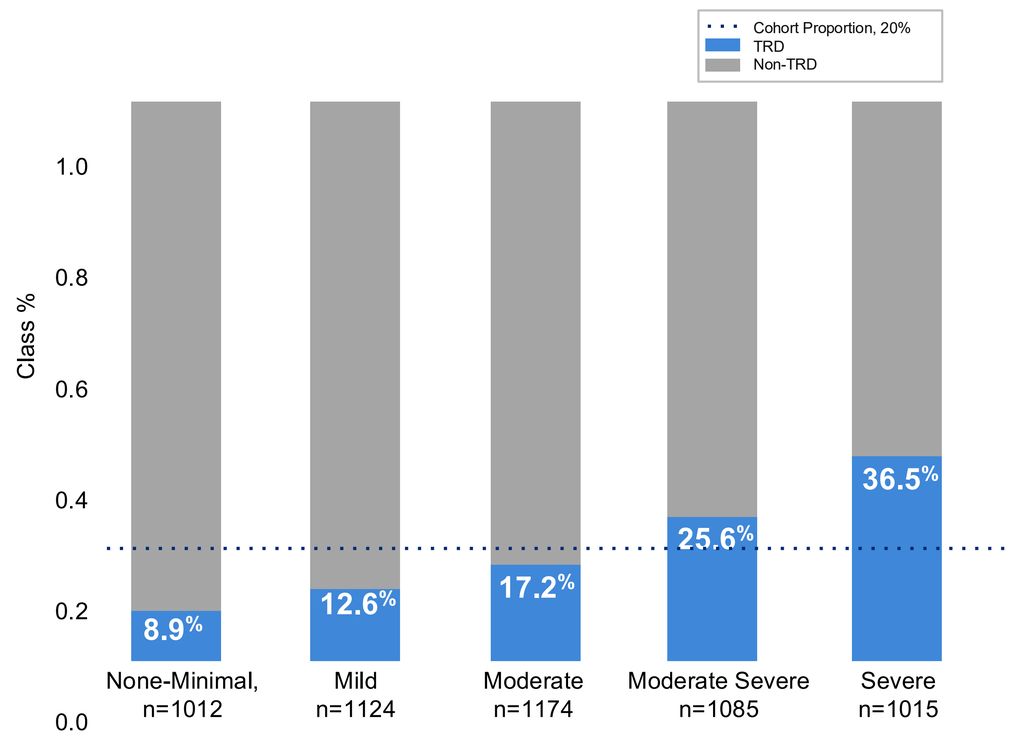
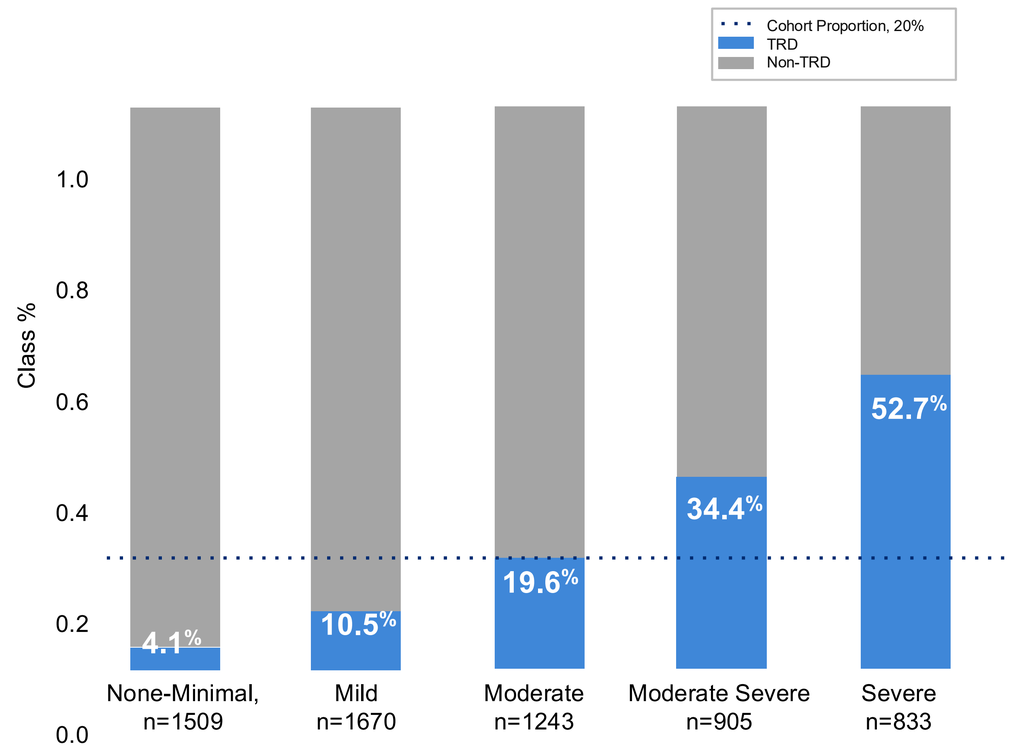


Figure 2. Physician-Attested TRD Proportion by ePHQ-9 Severity



Conclusions

- The ePHQ-9 is an AI based estimation model for the PHQ-9 that showed a stronger association with physician-attested TRD severity than observed PHQ-9 scores.
- The strong performance of the AI model may be linked to the fact that unlike the patient-reported observed PHQ-9, the ePHQ-9 is derived from clinician narratives, providing clinical consistency in the assessments of severity and treatment resistant status.
- The study helps validate the utility of the ePHQ-9 as a depression severity measure for use in RWD research.

Conference

Presented at ISPOR Europe Annual Conference, November 10-12; Scottish Event Campus, Glasgow Scotland UK

References

1. Alves, P, Marci CD, Cohen-Stavi C, Murray Whelan K, Boussios C. A Machine Learning Model Using Clinical Notes to Estimate PHQ-9 Symptom Severity Scores in Depressed Patients. Journal of Affective Disorders. 2025; 1(376):216-224.
2. McIntyre RS, Alsuwaidan M, Baune BT, Berk M, Demyttenaere K, Goldberg JF, et al. Treatment-resistant depression: definition, prevalence, detection, management, and investigational interventions. World Psychiatry Off J World Psychiatr Assoc WPA. 2023 Oct;22(3):394–412.