Addressing Gaps in Real-World Mental Health Data: Feasibility of AI-Based Estimation of the PHQ-9 and CGI-I in Major Depressive Disorder

Michelle Leavy, MPH[1], Zachary Bryant, PharmD[1], Ombretta Palucci, MS[2], Pedro Alves, MSc[1], Costas Boussios, PhD[1]
[1]OM1, Inc., Boston, MA, USA | [2]OM1, Inc., Pully, Switzerland

Background

The Patient Health Questionnaire 9-item (PHQ-9) and the Clinical Global Impression Scale–Improvement (CGI-I) are standard measures of depression severity and treatment response in major depressive disorder (MDD).[1,2]

Documentation of these measures in electronic medical records (EMRs) is inconsistent, which limits the utility of real-world data (RWD) for large-scale MDD studies.

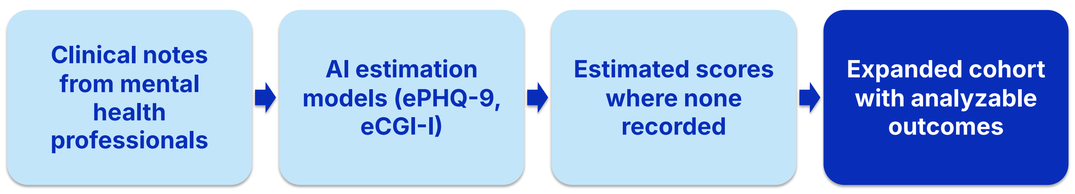
Artificial intelligence (AI) models have been developed and validated to estimate PHQ-9 (ePHQ-9) and CGI-I (eCGI-I) scores from routinely recorded clinical notes.[3,4]

Objectives

To assess the feasibility of applying the ePHQ-9 and eCGI-I models to expand the population of MDD patients with analyzable outcome data for RWD studies.

Methods

- **Data source:** OM1 MDD PremiOM Dataset, a large U.S.-based RWD source with structured and unstructured EMR data on >490,000 patients with MDD receiving care from mental health professionals.
- **eCGI-I cohort:** Patients initiating a new antidepressant with baseline and follow-up observations (n=182,750).
- **ePHQ-9 cohort:** Patients with MDD with or without an observed PHQ-9 score within the study window (n=77,871).
- **Estimation models:** Both models generate scores from clinical notes written by mental health professionals and were previously validated against recorded scores (Table 2).
- **Analysis:** The proportion of patients with analyzable outcome data was compared before and after applying each model.



Results

- **Documentation gap:** Only 21% (38,252/182,750) of the eCGI-I cohort had a recorded CGI-I score in the study timeframe, and 38% (29,608/77,871) of the ePHQ-9 cohort had an observed PHQ-9 score (Figure 1, Table 1).
- **eCGI-I:** Applying the eCGI-I model expanded the available study sample size by 4.3x to 6.0x across antidepressant drug classes.
- **ePHQ-9:** The ePHQ-9 model generated estimated scores for 61,794 patients, increasing the number with analyzable depression severity data from 29,608 to 77,871 (Figure 2).
- **Clinical validity:** In prior work, the ePHQ-9 correlated more strongly with physician-attested treatment-resistant depression severity (r=0.42) than observed PHQ-9 scores (r=0.31).[5]

Figure 1. Proportion of each cohort with a recorded outcome score

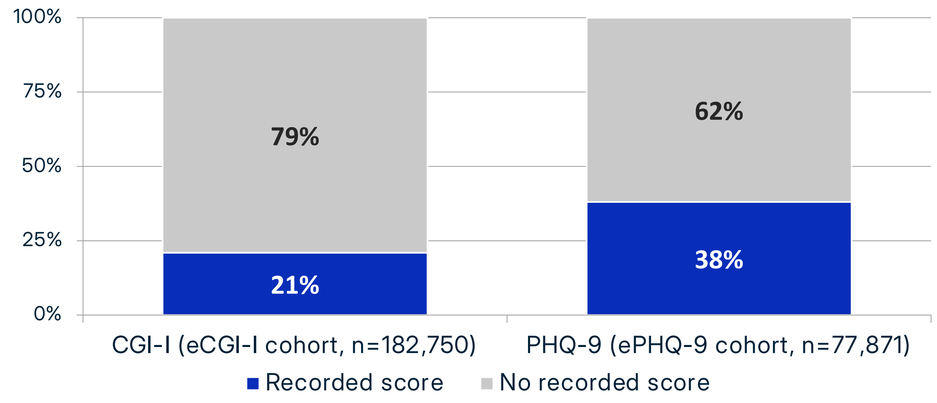


Table 1. Analyzable outcome data before and after estimation

	eCGI-I cohort	ePHQ-9 cohort
Patients in cohort, n	182,750	77,871
Recorded score, n (%)	38,252 (21%)	29,608 (38%)
No recorded score, n (%)	144,498 (79%)	48,263 (62%)
Sample size gain with estimation	4.3x–6.0x (by drug class)	2.6x

Figure 2. ePHQ-9 cohort: patients with depression severity data

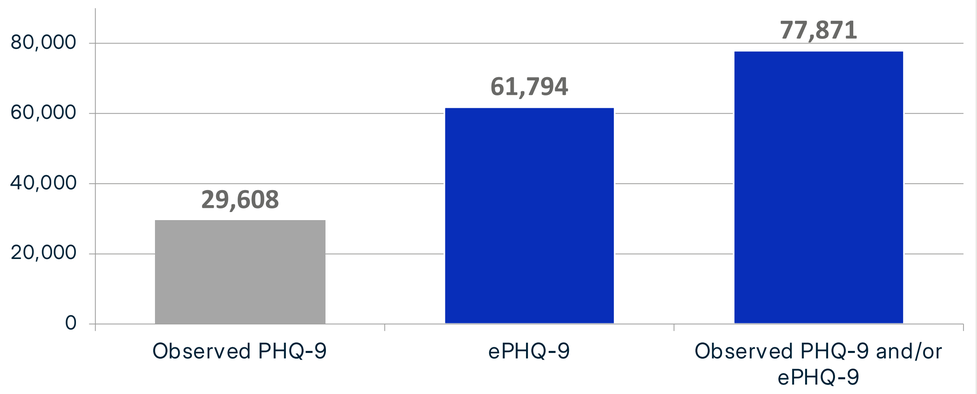


Table 2. Prior validation of the estimation models

Model	Validation metric	Result
ePHQ-9[3]	AUC (48,594 patients with recorded PHQ-9)	0.81
ePHQ-9[5]	Correlation with physician-attested TRD	0.42 (vs. 0.31 for observed PHQ-9)
eCGI-I[4]	AUC, improving vs. not improving (29,921 encounters)	0.71
eCGI-I[4]	Spearman correlation with recorded CGI-I	0.40

Conclusions

- AI-based estimation substantially increased the number of MDD patients with analyzable outcome data in RWD.
- Sample sizes grew by up to 6x for treatment response (eCGI-I) and 2.6x for depression severity (ePHQ-9).
- These validated models offer a scalable way to close critical gaps in structured mental health data.

Conference

Presented at ISPOR Europe 2026, 8–11 November 2026, Austria Center Vienna, Vienna, Austria. Poster RWD49.
Disclosures: All authors are employees of OM1, Inc.
Contact: Michelle Leavy, mleavy@om1.com

References

1. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-613.
2. Guy W. ECDEU Assessment Manual for Psychopharmacology. Rockville, MD: US Department of Health, Education, and Welfare; 1976.
3. 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. J Affect Disord. 2025;376:216-224. doi:10.1016/j.jad.2025.01.152
4. Severtson SG, Alves P, Boussios C, Marci CD. Development of a machine learning model for estimating CGI-I scores using electronic medical records from real-world data sources in schizophrenia, bipolar I and major depressive disorder. Poster presented at: ASCP Annual Meeting; May 2025; Scottsdale, AZ.
5. Alves P, Marci CD, Zabinski J, Batech M, Boussios C. Clinical relevance of a machine learning model for automated analyses of depression severity: the ePHQ-9 in treatment-resistant depression. Poster presented at: ISPOR Europe; November 2025; Glasgow, UK.