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

Stevan Geoffrey Severtson, Pedro Alves, Costas Boussios, Carl D. Marci | OM1, Inc, Boston, MA, USA

Background

The Clinical Global Impression Scale - Improvement (CGI-I) is a widely used, transtheoretical, clinician-reported measure of change in a patient's symptoms over time. Due to its simplicity, the CGI-I is used in both clinical trials and real-world settings to evaluate treatments for patients with major psychiatric disorders. Consistent capture of the CGI-I over time is useful for evaluating treatment effectiveness. However, documentation of the CGI-I is inconsistent in real-world data (RWD) sources such as electronic medical records (EMRs). This limits the potential role of these RWD sources for supporting large, heterogeneous effectiveness studies. This study applied machine learning methods to fill in missing data in three RWD cohorts.

Methods

A machine learning model was developed to generate estimated CGI-I (eCGI-I) scores for clinical encounters. Training data were drawn from the OM1 Mental Health Specialty Network, an EMR data source from mental health specialists across the United States (OM1, Boston, MA). Patients from three condition specific datasets were included – bipolar I disorder, major depressive disorder and schizophrenia. These datasets are continuously updated subsets of patients within the OM1 Mental Health Specialty Network who have linked claims and EMR data. Encounters used to generate an estimated score were filtered to eliminate those with low levels of clinically relevant features.

Patient encounters with both recorded CGI-I scores and clinical notes from mental health specialists were identified and randomly assigned to a training cohort (n=1,971,155) using an extreme gradient boosting (XGBoost) model. Information from each encounter and the previous encounter was used to estimate CGI-I scores. A validation cohort (n=29,921) of encounters was drawn from the same condition datasets with the same diagnoses. To assess model performance, the area under the receiver-operating-characteristic curve (AUC) was calculated using a binarized version of the outcome.

Methods (continued)

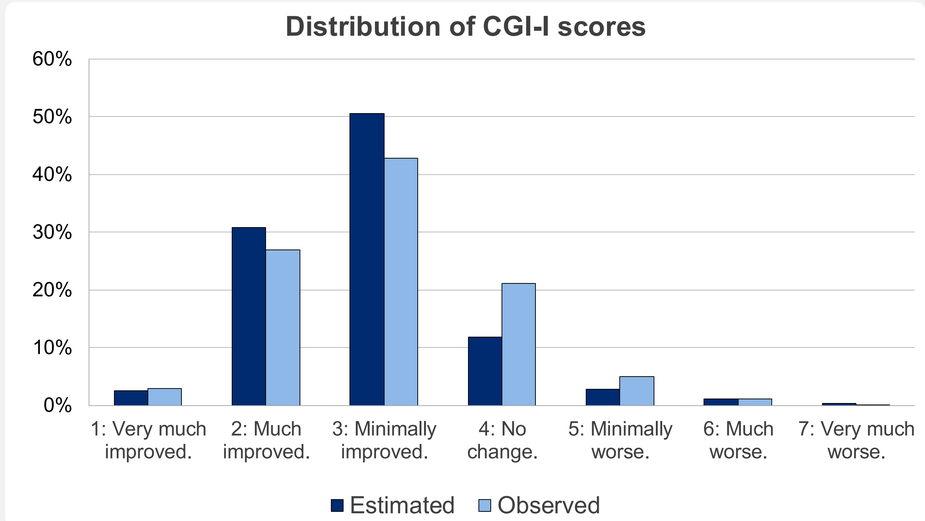
The binary outcome was a score of >=4 (unchanged or worsening) compared to scores <4 (improving). Continuous eCGI-I scores were evaluated using Spearman's and Pearson's correlation coefficients.

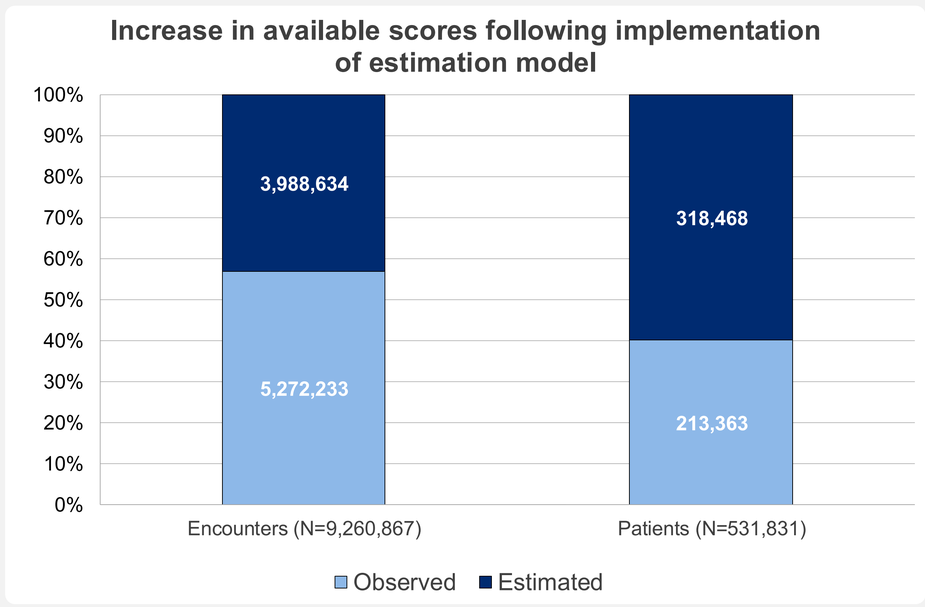
Results

The model had **an AUC of 0.71** when evaluating performance using the binarized version of the outcome in the validation cohort, a Spearman's r value of 0.40 and a Pearson's r value of 0.38 when evaluating performance using continuous scores. When applied to the encounters among patients with conditions of interest, there were 1.8x the number of CGI-I scores and 2.5x the number of patients with scores or 3,988,634 additional scores added (for a total of 9,260,867 encounters with recorded or estimated CGI-I scores).

Conclusions

- **A machine learning model can estimate CGI-I scores using information routinely recorded in EMR clinical notes from mental health professionals in three conditions – bipolar I disorder, major depressive disorder and schizophrenia.**
- **Use of the model provides a more complete view of changes in a patient's symptoms over time across three diagnoses.**
- **At the population level, application of the model to RWD sources increases the number of patient encounters with outcomes and expands available RWD patients for psychiatric research.**





Conference

Presented at ASCP Annual Conference. May 27-30 in Scottsdale, AZ, USA