Artificial Intelligence-Based Estimation of the CGI-I to Address Gaps in Real-World Data and Increase Study Sample Size

Michelle B. Leavy, MPH, Zachary Bryant, PharmD, RPh, Carl Marci, MD, and Costas Boussios, PhD | OM1, Inc, Boston, MA, USA

Background

The Clinical Global Impression Scale - Improvement (CGI-I) is a clinician-reported measure for global assessment of change in patients with psychiatric conditions.[1] The CGI-I is widely used in clinical trials for conditions such as major depressive disorder (MDD), bipolar I disorder, schizophrenia, and anxiety disorders due to its clinical relevance for capturing treatment response.

The brevity and ease of use of the CGI-I also make it suitable for use in routine clinical practice for monitoring treatment response and patient outcomes over time,[2] yet documentation of the CGI-I is inconsistent in real-world data sources such as electronic medical records (EMRs). This limits the utility of these data for supporting large, heterogeneous real-world studies. To address this gap, a previous effort applied artificial intelligence (AI) methods to estimate CGI-I scores for patients using routinely-recorded clinical notes data with very good performance.

Objective

This effort assessed the feasibility of using the CGI-I estimation model to increase the sample size for RWD studies of treatment response in different classes of medications prescribed for psychiatric conditions.

Methods

The CGI-I model was developed and validated using data from the OM1 Mental Health Specialty Network (**Table 1**). The model was applied to the OM1 MDD PremiOM Dataset, a RWD source containing data on over 490,000 MDD patients with a diagnosis of depression and receiving treatment from a mental health professional. Patients were included in this feasibility assessment if they met the following inclusion criteria:

- Diagnosis of MDD
- New drug initiation (index date) for a drug indicated for depression treatment
- Baseline observation within 90 days prior to 14 days after the index day and follow-up observation between 45-273 days post-index date.

Results

The cohort included 182,750 patients (**Figure 1, 2**). Of these, 38,252 had at least one observed CGI-I score in the study timeframe. The remaining 144,498 patients had estimated CGI-I (eCGI-I) scores generated by the AI model.

Increases of 4.3x to 6.0x in available study sample size were observed across drug classes (**Figure 3**). Specifically, sample sizes increased for:

- Serotonin-specific modulators (6.0x)
- Serotonin-norepinephrine reuptake inhibitors (5.1x)
- Norepinephrine dopamine reuptake inhibitors (5.0x)
- Serotonin antagonists and reuptake inhibitors (5.0x)
- Atypical antipsychotics (4.9x)
- Tetracyclic antidepressants (4.5x)
- Tricyclic antidepressants (4.4x)
- Selective serotonin reuptake inhibitors (4.3x)

Conclusions

- Use of an AI-based model to estimate CGI-I scores for patients with depression increased the number of patients available for RWD studies across drug classes.
- This novel approach to addressing missing data in studies using RWD could lead to a better understanding of MDD treatment response and patient outcomes over time.

Table 1. eCGI-I Model Performance

PPV	NPV	AUC	Spearman's R	Pearson's R
0.46	0.82	0.71	0.40	0.38

Figure 1. Age of Patients with Observed CGI-I Scores vs. Patients with Observed and Estimated eCGI-Scores

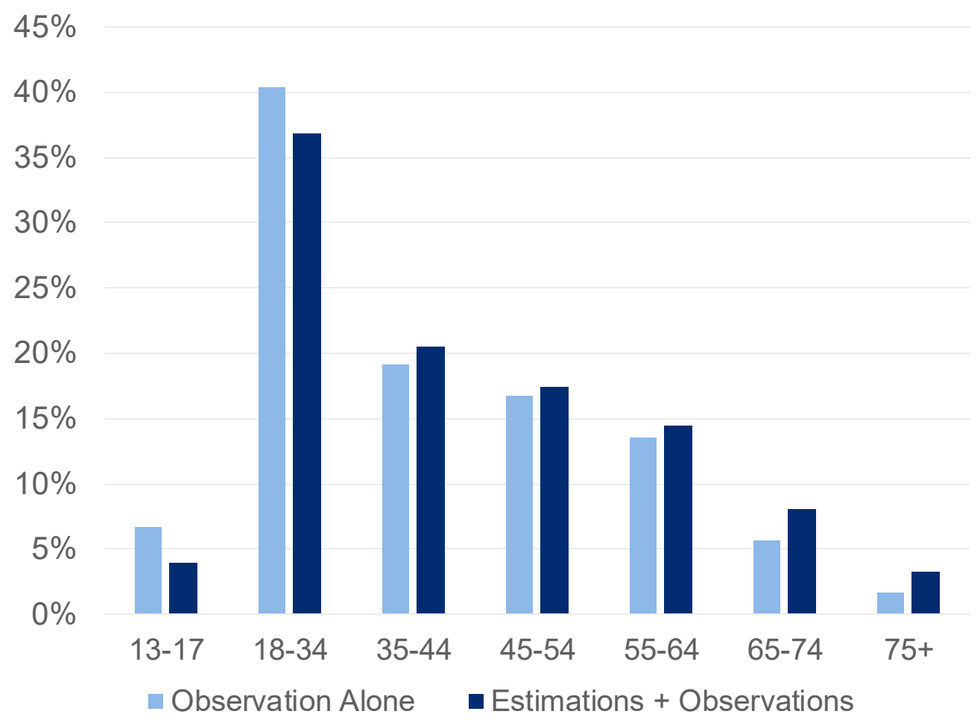


Figure 2. Sex of Patients with Observed CGI-I Scores vs. Patients with Observed and Estimated eCGI-Scores

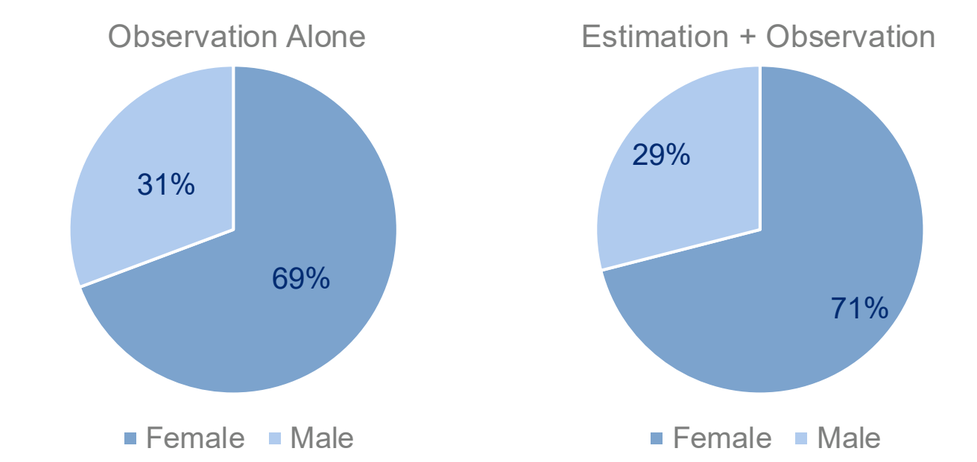
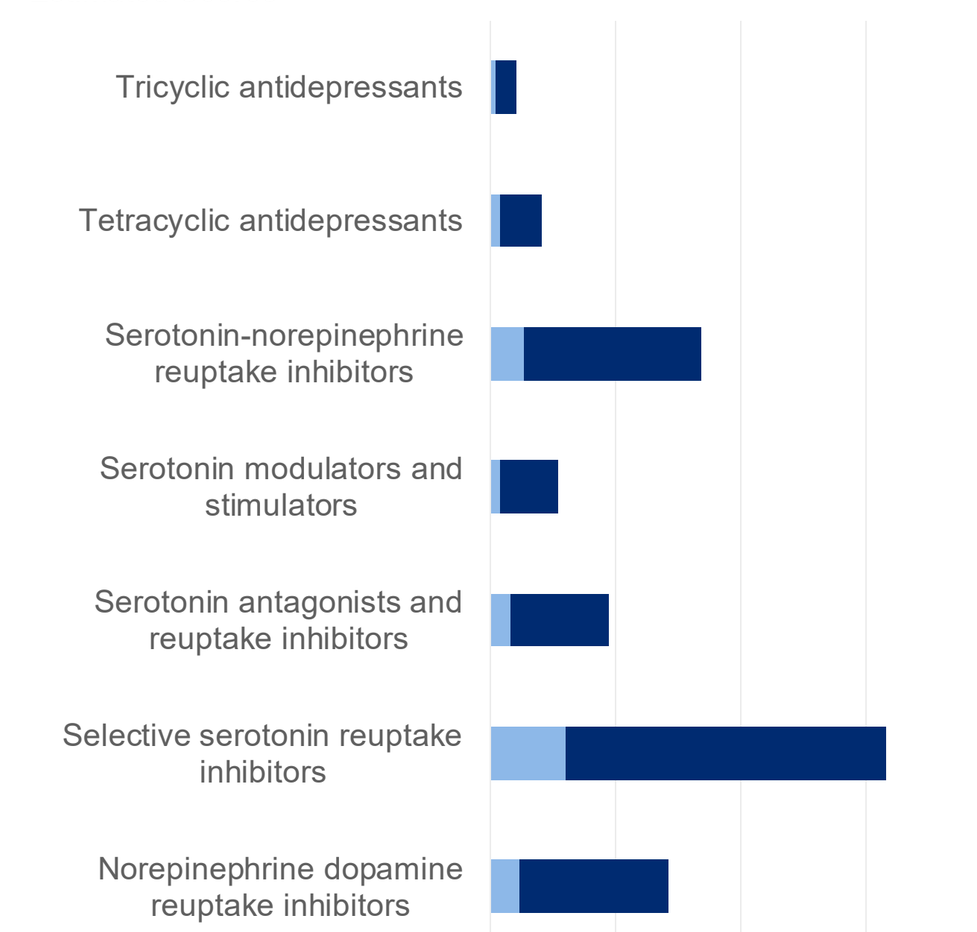


Figure 3. Available Patients for Inclusion in RWD Studies with Recorded CGI-I Scores Only vs. Patients with Recorded and Estimated Scores



Conference

Presented at the ISPE 2025 Annual Meeting. August 26, 2025. Washington, D.C., USA

References

1. Haro JM, Kamath SA, Ochoa S, Novick D, Rele K, Fargas A, et al. The Clinical Global Impression–Schizophrenia scale: a simple instrument to measure the diversity of symptoms present in schizophrenia. Acta Psychiatr Scand. 2003 May;107(s416):16–23.
2. Berk M, Ng F, Dodd S, Callaly T, Campbell S, Bernardo M, et al. The validity of the CGI severity and improvement scales as measures of clinical effectiveness suitable for routine clinical use. J Eval Clin Pract. 2008 Dec;14(6):979–83.
3. Alves P, Marci CD, Cohen-Stavi CJ, Whelan KM, Boussios C. A machine learning model using clinical notes to estimate PHQ-9 symptom severity scores in depressed patients. J Affect Disord. 2025 May 1;376:216–24.