Request for Copy of Published Material

Enclosure:

POSTER: Bell Lynum KS, Zhang Z, Atkins N et al. Presented at Psych Congress, September 17-19, 2025, San Diego, CA

Development of a Machine Learning Model Predicting Response to Aripiprazole Once-Monthly in Patients Diagnosed With Schizophrenia

Scan the QR code to access a digital copy of the poster and supplementary content

Karimah S Bell Lynum, PharmD[1]; Zhen Zhang, PhD[1]; Norman Atkins, Jr., PhD[1]; Anne Walker, PhD[2]; Hong Xiao, PhD[1]; Caitlin A Stamatis, PhD[1]; Murat Yildirim, MD, PhD[3]; Soma S Nag, PhD[1]; Mauricio Tohen, MD, DrPH[4]

[1]Otsuka Pharmaceutical Development & Commercialization Inc., Princeton, NJ, USA; [2]Lundbeck LLC, Deerfield, IL, USA; [3]H. Lundbeck A/S, Valby, Denmark; [4]Department of Psychiatry and Behavioral Sciences, University of New Mexico Health Sciences Center, Albuquerque, NM, USA

Background

- The antipsychotic aripiprazole is available as a once-daily oral tablet and as extended-release suspensions of aripiprazole monohydrate administered intramuscularly either once monthly or once every 2 months.[1-3]
- In the United States, aripiprazole once-monthly 400 mg (AOM 400) is approved for the treatment of schizophrenia in adults and for the maintenance monotherapy treatment of bipolar I disorder (BP-I) in adults.[2]
- Evidence from clinical trials and real-world studies shows that maintenance treatment with AOM 400 in patients diagnosed with schizophrenia delays relapse, improves functioning and health-related quality of life, and is well tolerated.[4-8]
- Understanding factors that might predict response to AOM 400 in the treatment of schizophrenia may help clinicians to provide more individualized care to patients initiating treatment, with the aim of optimizing outcomes. This may be especially beneficial for patients who are earlier in their treatment journey, since recurring relapse is associated with poorer outcomes.[9,10]
- Traditionally, statistical techniques such as regression modeling have been used to explore treatment-response relationships, relying on *a priori* hypotheses about which factors to investigate. Newer methodologies involving machine learning may improve this process by detecting complex, data-driven patterns that are not limited to predefined assumptions.[11,12]
- In prior studies, machine learning demonstrated clinical utility for predicting response to antipsychotics, including aripiprazole, in patients diagnosed with schizophrenia.[13,14]
- The current study builds on these findings by using machine learning to identify factors that predict response to AOM 400, with a view to identifying factors that clinicians may be able to monitor and target to improve treatment outcomes.

Here, we describe the development of a machine learning model to identify baseline factors predictive of response to AOM 400 in patients diagnosed with schizophrenia using data from a clinical trial.

A separate model to identify baseline factors predictive of response to AOM 400 versus placebo in patients diagnosed with BP-I has also been developed using data from a clinical trial, with results reported in poster 64.

Methods

Source data

- Baseline information used for the development of the model was derived from patients enrolled in a pre-post study that compared hospitalization rates in the 6-month periods before and after switching from oral antipsychotic treatment to AOM 400[15] (**Supplementary Figure 1**; please scan the QR code to access supplementary content).
- The study included 433 patients diagnosed with schizophrenia. Of these, 38 (8.8%) were hospitalized in the 6-month period following a switch to AOM 400.[15]

Variable screening

- A list of potential variables, encompassing demographic characteristics, vital signs, medical/disease history, and data from clinician-rated and patient-reported outcome measures, are shown in **Box 1**. These variables reflected the entirety of baseline data collected during the study.
- A univariate analysis was conducted to associate each variable, one at a time, to the outcome of hospitalization in the 6-month period following the switch to AOM 400. Variables meeting predefined thresholds were carried forward to the predictive modeling step.
- The goal of variable screening was to narrow the dataset to only the most relevant baseline factors, reducing 'noise' and improving the model's ability to identify meaningful predictors of treatment response.[16]

Box 1: Baseline variables screened in univariate analysis

Demographics	• Age • Gender • Race • Ethnicity • Height
Vital signs	• BMI/weight • Heart rate • Standing/supine diastolic/systolic blood pressure
Medical and disease history	• Hospitalization/ER visits in the 6-month preswitch period • Diagnosis history of prevalent diseases[a] including hypertension, insomnia, anxiety, asthma, depression, seasonal allergies, and obesity
Clinician-rated and patient-reported outcome measures	• PANSS including matched BPRS • CGI-S • QLS • SWN-S • DAI • AIMS • BARS • SAS • CSSRS

[a]Present in >30 patients

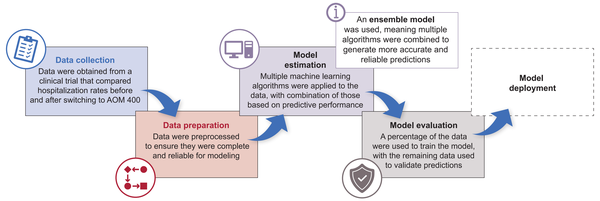
AIMS, Abnormal Involuntary Movement Scale; BARS, Barnes Akathisia Rating Scale; BMI, body mass index; BPRS, Brief Psychiatric Rating Scale (older version); CGI-S, Clinical Global Impression – Severity; CSSRS, Columbia Suicide Severity Rating Scale; DAI, Drug Attitude Inventory; ER, emergency room; PANSS, Positive and Negative Syndrome Scale; QLS, Quality of Life Scale; SAS, Simpson-Angus Scale; SWN-S, Subjective Well-being under Neuroleptic Treatment – Short Form

Methods (continued)

Predictive modeling

- Key steps in the predictive modeling process are shown in **Figure 1**.
- Premodeling steps were undertaken to address missing data and the overall low frequency of hospitalization events in the 6-month post-switch period.
- Commonly used binary machine learning algorithms with demonstrated predictive capability were evaluated for inclusion in an ensemble model. The ensemble was refined through iterative exclusion of classifiers with a weight of zero (indicating a trivial contribution to the model's predictive capability).
- Similar to the work of others,[14] data were randomly split into 70% for training the ensemble model and 30% for out-of-sample validation.
- Model performance was assessed using standard metrics, including those appropriate for imbalanced data. The importance of each variable in the final model was reported using SHapley Additive exPlanations.
- The established model was used to generate partial dependence plots (PDPs) illustrating the likelihood of nonhospitalization according to influential variables, while holding all other variables constant at population mean values.

Figure 1: Key steps in the predictive modeling process



AOM 400, aripiprazole once-monthly 400 mg

Results

- Overall, 163 baseline variables were considered (**Supplementary Table 1**; please scan the QR code to access supplementary content); of these, 37 were carried forward to the predictive modeling step.
- Three classifiers were included in the final ensemble model, each contributing to prediction accuracy based on their nonzero weighting (**Table 1**).
- The final model demonstrated a strong performance (**Table 2** and **Figure 2**).
- The importance ranking of each of the 37 variables included in the final predictive model is shown in **Figure 3**.
- A PDP depicting the model-predicted likelihood of nonhospitalization as a function of the top two ranked variables (ie, Positive and Negative Syndrome Scale [PANSS] items 'G12, Lack of judgment and insight' and 'P7, Hostility') is shown in **Figure 4**.
- A PDP depicting the model-predicted likelihood of nonhospitalization as a function of influential Quality of Life Scale (QLS) variables (ie, 'Extent of occupational role functioning' and 'Sociosexual relations') is shown in **Supplementary Figure 2** (please scan the QR code to access supplementary content).

Table 1: Final ensemble model

Final ensemble model	Weight
Random forests	0.001
eXtreme Gradient Boosting	0.406
Kernel-based Support Vector Machine	0.593

- In total, nine machine learning algorithms were evaluated for inclusion, with **predictions from three of the algorithms combined in an ensemble model.**
- Each algorithm was assigned a weight reflecting how much it contributed to improving the overall accuracy of the combined model **(a higher weight = a greater contribution to prediction accuracy).**

All other binary classifiers tested had a weight of zero

Table 2: Model performance metrics

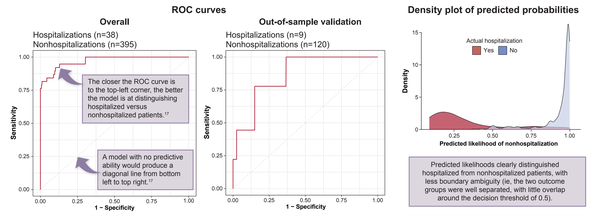
	Accuracy	Sensitivity	Specificity	F1-score	AUC (ROC)	AUC (PRC)
Overall N=433 (38 hospitalizations, 395 nonhospitalizations)	0.9	0.92	0.9	0.62	0.97	0.89
Out-of-sample validation, random 30% split, n=129 (9 hospitalizations, 120 nonhospitalizations)	0.84	0.78	0.84	0.4	0.86	0.44

- AUC-PRC in **out-of-sample validation was 0.44, which was considered robust** for highly imbalanced, rare-event binary classification (ie, 36 hospitalizations in 433 patients).
- An AUC-PRC value of 0.44 indicates **the model was ~six times better than that expected from random chance.**
- The model showed **strong overall accuracy** and an excellent ability to distinguish between patients who were hospitalized and those who were not.

AUC, area under the curve; AUC-PRC, area under the precision-recall curve; PRC, precision-recall curve; ROC, receiver-operating characteristic

Results (continued)

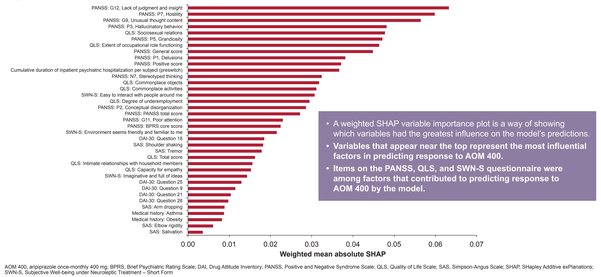
Figure 2: Model performance plots



Visualizations of model performance demonstrated robust discrimination between hospitalized and nonhospitalized patients.

ROC, receiver-operating characteristic

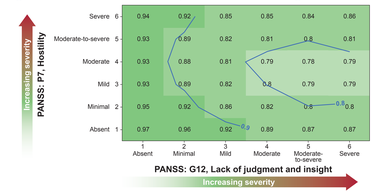
Figure 3: Variable importance ranking



- A weighted SHAP variable importance plot is a way of showing which variables had the greatest influence on the model's predictions.
- **Variables that appear near the top represent the most influential factors in predicting response to AOM 400.**
- **Items on the PANSS, QLS, and SWN-S questionnaire were among factors that contributed to predicting response to AOM 400 by the model.**

AOM 400, aripiprazole once-monthly 400 mg; BPRS, Brief Psychiatric Rating Scale; DAI, Drug Attitude Inventory; PANSS, Positive and Negative Syndrome Scale; QLS, Quality of Life Scale; SAS, Simpson-Angus Scale; SHAP, SHapley Additive exPlanations; SWN-S, Subjective Well-being under Neuroleptic Treatment – Short Form

Figure 4: Likelihood of nonhospitalization according to influential variables related to the PANSS



- The estimated probability of nonhospitalization reduced as the baseline severity of 'G12, Lack of judgment and insight' increased.
- A general positive efficacy signal was observed following a switch to AOM 400 regardless of item scores, with all predicted probabilities exceeding 0.5 (ie, indicating a likelihood of nonhospitalization).

No patients scored 7 (extreme) for PANSS items 'P7, Hostility' and 'G12, Lack of judgment and insight.'

AOM 400, aripiprazole once-monthly 400 mg; PANSS, Positive and Negative Syndrome Scale

Limitations

- The input data included a limited number of transition events (ie, 38 out of 433 subjects hospitalized), although methods appropriate for imbalanced data were applied to boost the model's predictive performance.
- The input data were derived from a trial that was not designed with predictive modeling in mind. As such, variable selection was limited to what was collected during the trial rather than being guided by a theory-driven framework. This may impact model predictive power and real-world applicability.
- Variables included in the predictive model were limited to the baseline data collected in the clinical study; other potentially relevant factors, such as substance use and sleep, were absent and are yet to be studied.
- Although variables captured in the PANSS were identified as influential, this tool is not typically used in routine clinical practice. This highlights the need for expanded modeling that includes real-world evidence.

Conclusions

- A robust machine learning model has been developed using clinical trial data to identify baseline factors predictive of response to AOM 400 in patients diagnosed with schizophrenia, where response was defined by hospitalizations
- Preliminary findings support the potential utility of current symptom severity for predicting response to AOM 400. In this model, influential PANSS items suggested that overall symptom severity, positive symptoms, and insight may be particularly important predictors.
- Quality of life and well-being related to social engagement also appear relevant for predicting response to AOM 400, indicating that environmental context matters and may not be merely a secondary concern. In this model, influential QLS items included 'Sociosexual relations' and 'Extent of occupational role functioning', along with 'Easy to interact with people around me' from the Subjective Well-being under Neuroleptic Treatment – Short Form (SWN-S) questionnaire. These findings highlight the importance of incorporating the subjective patient experience into treatment planning.
- Results of the model may be used to inform clinician-led monitoring and supportive interventions (eg, psychoeducation, therapy, medication adjustment) in patients who are being considered for AOM 400 treatment, with a view to optimizing treatment outcomes. This may have the greatest benefit early in the patient's treatment journey, when relapse prevention has the greatest potential to preserve long-term outcomes.[9,10]
- Further validation of the model using real-world data is planned, with the potential to explore a broader range of variables, such as social determinants of health, substance use, sleep, and socioeconomic status.

References

1. Otsuka America Pharmaceutical, Inc. Abilify (aripiprazole). Prescribing information. January 2025.
2. Otsuka America Pharmaceutical. Inc. Abilify Maintena (aripiprazole). Prescribing information. March 2025.
3. Otsuka America Pharmaceutical, Inc. Abilify Asimtufi (aripiprazole). Prescribing information. March 2025.
4. Kane et al. J Clin Psychiatry 2012; 73 (5): 617–624.
5. Fleischhacker et al. Br J Psychiatry 2014; 205 (2): 135–144.
6. Naber et al. Schizophr Res 2015; 168 (1-2): 498–504.
7. Naber et al. Schizophr Res 2018; 192: 205–210.
8. Mustafa et al. BMC Psychiatry 2019; 19 (1): 114.
9. Correll et al. Int J Neuropsychopharmacol 2025; 28 (5): pyaf018.
10. Lin et al. Front Psychiatry 2021; 12: 695672.
11. Chekroud et al. World Psychiatry 2021; 20 (2): 154–170.
12. Goldstein et al. Eur Heart J 2017; 38 (23): 1805–1814.
13. Wang et al. Front Genet 2022; 13: 848205.
14. Wu et al. JAMA Netw Open 2020; 3 (2): e1921660.
15. Kane et al. J Med Econ 2015; 18 (2): 145–154.
16. Michel et al. BMC Med Inform Decis Mak 2021; 21 (Suppl 4): 130.
17. Jiao & Du. Quant Biol 2016; 4: 320–330.

Disclosures

KSBL, ZZ, HX, CAS, SN: employees of Otsuka Pharmaceutical Development & Commercialization Inc.
NA: consultant for Otsuka Pharmaceutical Development & Commercialization Inc.
AW: employee of Lundbeck LLC.
MY: employee of H. Lundbeck A/S.
MT: received honoraria or consultation fees from Abbott, AbbVie, Alkermes, AstraZeneca, Elan, Gedeon Richter, Intracellular Therapies, Johnson & Johnson, Lilly, Lundbeck, Merck, Minerva, Neurocrine Biosciences, Otsuka, Pfizer, Roche, Sunovion, and Teva. MT was an employee at Lilly (1997-2008); his spouse was an employee at Lilly (1998-2013).

Acknowledgments

Medical writing and editorial assistance were provided by Lyndal Staples, BSc, and colleagues of Cambridge (a division of Prime, Cambridge, UK), with funding from Otsuka Pharmaceutical Development & Commercialization Inc. (Princeton, NJ, USA) and H. Lundbeck A/S (Valby, Denmark).
The sponsors thank the patients and their families who participated in this study.

Sponsorship

This work was supported by Otsuka Pharmaceutical Development & Commercialization Inc. (Princeton, NJ, USA) and H. Lundbeck A/S (Valby, Denmark).