Check for updates

Developing prediction models for symptom severity around the time of discharge from a tertiary-care program for treatment-resistant psychosis

Lik Hang N. Lee[1], Ric M. Procyshyn[2,3], Randall F. White[2], Kristina M. Gicas[4], William G. Honer[2,3] and Alasdair M. Barr[3,5]*

[1]Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada, [2]Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada, [3]British Columbia Mental Health and Substance Use Services Research Institute, Vancouver, BC, Canada, [4]Department of Psychology, York University, Toronto, ON, Canada, [5]Department of Anesthesiology, Pharmacology & Therapeutics, University of British Columbia, Vancouver, BC, Canada

Antipsychotics are the only therapeutic class indicated in the symptomatic management of psychotic disorders. However, individuals diagnosed with schizophrenia or schizoaffective disorder may not always benefit from these first-line agents. This refractoriness to conventional treatment can be difficult to address in most clinical settings. Therefore, a referral to a tertiary-care program that is better able to deliver specialized care in excess of the needs of most individuals may be necessary. The average outcome following a period of treatment at these programs tends to be one of improvement. Nonetheless, accurate prognostication of individual-level responses may be useful in identifying those who are unlikely to improve despite receiving specialized care. Thus, the main objective of this study was to predict symptom severity around the time of discharge from the Refractory Psychosis Program in British Columbia, Canada using only clinicodemographic information and prescription drug data available at the time of admission. To this end, a different boosted beta regression model was trained to predict the total score on each of the five factors of the Positive and Negative Syndrome Scale (PANSS) using a data set composed of 320 hospital admissions. Internal validation of these prediction models was then accomplished by nested cross-validation. Insofar as it is possible to make comparisons of model performance across different outcomes, the correlation between predictions and observations tended to be higher for the negative and disorganized factors than the positive, excited, and depressed factors on internal validation. Past scores had the greatest effect on the prediction of future scores across all 5 factors. The results of this study serve as a proof of concept for the prediction of symptom severity using this specific approach.

KEYWORDS

schizophrenia, psychosis, prediction model, tertiary care, treatment resistance

Introduction

Schizophrenia is a neurodevelopmental disorder characterized by altered brain connectivity and synaptic dysregulation (1–5). Antipsychotics are the primary therapeutic class indicated in the symptomatic management of psychotic disorders such as schizophrenia. Unfortunately, while most individuals show some therapeutic response to these drugs, an inadequate response to this first-line treatment is observed in approximately one-quarter of patients with first-episode psychosis (6–10). Moreover, there also exists a subset of patients who become refractory to treatment following a variable period of response (11). Consequently, specialized approaches, like the timely initiation of the atypical antipsychotic clozapine (12) may be warranted if there is an ongoing lack of clinical improvement despite multiple antipsychotic trials of adequate dose and duration.

Tertiary-care programs have shown promise in the effective delivery of these approaches under real world conditions (13–16). In a retrospective study investigating antipsychotic prescribing patterns at a tertiary-care program for treatment-resistant psychosis, we observed fewer instances of antipsychotic polypharmacy and a greater use of clozapine when comparing the drugs prescribed at discharge to those prescribed at admission (17). This trend towards a more evidence-based approach to pharmacotherapy coincided with decreases in symptom severity as quantified using the original three-factor Positive and Negative Syndrome Scale (PANSS) (18). Nevertheless, some individuals discharged from the program remained without clinical improvement. Given both the dearth of prognostic models in this patient population and the potential utility in identifying non-responders to specialized treatment approaches as early into their admission as possible, it was of interest to determine if symptom severity around the time of discharge can be predicted using only the information available at, or shortly after the time of admission, to a tertiary-care program. To this end, the current study aimed to develop and internally validate models for the prediction of PANSS scores around the time of discharge following a re-scoring of values to more accurately reflect the multidimensional symptomatology of treatment-resistant psychosis (19). An exploratory data analysis was then performed to identify predictors routinely selected by the model fitting procedure for further study in the future.

Materials and methods

The clinical research ethics board at the University of British Columbia approved the study protocol and granted a waiver of consent for the secondary use of retrospective chart review data, as previously (20, 21). Data processing, statistical analysis, and data visualization were accomplished using R (v. 3.5.1). Add-on packages facilitating these tasks include tidyverse (22), gamlss (23), gamboostLSS (24), rsample (25), ggplot2 (26), viridis (27), cowplot (28), and extrafont (29).

Overview of the model fitting procedure

A component-wise gradient boosting approach to beta regression (30, 31) was used to predict symptom severity around the time of discharge from a tertiary-care program for treatment-resistant psychosis. This involved the training of five distinct models, one for each symptom domain routinely identified by previous factor analyses of PANSS scores (32).

This approach to predictive modelling has the following advantages. If n is the number of items included in a given factor of the PANSS, then the total score on that factor can only take on integer values in the interval $[n,7n]$. These lower and upper bounds limit the utility of linear regression models (33), but a rescaling of the scores as decimal proportions of the upper bound would enable their prediction by beta regression models instead. These rescaled values would normally need to be continuous (34), but discrete proportions can be predicted in the same way if they reflect an underlying continuous beta distribution (35). We consider this to be a tenable assumption because the routine practice of reporting total scores only makes sense if they approximate the latent variables of interest. Furthermore, while maximum likelihood can be used to estimate parameters in beta regression, component-wise gradient boosting offers two distinct advantages. First, the functional form of non-linear relationships do not need to be specified beforehand if they are modelled using penalized regression spline (P-spline) base learners (36). Second, early stopping of the boosting algorithm can result in both shrinkage and variable selection since the additive predictor (η) is only updated with a fixed proportion of the fit of the base learner that best fits the negative gradient vector at each iteration (24). These are both desirable properties given how little is known about predictors of symptom severity in treatment-resistant psychosis following hospitalization at a tertiary care program.

Working within the framework of generalized additive models for location, scale and shape (23), the location (μ) and scale (σ) parameters of each beta distribution were related to their respective additive predictors (η_μ and η_σ) by a logit link function. The ensuing model was then trained using a non-cyclic algorithm capable of updating only one of these additive predictors in each iteration (37). Specifically, the additive predictor that most improved the empirical risk of the model following the inclusion of its best fitting base learner was updated with one-tenth of this fit. This intentional slowing of the learning rate enabled both shrinkage estimation and variable selection to occur should the algorithm be allowed to come to an early stop (24).

The optimal number of iterations with which to update both additive predictors in a specific model was identified by searching a vector of values for the number of iterations that minimizes the average predictive risk in 10-fold cross-validation. This hyperparameter, hereafter referred to as m_{stop}, was arbitrarily limited to a maximum of 1,000 iterations should earlier stopping not occur. Once the model has been updated to m_{stop} iterations, the predicted values of the distribution parameters were used to estimate the mean and the variance of the beta distribution based on the relationships: $E(X) = \mu; \mathrm{Var}(X) = \mu(1-\mu)\sigma^2$.

Nested cross-validation was then used in the absence of an independent testing data set to internally validate this model fitting procedure (38, 39). Model selection (i.e., the tuning of m_{stop}) was accomplished in the inner loop by 10-fold cross-validation. Model assessment (i.e., the estimation of prediction error) was accomplished in the outer loop by 100 times 10-fold cross-validation. In essence, the model fitting procedure was repeated 1,000 times, but instead of using all of the data to train each model, a different nine-tenths of the data was selected as the training data set in each iteration. This freed up the remaining one-tenth of the data to be used as a testing data set for the model in question.

Study population

Data were retrospectively abstracted from the charts of individuals admitted to the Refractory Psychosis Program at Riverview Hospital in Coquitlam, British Columbia, Canada for psychosis not improved by conventional treatment despite multiple antipsychotic trials of adequate dose and duration as documented by the referring physician (17). Those without antipsychotic use at admission or discharge, PANSS scores around the time of admission or discharge, or a primary diagnosis of schizophrenia or schizoaffective disorder based on the criteria outlined in either the Diagnostic and Statistical Manual III-R or IV were excluded upon identification. If an individual had multiple admissions to this program over the observation period, then data from only the most recent admission were included in the data set because later admissions were hypothesized to reflect higher degrees of treatment resistance. Following listwise deletion of admissions with missing data, 320 of 325 hospitalizations meeting the above selection criteria remained in the data set.

Variables

No actions were taken to blind the assessment of predictors for the outcome. While information routinely available around the time of admission had already been gathered, data for which the veracity was dependent on the reliability of the historian were excluded owing to the absence or inaccessibility of collateral information. The list of candidate predictors was therefore solely derived from the following set of variables: sex, age, diagnosis, year of admission, history of a prior clozapine trial, medications, and PANSS scores. A more detailed discussion of these variables is available elsewhere (17).

Sex, psychiatric diagnosis, and history of a prior clozapine trial were treated as binary data. Age and year of admission were treated as numeric data precise to one decimal place. The list of regularly prescribed medications (i.e., those not prescribed *pro re nata*) were initially processed by sorting the individual drugs into their respective therapeutic classes. Antipsychotic utilization was then quantified by the prescribed daily dose (PDD) to defined daily dose (DDD) ratio (i.e., PDD:DDD) (40) as was done previously (17). Patterns of antipsychotic use were further characterized using two binary variables to separately denote the use of clozapine and antipsychotic polypharmacy at admission. Contrary to the approach taken for antipsychotics as a therapeutic class, the use of antidepressants, mood stabilizers, benzodiazepines, and anticholinergics was strictly limited to qualitative representation by way of binary variables since these were not always prescribed for their primary indication [i.e., antidepressants and mood stabilizers may instead be used as off-label adjunctive agents to augment antipsychotic response (41)]. All remaining therapeutic classes were excluded from further consideration as candidate predictors since these were used to treat comorbidities present in only a minor subset of the study population. Factor-specific PANSS scores were obtained by summing the scores of individual items that had been grouped together according to the consensus five-factor model developed by Wallwork et al. (32) to approximate symptom severity in a manner that better reflects the underlying symptom dimensions in psychotic disorders.

After dummy coding the binary data and mean-centering the numeric data, a linear base learner was specified for each candidate predictor to model its effect on the distribution parameter of interest. A P-spline base learner was also specified for each numeric predictor to entertain the possibility of non-linear effects. Both linear and P-spline base learners were assigned one degree of freedom to mitigate any undue preference for increasing model complexity (42).

Outcomes

No actions were taken to blind the assessment of the different outcomes. Factor-specific PANSS scores around the time of discharge were derived in the same manner and for the same reason as those approximating symptom severity around the time of admission. However, the summed totals required further rescaling using the formula: $y_{\text{rescaled}} = (y - n + 1)/(7n - n + 1)$, where n is the number of PANSS items assigned to a particular factor, y is the total score on the original scale, and y_{rescaled} is an approximation of that score as a decimal proportion of $7n$ (i.e., the maximum score if all items in that factor were rated "extreme") (35). No additional steps were necessary to constrain the rescaled values to the interval $(0,1)$ as is required for beta regression since the numerator precludes values of 0 and the maximum score of $7n$ was not observed across any of the factors.

Statistical analysis

Agreement between model predictions and their respective observations was quantified within the outer loop of the nested cross-validation by pseudo R^2 (34, 43) and root mean squared error (RMSE). Differences between training and testing data sets in terms of these two metrics were interpreted as optimism in the apparent performance of boosted beta regression. A qualitative assessment was facilitated by the fitting of locally weighted least squares regression smoothers using the *loess* function in R. Probabilistic calibration was also assessed using probability integral transform (PIT) histograms (44, 45) following the discretization of predicted beta distributions (35). The five models trained on all data were not updated in light of the findings of this internal validation.

Exploratory data analysis

Since the refitting of each model in the outer loop of the nested cross-validation involved training data sets that were largely composed of the same observations (i.e., eight-ninths of the data are shared between any two training data sets in a single 10-fold cross-validation), the set of candidate predictors selected by the non-cyclical algorithm was expected to share some similarities between runs and repeats. To this end, the proportion of refitted models in which the base learner or base learners specified for a given variable were used to update an additive predictor at least once before termination of the algorithm was quantified and reported as the variable selection frequency for that particular additive predictor.

A series of partial dependence plots were constructed from the refitted models trained in the outer loop of the nested cross-validation to examine the marginal effect of each variable on predicted values of the distribution parameters (46). A second set of partial dependence plots was constructed from the five models trained on the full data set

following the discretization of predicted beta distributions to examine the marginal effect of each included variable on the probability of obtaining a specific score around the time of discharge.

Results

The variables available at or around the time of admission that were given consideration as candidate predictors of symptom severity around the time of discharge are summarized in Table 1. Briefly, 58.1% of the 320 individuals admitted to the Refractory Psychosis Program for management of treatment-resistant psychosis had previously been treated with clozapine, but only 20.3% of the study population were prescribed clozapine at admission. This was comparable in rate to the prescription of oral anticholinergics for the prophylactic or symptomatic management of extrapyramidal symptoms (20.6%). Total antipsychotic utilization generally exceeded the defined daily dose for conventional management of psychotic disorders as evidenced by a median PDD:DDD of 1.81, but this tendency toward higher doses was observed even in the absence of antipsychotic polypharmacy in 48.1% of the study population. The concurrent prescription of other psychotropics was relatively common despite the selection criteria excluding individuals with treatment-resistant psychosis attributable to either depression or bipolar disorder. For example, antidepressants, though less frequently prescribed compared to mood stabilizers and benzodiazepines (21.9% versus 37.5 and 31.9%), were also similar in rate to clozapine at admission.

Histograms depicting the total score on each factor of the PANSS around the time of discharge are presented alongside histograms depicting individual scores on their constituent items in Figure 1. Relative to the median total scores serving as approximations of symptom severity around the time of admission (Table 1), the corresponding values around the time of discharge were unanimously lower (i.e., less severe) across all five factors. In fact, there were: 6 individuals (1.9%) without items assigned to the positive factor (i.e., delusions [P1], hallucinatory behavior [P3], grandiosity [P5], and unusual thought content [G9]); 4 (1.3%) without items assigned to the negative factor (i.e., blunted affect [N1], emotional withdrawal [N2], poor rapport [N3], passive-apathetic social withdrawal [N4], lack of spontaneity and flow of conversation [N6], and motor retardation [G7]); 6 (1.9%) without items assigned to the disorganized factor (i.e., conceptual disorganization [P2], difficulty in abstract thinking [N5], and poor attention [G11]); 31 (9.7%) without items assigned to the excited factor (i.e., excitement [P4], hostility [P7], uncooperativeness [G8], and poor impulse control [G14]); 39 (9.1%) without items assigned to the depressed factor (i.e., anxiety [G2], guilt feelings [G3], and depression [G6]).

Model evaluation

As expected, the pseudo R^2 for each model trained on the full data set was better approximated by the cross-validated value on training data than that on testing data (Table 2). Specifically, the average optimism in pseudo R^2 for models predicting the total score on the positive, negative, disorganized, excited, and depressed factors was 0.05, 0.05, 0.05, 0.03, and 0.04, respectively. Optimism was likewise observed in terms of RMSE with each factor having an average difference between cross-validated values of 0.02 (Table 3). Visualization of this optimism

TABLE 1 Variables available at or around the time of admission to a tertiary care program for treatment-resistant psychosis given consideration as candidate predictors of symptom severity around the time of discharge.

Candidate predictors	Median [Q1, Q3] or Count (%)
Year of admission	2001.6 [1998.1, 2005.3]
Age	36.5 [27.0, 44.0]
Sex	
Male	190 (59.4%)
Female	130 (40.6%)
Psychotic disorder	
Schizophrenia	221 (69.1%)
Schizoaffective disorder	99 (30.9%)
Previous pharmacological interventions	
Prior clozapine trial	186 (58.1%)
Current pharmacological interventions	
Total antipsychotic utilization	1.81 [1.09, 2.75]
Clozapine	65 (20.3%)
Antipsychotic polypharmacy	166 (51.9%)
Antidepressant(s)	70 (21.9%)
Mood stabilizer(s)	120 (37.5%)
Benzodiazepine(s)	102 (31.9%)
Anticholinergic(s)	66 (20.6%)
Symptom severity by total PANSS scores	
Positive	16 [13, 19]
Negative	19 [16, 22.25]
Disorganized	11 [9, 13]
Excited	11 [8, 14]
Depressed	7 [5, 9]

$N = 320$. PANSS, Positive and Negative Syndrome Scale. Antipsychotic utilization was quantified by the prescribed daily dose (PDD) to defined daily dose (DDD) ratio (World Health Organization, 2020). PDD:DDD ratios were summed if antipsychotic polypharmacy was observed at admission. Drugs prescribed *pro re nata* have been excluded from the analysis. The theoretical range of total scores on each of the five factors of the PANSS is as follows: positive (4, 28), negative (6, 42), disorganized (3, 21), excited (4, 28), and depressed (3, 21). Individual items of the original PANSS Kay et al. (18) were assigned to these factors based on a consensus five-factor model developed by Wallwork et al. (32).

by density curves superimposed on histograms indicates a lesser degree of skewness for RMSE than pseudo R^2 (Figure 2).

Scatter plots depicting the extent of agreement between observations and predictions are presented in Figure 3. Among models trained on the full data set, deviation of the smooth curve from the identity line was most notable for the positive and excited factors. Though this was also reproduced in the cross-validated predictions of training data, the severity of the deviation appeared to be of lesser consequence when examining cross-validated predictions of testing data. The interpretation of PIT histograms derived from predictions on individual testing data sets of the nested cross-validation was limited by the small sample size ($N = 28$), but pooling of these predictions from each repeat in the outer loop yielded histograms that, if noticeably deviated from a uniform distribution, were most concerning for over-forecasting as suggested by low counts in the highest bins. This risk of over-forecasting at the highest bin is consistent with the predilection
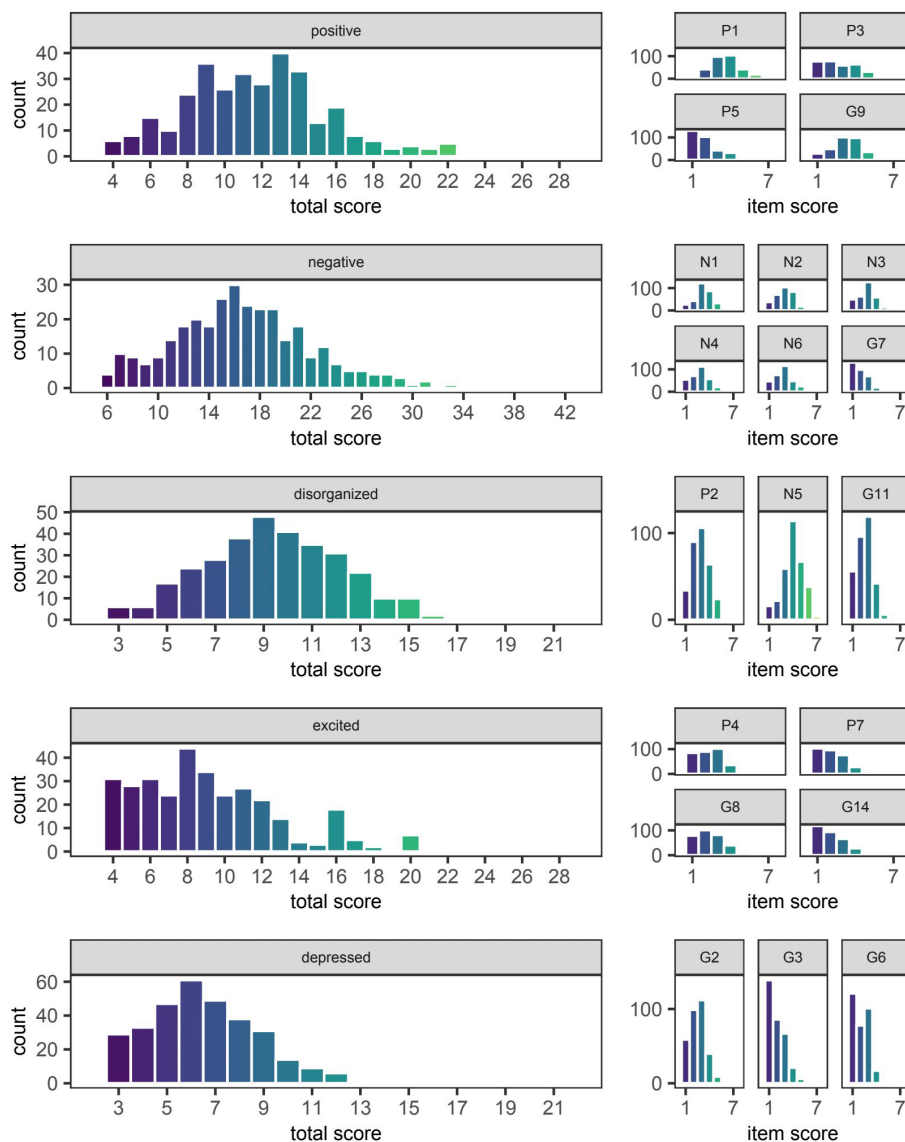
FIGURE 1
Symptom severity around the time of discharge as quantified by total scores on five different factors of the Positive and Negative Syndrome Scale
(PANSS). The median [Q1, Q3] score on the positive, negative, disorganized, excited, and depressed factors of the PANSS was 12 [9, 14], 16 [13, 20],
9 [7, 11], 9 [6, 11], and 6 [5, 8], respectively. The distribution of scores on individual items of the PANSS are shown to the right of the factor to which they
were assigned based on the consensus five-factor model developed by Wallwork et al. (32). The naming convention for individual items remains the
same as that used by Kay et al. (18) in the original three-factor model (i.e., P1-P7, N1-N7, and G1-G16 denote items numbered 1 through 7 or 1 through
16 of the positive, negative, and general psychopathology scales, respectively).

for overestimation seen in the previous set of scatter plots for cross-validated predictions on testing data (Figure 3).

Exploratory analysis

The frequency at which individual variables were selected by the algorithm spanned the range from 0–100% of models trained in the outer loop of the nested cross-validation (Figure 4). No variable was as ubiquitous in its inclusion across all five factors than the year of admission, which had at least one of its specified base learners update 95.8–100% and 99.8–100% of additive predictors for the location and scale parameters, respectively. That said, the total scores around the time of admission were invariably selected when predicting the location parameter of their respective scores around the time of discharge. Total scores on the negative and disorganized factors were also invariably selected when predicting the location parameter of another score around the time of discharge, specifically those of the disorganized and negative factors, respectively.

However, even if a variable were to be regularly selected by the algorithm between runs and repeats of the outer loop of the nested cross-validation, the marginal effect of that variable may be inconsequential if it was used only sparingly to update the additive predictor. As illustrated in Figures 5, 6, the difference in estimates of a distribution parameter between two values of any binary variable is typically smaller than that between two values of any numeric variable

TABLE 2 Pseudo R^2 of boosted beta regression models trained to predict symptom severity around the time of discharge as quantified by total scores on five different factors of the Positive and Negative Syndrome Scale (PANSS).

Factor	Pseudo R^2 on all data	Cross-validated pseudo R^2 on training data	Cross-validated pseudo R^2 on testing data
Positive	0.23	0.25	0.20
Negative	0.45	0.46	0.41
Disorganized	0.44	0.45	0.40
Excited	0.21	0.22	0.20
Depressed	0.24	0.24	0.21

Each model was trained to predict the total score on only one of the five factors. Pseudo R^2 was defined as the squared correlation coefficient between observed and predicted values. The first column of values contains in-sample estimates of pseudo R^2 from models trained on the full data set ($N = 320$). The cross-validated estimates in the other two columns represent the mean pseudo R^2 on training or testing data sets partitioned by 100-times 10-fold cross-validation in the outer loop of the nested cross-validation.

TABLE 3 Root mean squared error (RMSE) of boosted beta regression models trained to predict symptom severity around the time of discharge as quantified by total scores on five different factors of the Positive and Negative Syndrome Scale (PANSS).

Factor	RMSE on all data	Cross-validated RMSE on training data	Cross-validated RMSE on testing data
Positive	0.72	0.70	0.72
Negative	0.62	0.62	0.64
Disorganized	0.56	0.55	0.57
Excited	0.90	0.90	0.92
Depressed	0.67	0.67	0.69

Outcomes were rescaled such that the ensuing values were constrained to the interval (0,1) and approximated the proportion of the maximum attainable score on a given factor of the PANSS. RMSE was therefore on the scale of this rescaled and logit-transformed outcome. The first column of values contains in-sample estimates of RMSE from models trained on the full data set ($N = 320$). The cross-validated estimates in the other two columns represent the mean RMSE on training or testing data sets partitioned by 100-times 10-fold cross-validation in the outer loop of the nested cross-validation.

taken over the range of observations. These discrepancies in marginal effect are also apparent in the partial dependence plots generated from models trained on the full data sets following the discretization of predicted beta distributions (Figures 7–11). The fitted probability mass functions tended to be most varied in form when adjusting the value of numeric predictors. Beyond the sizeable effect that previous scores have on the prediction of future scores, the effect that disorganized symptoms around the time of admission have on negative symptoms around the time of discharge is also readily appreciable on inspection of the partial dependence plots and straightforward in its interpretation as it has a comparatively larger effect on the location parameter than the scale parameter.

Discussion

Expectations for the performance of any prediction model should consider potential limitations inherent to the data set. First, the

current data used to predict psychosis at discharge are taken from a relatively short period in time taken weeks or months in advance of the discharge date. For that reason, any semblance of an early clinical response to clozapine suggestive of a more favorable trajectory (47) would have gone undetected because there is no follow-up assessment, save for the one around discharge, with which to compare. Second, the self-imposed restriction to use only reliable and readily available information found in most medical records means that the scope of the data is largely limited to routine assessments; more complex potential predictors, such as MRI brain scans, biomarkers or genetic data as might be gathered in a research study were not available. Thus, this restriction, though a potential hindrance to model performance, was deliberately introduced in hopes of maximizing the utility of these models in the setting of routine clinical practice. Third, individuals responsible for scoring the PANSS were not blinded to the treatments received at the program. Inter-rater reliability was also not assessed in this study. That said, the performance of each rater was reviewed annually using standardized interviews to ensure that ratings on individual items were all within one point of the standard score (19).

Efforts were made to adhere to the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (48), but the decision to use boosted beta regression precluded conventional estimation of confidence intervals and *p*-values. Though it has been established that cross-validated estimates on testing data reflect the expected performance of the model fitting procedure on hypothetical data sets sampled from a distribution identical to that of the observations, work aimed at quantifying the confidence interval of these estimates is still in progress and has yet to be extended beyond linear models (49). Even so, the degree of optimism associated with boosted beta regression as depicted in Figure 2 can be helpful in visualizing the spread of values around cross-validated estimates of pseudo R^2 and RMSE because the corresponding estimates on training data were largely consistent between runs.

Insofar as it is possible to compare performance across different outcomes, it would appear that the negative and disorganized factors of the PANSS are more amenable to prediction using data that were readily available around the time of admission. Heterogeneity in treatment response may explain the comparatively poor performance on the positive factor. While the latent variable corresponding to positive symptoms has previously been shown to be the most improved following treatment at this tertiary program (19), it is important not to conflate the response of the group with that of the individual. Owing to the clinical phenomenon that is ultra-treatment-resistant psychosis, clozapine may prove beneficial in the majority of patients yet still be unable to alleviate symptoms to an acceptable degree in approximately 40% of those meeting the criteria for treatment-resistant schizophrenia (50–53). Unless the models are trained using data capable of distinguishing clozapine responders from non-responders, the ability to predict total scores on the positive factor will be limited given the integral role of this antipsychotic at the tertiary program (17), despite its multitude of side-effects (54–57). Unfortunately, differences between these two groups have been elusive. It is unlikely that this situation would be much improved had biomarkers been available as much of the research is in its infancy (58).

That said, two clinicodemographic characteristics were associated with a sustained response to clozapine monotherapy in a retrospective chart review of patients who had previously taken or were currently
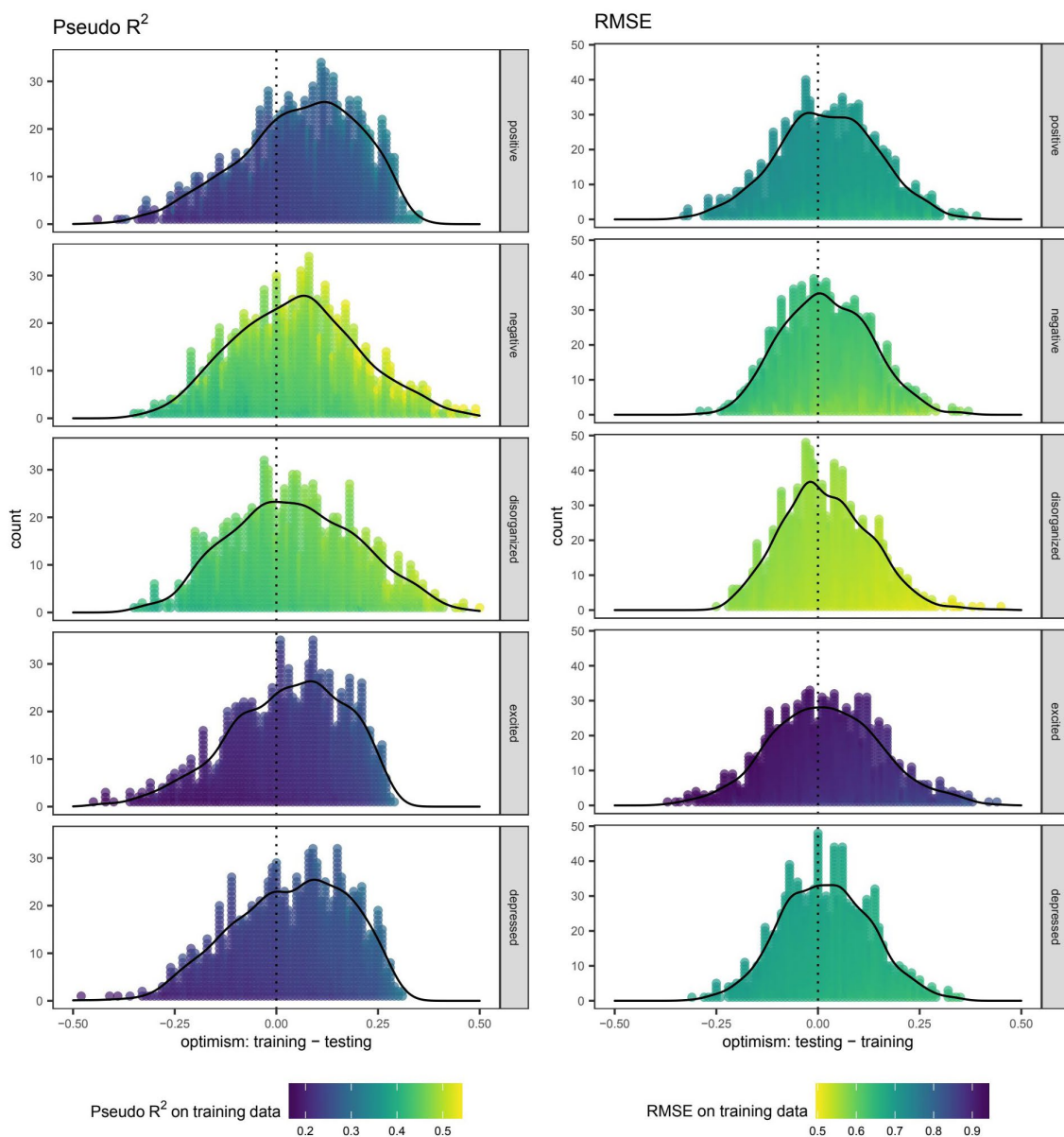
FIGURE 2
Optimism in the performance of boosted beta regression in predicting symptom severity around the time of discharge as quantified by total scores on five different factors of the Positive and Negative Syndrome Scale (PANSS). Internal validation was accomplished by nested cross-validation. Each point depicts the difference (i.e., optimism) in pseudo R^2 or root mean squared error (RMSE) when using the same boosted beta regression model trained in the outer loop to make predictions on the training and testing data sets. Outcomes were rescaled such that the ensuing values were constrained to the interval (0,1) and approximated the proportion of the maximum attainable score on a given factor of the PANSS. RMSE was therefore on the scale of this rescaled and logit-transformed outcome.

taking this antipsychotic for treatment-resistant schizophrenia or schizoaffective disorder: (1) a shorter delay to the initiation of treatment and (2) fewer hospitalizations over that same period of time (59). This first association appears to be particularly well established since the same research group had previously found longer delays to be associated with poorer outcomes in a systematic review of four other retrospective studies on the subject (60). Unfortunately, neither characteristic was readily available for this analysis because they were either missing or recalled with questionable reliability with no alternate means for verification. Even if they were available, there is no guarantee that their inclusion would have greatly improved the

performance of the models because a majority of patients (58.1%) had already tried clozapine before their admission. Considering the reason for their referral, it is likely that a sizeable proportion of this majority would have already met the criteria for ultra-resistant-treatment psychosis at admission if they had otherwise tolerated this antipsychotic well in the past. Having a prior history of clozapine use, let alone a history of non-response, has been grounds for exclusion from studies examining potential predictors of response to clozapine monotherapy. The overall lack of research involving this patient population also explains the continued reliance on expert consensus when the augmentation of clozapine monotherapy is necessary (61).
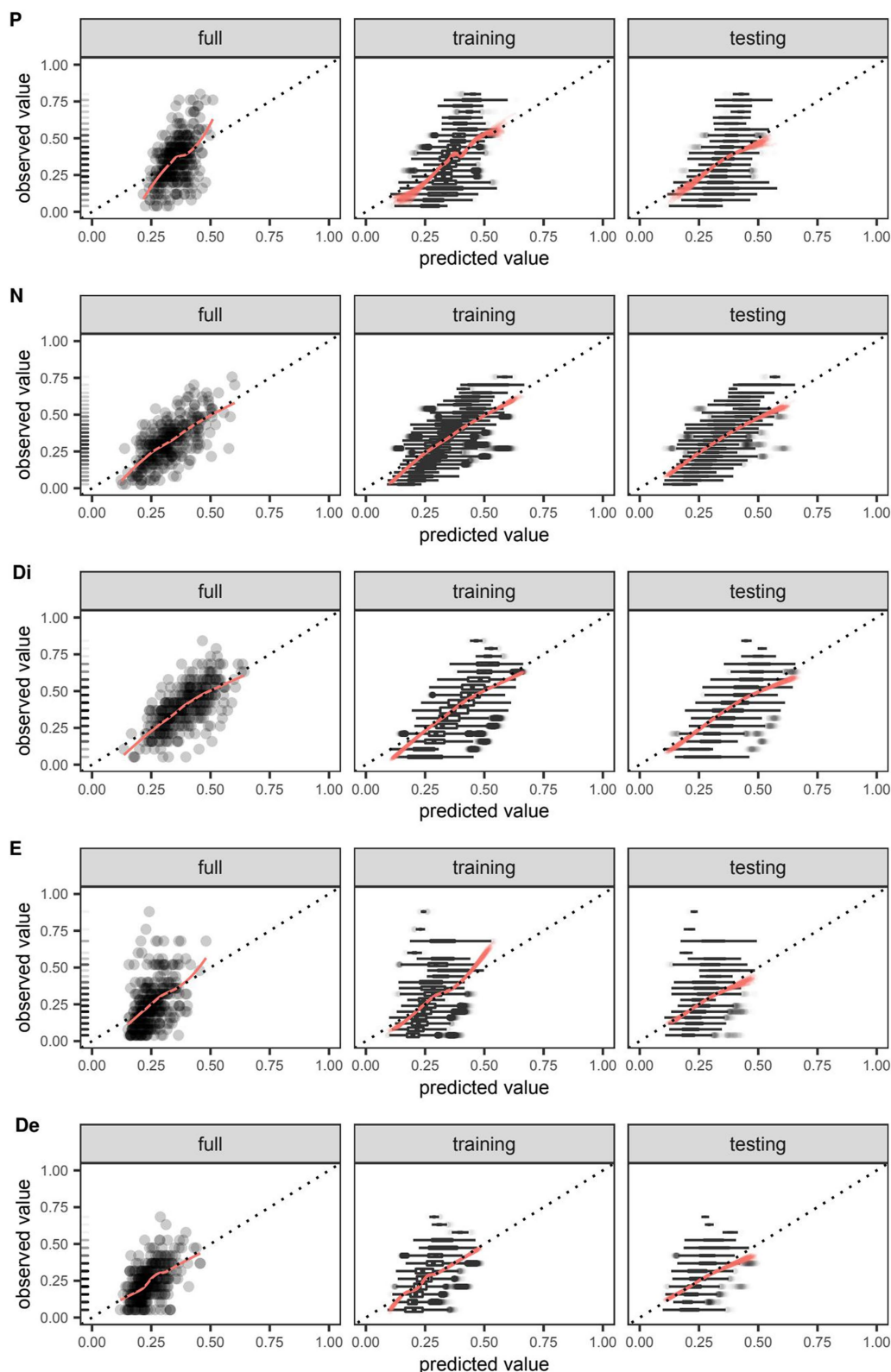
FIGURE 3
Agreement between observed and predicted values of symptom severity around the time of discharge as quantified by total scores on five different factors of the Positive and Negative Syndrome Scale (PANSS) and predicted by boosted beta regression. Outcomes were rescaled such that the ensuing values were constrained to the interval (0,1) and approximated the proportion of the maximum attainable score on a given factor of the PANSS. For models trained on the full data set and those trained in the outer loop of the nested cross-validation, the extent of agreement between observed and predicted values on the positive (P), negative (N), disorganized (Di), excited (E), and depressed (De) factors were visualized using smooth red curves fitted by locally weighted least squares regression and compared against the dotted identity line. Out-of-sample calibration was uniquely assessed in models trained on the outer loop using the testing data set. Boxplots of observed values were plotted to enhance visualization in cases where it was not feasible to plot individual data points. Rug plots added to the leftmost figure in each row depict the relative frequency of each observed value in the full data set.
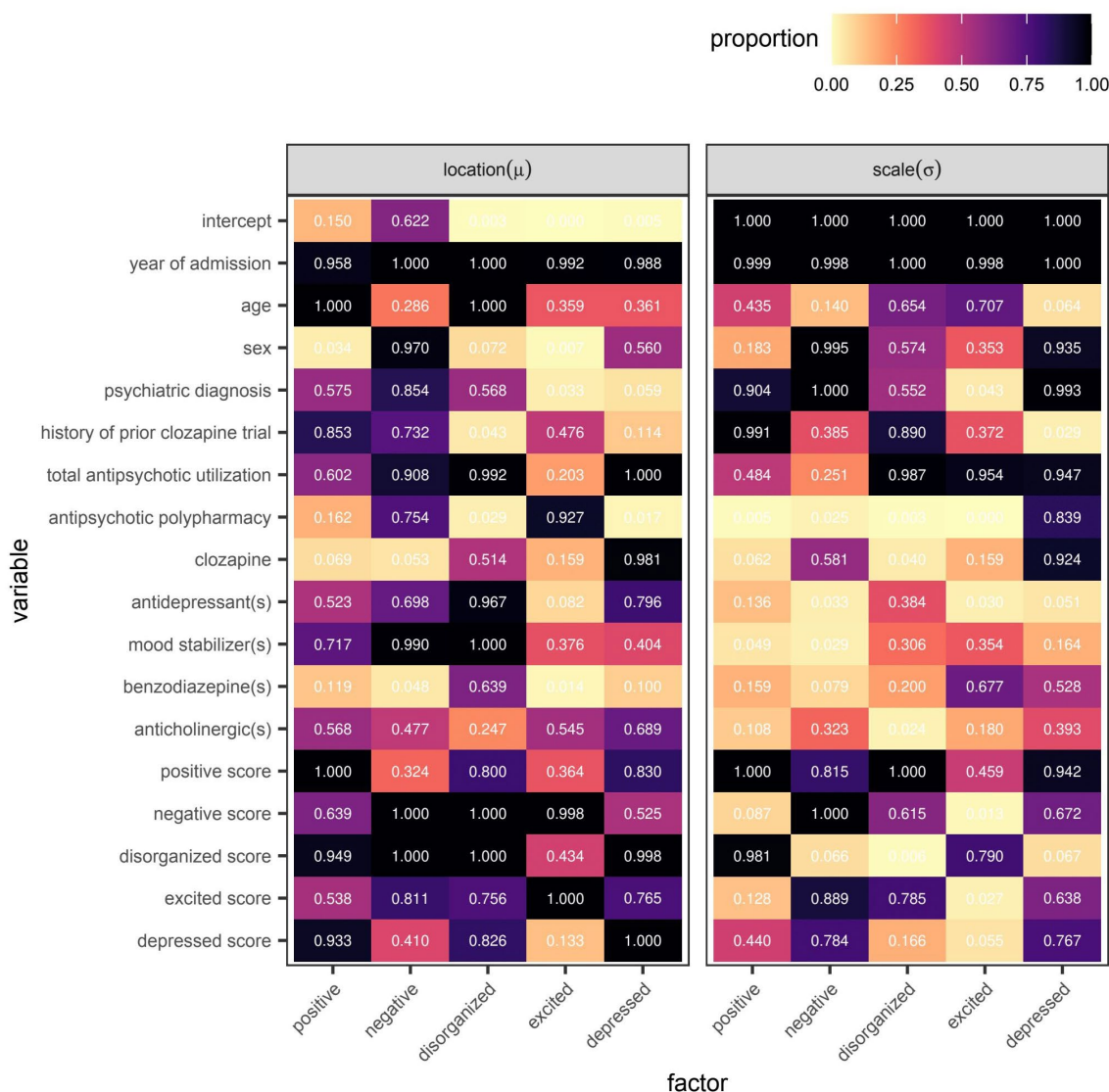
FIGURE 4

variable	location(μ) positive	negative	disorganized	excited	depressed	scale(σ) positive	negative	disorganized	excited	depressed
intercept	0.150	0.622	0.003	0.006	0.005	1.000	1.000	1.000	1.000	1.000
year of admission	0.958	1.000	1.000	0.992	0.988	0.999	0.998	1.000	0.998	1.000
age	1.000	0.286	1.000	0.359	0.361	0.435	0.140	0.654	0.707	0.064
sex	0.034	0.970	0.072	0.007	0.560	0.183	0.995	0.574	0.353	0.935
psychiatric diagnosis	0.575	0.854	0.568	0.033	0.059	0.904	1.000	0.552	0.043	0.993
history of prior clozapine trial	0.853	0.732	0.043	0.476	0.114	0.991	0.385	0.890	0.372	0.029
total antipsychotic utilization	0.602	0.908	0.992	0.203	1.000	0.484	0.251	0.987	0.954	0.947
antipsychotic polypharmacy	0.162	0.754	0.029	0.927	0.017	0.005	0.025	0.003	0.000	0.839
clozapine	0.069	0.053	0.514	0.159	0.981	0.062	0.581	0.040	0.159	0.924
antidepressant(s)	0.523	0.698	0.967	0.082	0.796	0.136	0.033	0.384	0.030	0.051
mood stabilizer(s)	0.717	0.990	1.000	0.376	0.404	0.049	0.029	0.306	0.354	0.164
benzodiazepine(s)	0.119	0.048	0.639	0.014	0.100	0.159	0.079	0.200	0.677	0.528
anticholinergic(s)	0.568	0.477	0.247	0.545	0.689	0.108	0.323	0.024	0.180	0.393
positive score	1.000	0.324	0.800	0.364	0.830	1.000	0.815	1.000	0.459	0.942
negative score	0.639	1.000	1.000	0.998	0.525	0.087	1.000	0.615	0.013	0.672
disorganized score	0.949	1.000	1.000	0.434	0.998	0.981	0.066	0.000	0.790	0.067
excited score	0.538	0.811	0.756	1.000	0.765	0.128	0.889	0.785	0.027	0.638
depressed score	0.933	0.410	0.826	0.133	1.000	0.440	0.784	0.166	0.055	0.767

proportion: 0.00 0.25 0.50 0.75 1.00

factor

Variable selection frequency within the outer loop of the nested cross-validation for boosted beta regression models predicting symptom severity around the time of discharge as quantified by total scores on five different factors of the Positive and Negative Syndrome Scale (PANSS). Variable selection frequency was defined as the proportion of models re-fitted in the outer loop of the nested cross-validation in which the base learner or base learners specified for a given variable were used to update the additive predictor of interest at least once before termination of the algorithm.

Interestingly, of the 119 patients in the aforementioned chart review who were retrospectively classified as clozapine non-responders at the first time point (i.e., during an assessment made within the first 2 years of treatment initiation), 15 (12.6%) were classified as clozapine responders at the second time point (i.e., during the most recent assessment in which clozapine was still prescribed to the patient) (59). Conversely, of the 122 patients classified as clozapine responders at the first time point, 20 (16.4%) were classified as clozapine non-responders at the second time point. Together, these two trajectories give the impression that ultra-treatment resistance is potentially reversible, but may also present later in individuals currently responding to clozapine monotherapy. Thus, while it may appear reasonable to expect individuals with ultra-treatment-resistant psychosis at admission to require treatment augmentation at discharge and hence have a higher symptom burden relative to those adequately treated with clozapine

alone, this is unlikely to be true in all cases. Limited sample sizes prevented meaningful assessment of the differences that may exist between stable non-responders and converted responders. Again, further research in this area is required to improve prognostication.

Should the above explanation for the comparatively poor performance of the model predicting the total score on the positive factor be correct, it then raises the question of whether this heterogeneity in treatment response also extends to the excited and depressed factors. The therapeutic benefits of antipsychotics may not be restricted to the improvement of positive symptoms as has been previously speculated. Based on the results of a post-hoc analysis, improvement on a general second-order factor may underlie improvement on each of the five factors of the PANSS in patients not previously known to have either treatment-resistant or ultra-treatment-resistant psychosis (62). If this hierarchical structure of symptom
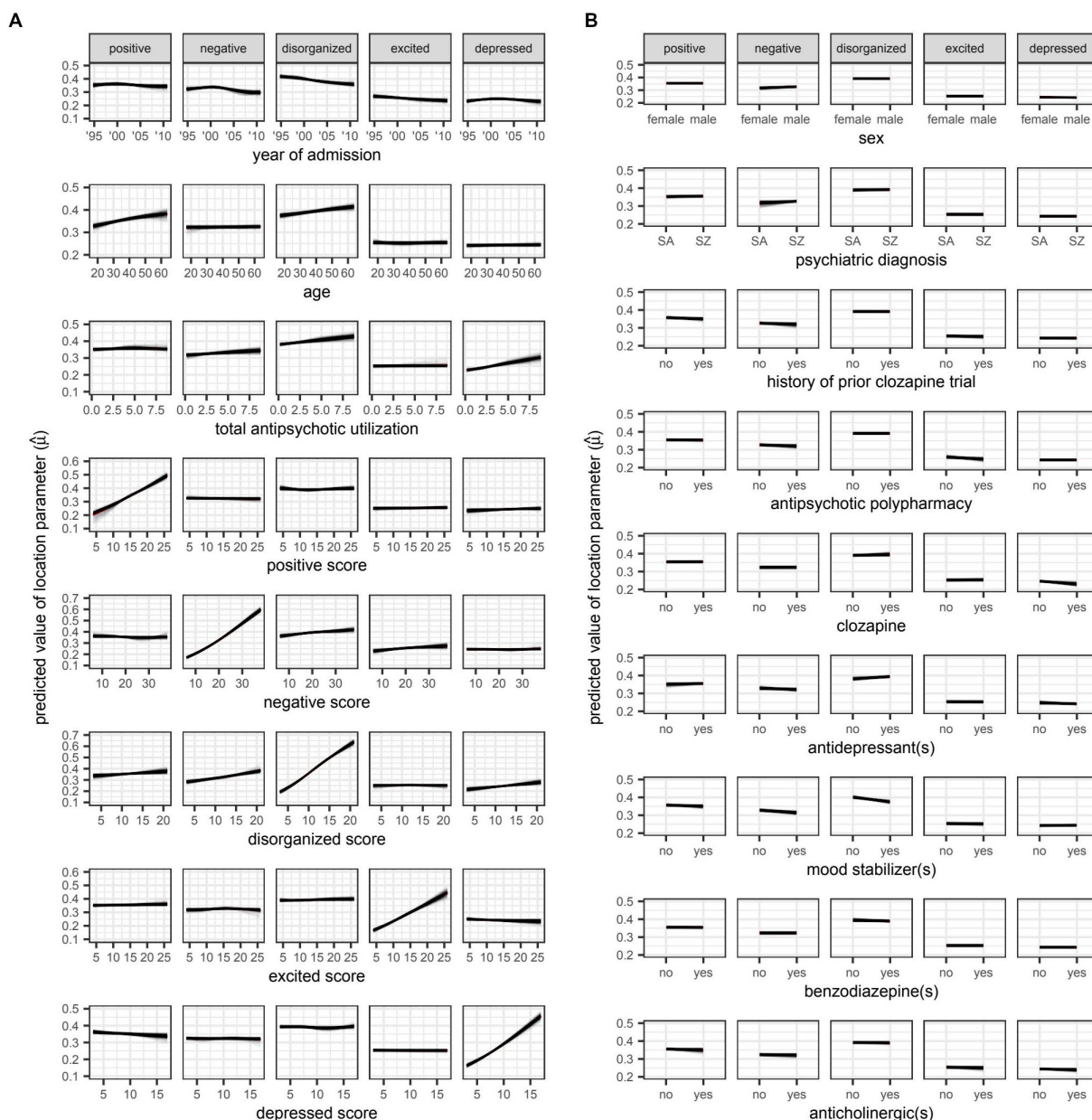
FIGURE 5
Plots for **A**) numeric and **B**) variables, respectively. SA: schizoaffective disorder. SZ: schizophrenia. Partial dependence plots for candidate predictors of the location parameter [(μ)] of beta distributions reflecting symptom severity around the time of discharge as quantified by total scores on five different factors of the Positive and Negative Syndrome Scale (PANSS). Outcomes were rescaled such that the ensuing values were constrained to the interval (0,1) and approximated the proportion of the maximum attainable score on a given factor of the PANSS. Locally weighted least squares regression was used to fit smooth curves for each model trained in the outer loop of the nested cross-validation. The red curve represents the average predicted value across all models at a given value of the candidate predictor. The location parameter is equal to the expected value of the beta distribution (i.e., $E(X) = \mu$).

improvement is also applicable in the setting of treatment-resistant and ultra-treatment-resistant psychosis, then it stands to reason that the prediction of total scores on the other four factors could be similarly impacted. To our knowledge, the generalizability of this finding to the patient population of interest has yet to be formally tested. Furthermore, an additional explanation would be needed to reconcile the apparent difficulty in predicting treatment response with the seemingly better predictions obtained on the negative and disorganized factors of the PANSS. Perhaps response to clozapine can be predicted to some extent given the inclusion of age and baseline negative symptom severity,

albeit on the five-factor rather than the three-factor PANSS (63), but the positive, excited, and depressed factors load less strongly on the general second-order factor, resulting in a weaker correlation between the first-and second-order latent variables.

The decision to predict the total scores on each of the five factors of the PANSS is uncommon in that models are more frequently trained to predict response in refractory psychosis. However, a 20% reduction in symptom severity from baseline values may not even begin to reflect the smallest increment of change that can be routinely detected by clinicians (64). In
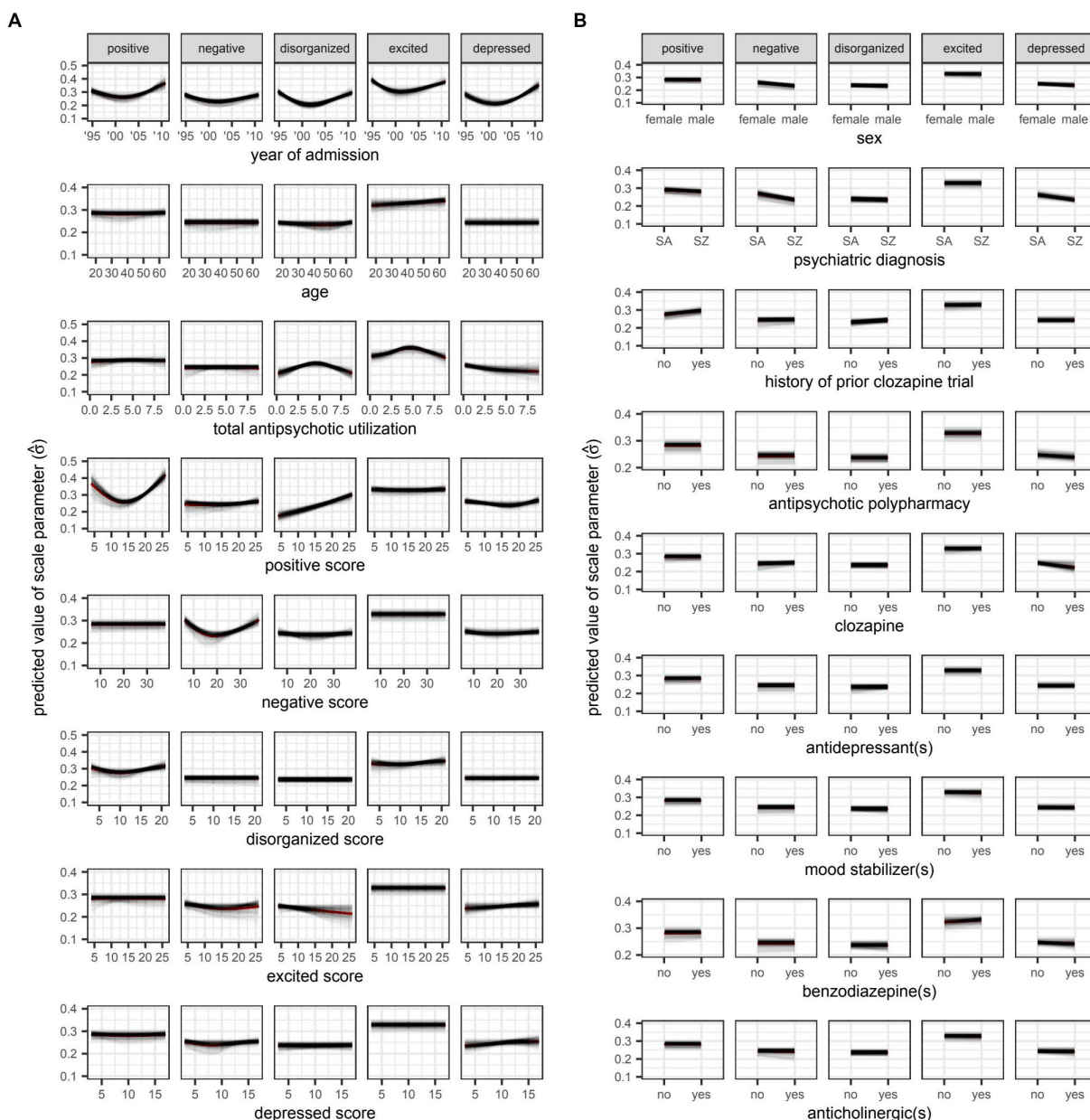
FIGURE 6
Plots for **A**) numeric and **B**) variables, respectively. SA: schizoaffective disorder. SZ: schizophrenia. Partial dependence plots for candidate predictors of the scale parameter [(σ)] of beta distributions reflecting symptom severity around the time of discharge as quantified by total scores on five different factors of the Positive and Negative Syndrome Scale (PANSS). Outcomes were rescaled such that the ensuing values were constrained to the interval (0,1) and approximated the proportion of the maximum attainable score on a given factor of the PANSS. Locally weighted least squares regression was used to fit smooth curves for each model trained in the outer loop of the nested cross-validation. The red curve represents the average predicted value across all models at a given value of the candidate predictor. The square of the scale parameter is proportional to the variance of the beta distribution (i.e., $\mathrm{Var}(X) = \mu(1-\mu)\sigma^2$).

addition, a focus on this threshold discourages a more thorough investigation into the underlying dimensions of psychotic disorders. Should the prediction of individual symptoms become even more important given the growing interest in applying network analysis to mental disorders (65), the current approach could be used to predict scores on individual items instead.

Lastly, it is worth reiterating that the primary objective of this study was to train models for prediction rather than inference. For that reason, caution must be exercised when interpreting the variables selected by the non-cyclic algorithm. Variables may be inconsistently

selected within the refitted models of the nested cross-validation because of algorithmic instability or the presence of mutually correlated variables (66, 67). Stated differently, the relationships between predictors and the outcome reflect association do not imply causality. Thus, the goal was never to create models to guide decision-making, but to identify individuals likely to have more severe symptoms around the time of discharge, allowing for an earlier and more thorough assessment of their needs at the earliest time point possible.

In summary, patient-level responses to personalized treatments offered at a tertiary-care program for treatment-resistant psychosis
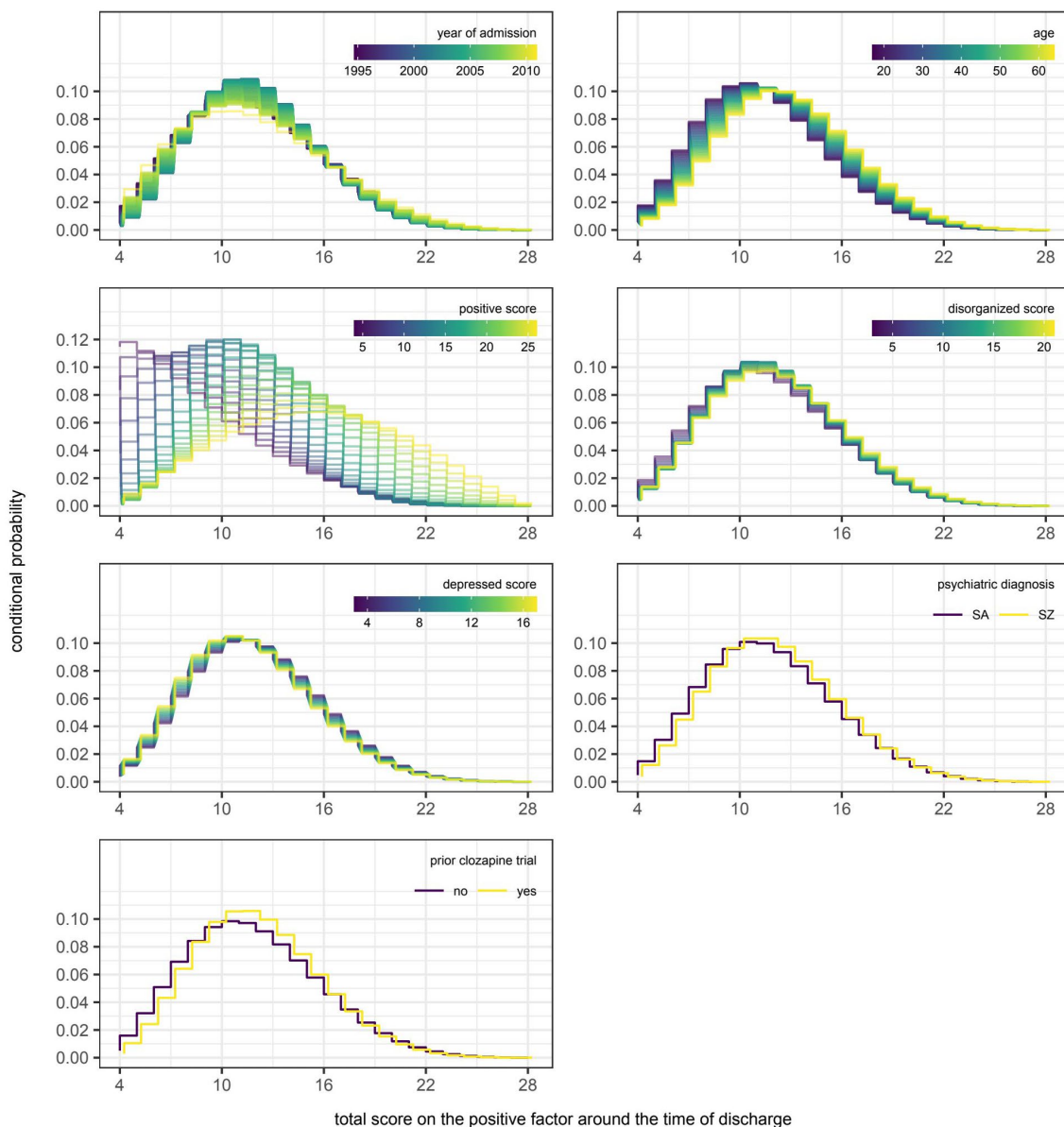
FIGURE 7
Partial dependence plots for variables included in the boosted beta regression model predicting the total score on the positive factor of the Positive and Negative Syndrome Scale (PANSS) around the time of discharge. The boosted beta regression model was trained on all available data (N=320). The plotting of conditional probabilities required discretization of the predicted beta distributions. Individual items of the PANSS assigned to the positive factor include delusions (P1), hallucinatory behavior (P3), grandiosity (P5), and unusual thought content (G9). The minimum and maximum theoretical scores were therefore 4 and 28, respectively.

were approximated by the total scores on individual factors of the PANSS around the time of discharge. These values were predicted using a novel approach that better lends itself to the modelling of integer scores from Likert scales commonly encountered in psychiatry. Safeguards in the form of data resampling methods were put in place to mitigate the risk of overfitting the models to the noise in the training data. This proved successful since the average optimism in predictive ability was kept to a minimum when comparing the performance on data that were used to train each model to that on data held out from this procedure. Candidate predictors derived from clinicodemographic information and prescription drug data were of varying and occasionally

inconsistent prognostic value. That said, symptom severity around the time of admission proved most influential in estimating the corresponding value around the time of discharge.

Few investigations into the prognostication of outcomes in treatment-resistant psychosis exist, let alone those employing techniques traditionally associated with machine learning (68). Thus, ongoing work is required should precision medicine ever have a hope of becoming the standard of practice in psychiatry. Progress toward this endeavour may be helped by the adoption of statistical tools adapted to the handling Likert scales given the loss of information when converting raw scores into categories for the sole purpose of classification (e.g., responder versus non-responder). The approach
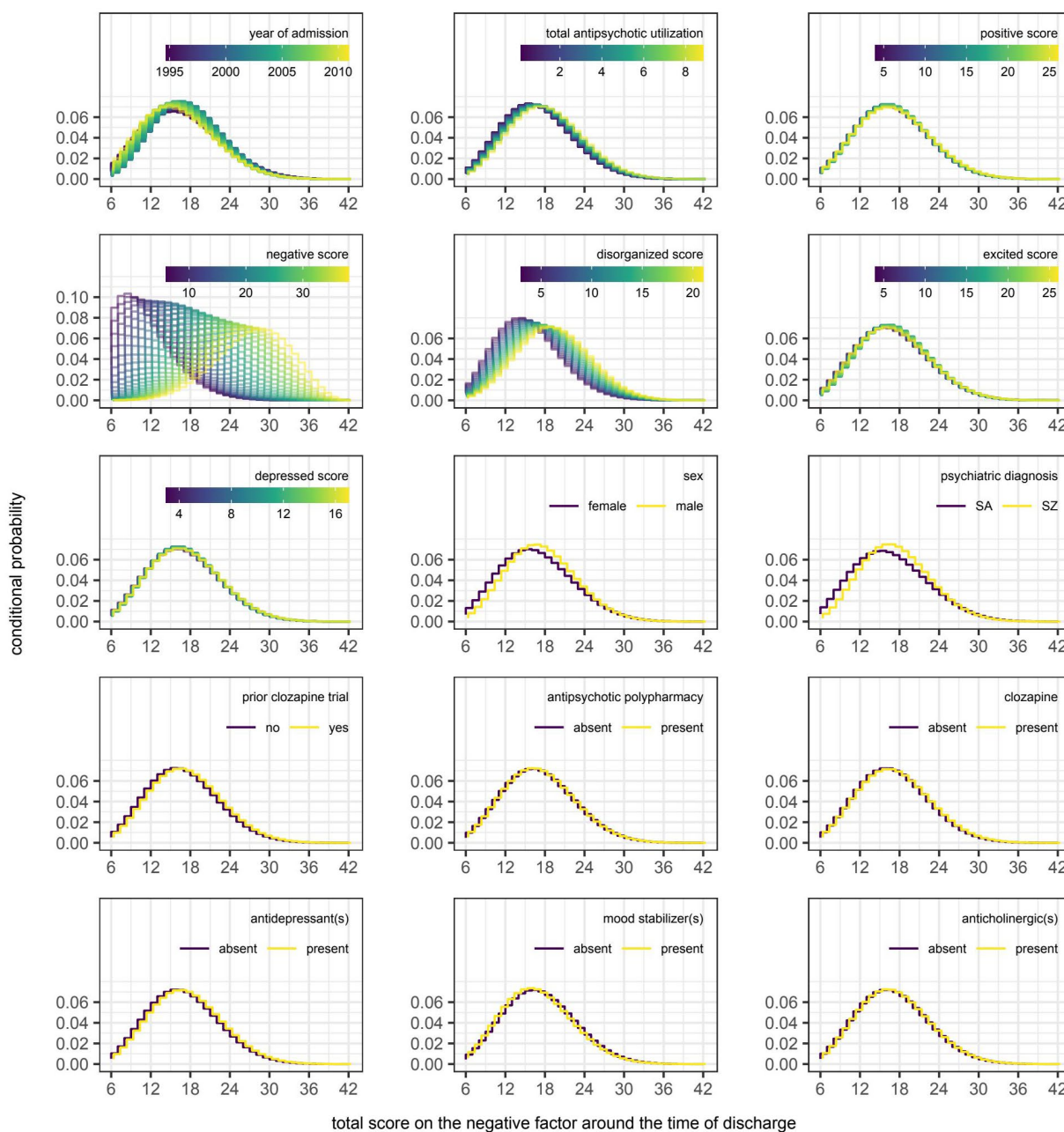
FIGURE 8
Partial dependence plots for variables included in the boosted beta regression model predicting the total score on the negative factor of the Positive and Negative Syndrome Scale (PANSS) around the time of discharge. The boosted beta regression model was trained on all available data (N=320). The plotting of conditional probabilities required discretization of the predicted beta distributions. Individual items of the PANSS assigned to the negative factor include blunted affect (N1), emotional withdrawal (N2), poor rapport (N3), passive-apathetic social withdrawal (N4), lack of spontaneity and flow of conversation (N6), and motor retardation (G7). The minimum and maximum theoretical scores were therefore 6 and 42, respectively.

detailed here represents one method in which clinical data can be used to their fullest potential.

Another key point to be emphasized from this investigation is the importance of including the PANSS in routine clinical practice – especially for complex patients maintained in a tertiary care, inpatient setting. There are multiple reasons for the clinical underutilization of this "gold standard" comprehensive measure of psychosis, which – as noted above – assesses multiple different and independent domains. Opler and colleagues (69) note that the PANSS, compared to other rating scales, has "many items, evaluates a multidimensional array of symptoms … and involves the use of data from patient reports, caregiver reports, and clinical observations. Consequently, the PANSS takes up more time during training and requires a greater amount of time for one to master it compared to many other instruments." Similarly, Østergaard and colleagues (70) emphasize the time required to complete the PANSS, as well as challenges administering the scale to severely ill patients. Nevertheless, the comprehensive profile of psychosis in patients provided by the PANSS would suggest that it
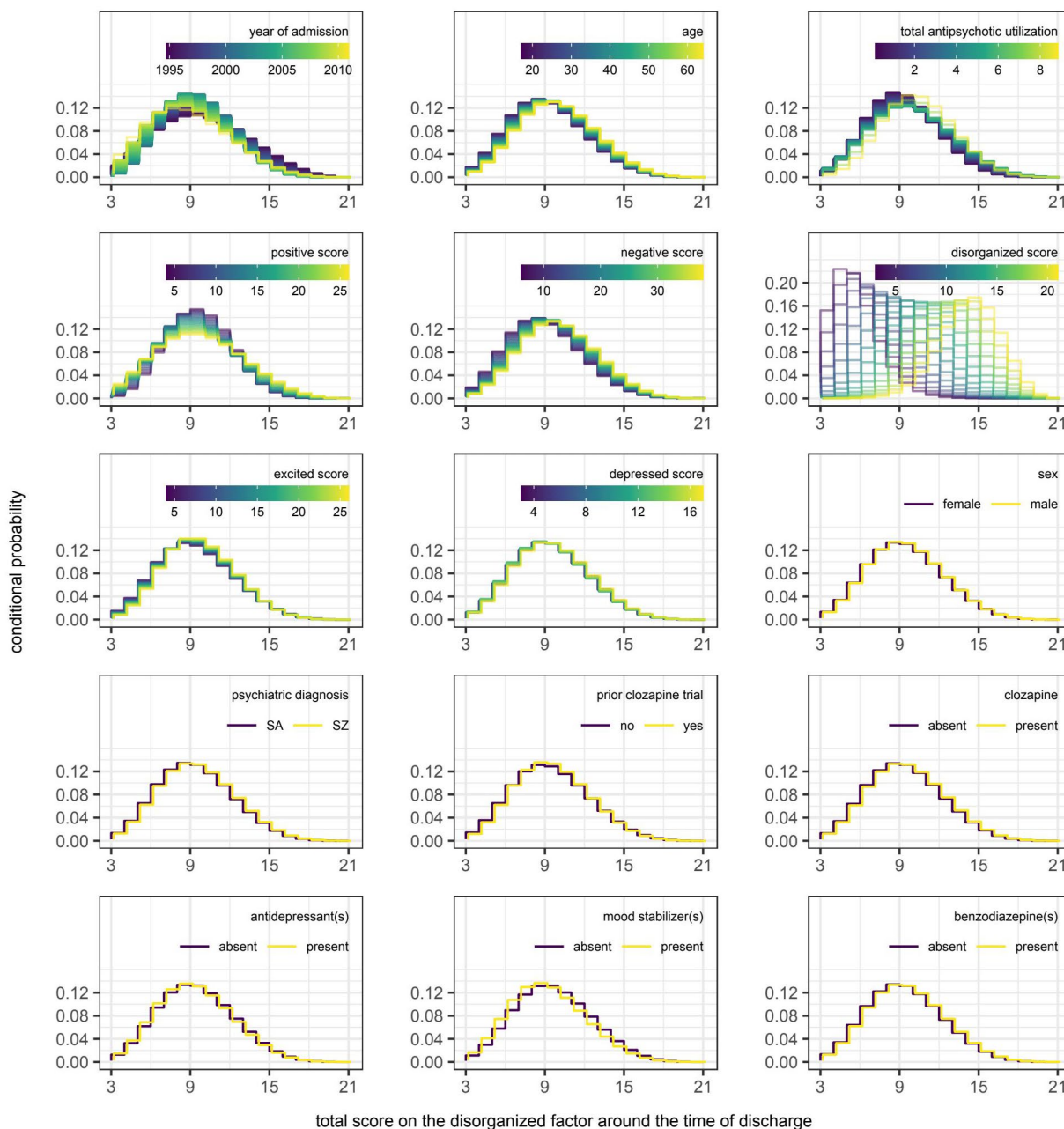
FIGURE 9
Partial dependence plots for variables included in the boosted beta regression model predicting the total score on the disorganized factor of the Positive and Negative Syndrome Scale (PANSS) around the time of discharge. The boosted beta regression model was trained on all available data (*N*=320). The plotting of conditional probabilities required discretization of the predicted beta distributions. Individual items of the PANSS assigned to the disorganized factor include conceptual disorganization (P2), difficulty in abstract thinking (N5), and poor attention (G11). The minimum and maximum theoretical scores were therefore 3 and 21, respectively.

should be used more frequently in clinical practice, which may require programs to implement specialized training for the scale (69) or consider using an abbreviated version, such as the PANSS-6 (70).

Conclusion

Boosted beta regression—selected for its ease in modelling non-linear effects with P-spline base learners as well as its ability to perform shrinkage estimation and variable selection—showed promise in predicting symptom severity around the time of discharge from a tertiary care program for treatment-resistant psychosis. Correlation between predictions and observations were highest for the negative and disorganized factors of the PANSS when using only reasonable reliable and readily available information around the time of admission. Future studies may benefit from a richer training data set with even more clinicodemographic variables.
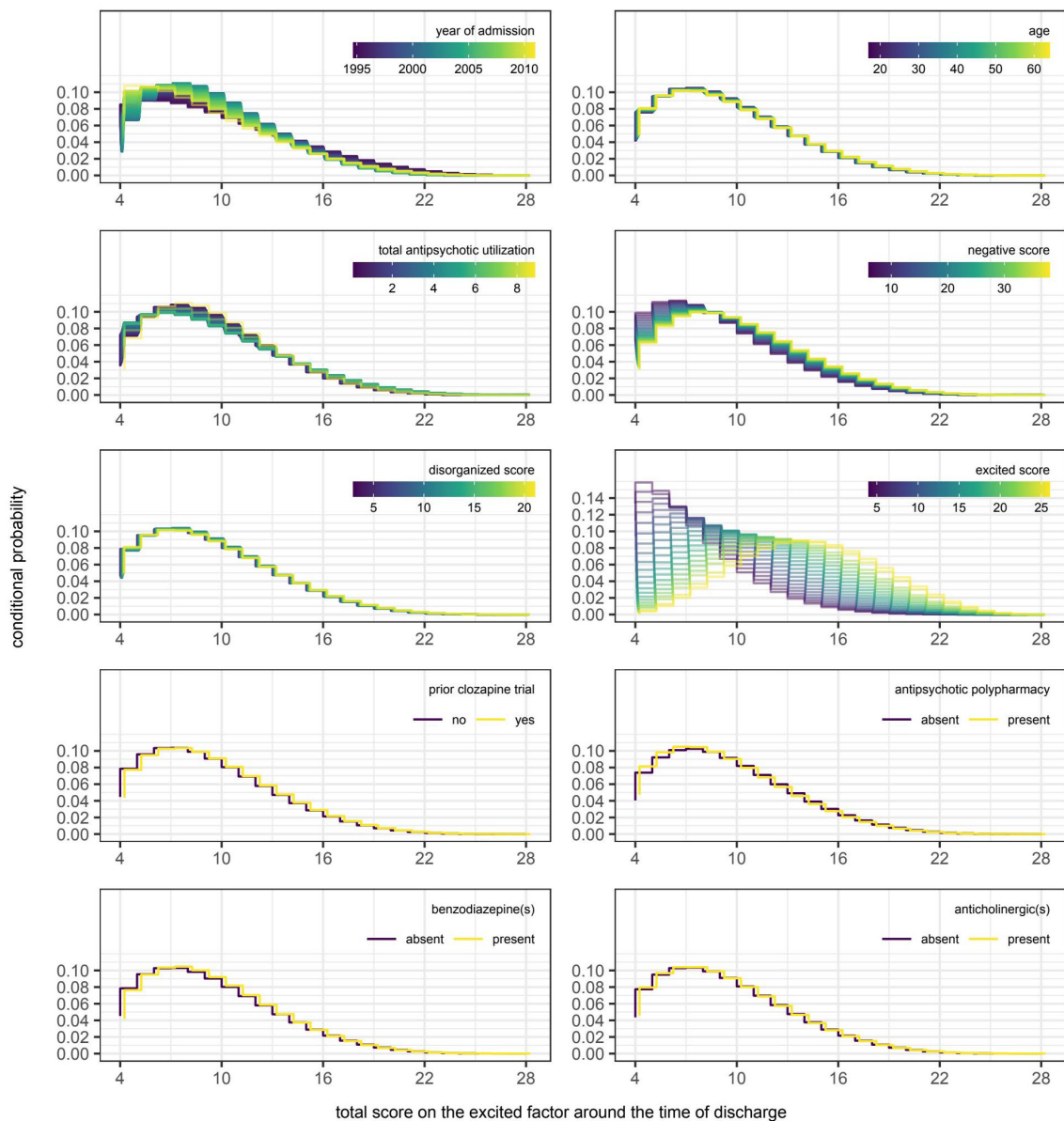
FIGURE 10
Partial dependence plots for variables included in the boosted beta regression model predicting the total score on the excited factor of the Positive and Negative Syndrome Scale (PANSS) around the time of discharge. The boosted beta regression model was trained on all available data (N=320). The plotting of conditional probabilities required discretization of the predicted beta distributions. Individual items of the PANSS assigned to the excited factor include excitement (P4), hostility (P7), uncooperativeness (G8), and poor impulse control (G14). The minimum and maximum theoretical scores were therefore 4 and 28, respectively.

Data availability statement

The datasets presented in this article are not readily available because Data underlying the study are not available due to restrictions imposed by the University of British Colombia Clinical Research Ethics Board. Data are stored at the British Columbia Psychosis Program (BCPP) in Vancouver, Canada, which is an inpatient medical unit of the BC Provincial Health Services Authority (PHSA). Data are available to PHSA employees who have had required background checks, as well as affiliated members of the BCPP, pending addition to the study protocol. As the data are not fully anonymized, total free access to data by all is not permitted. Requests to access the datasets should be directed to al.barr@ubc.ca

Ethics statement

The studies involving human participants were reviewed and approved by the University of British Columbia. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

Author contributions

LL was responsible for data collection, statistical analysis, and writing the first draft of the manuscript. AB was responsible for study
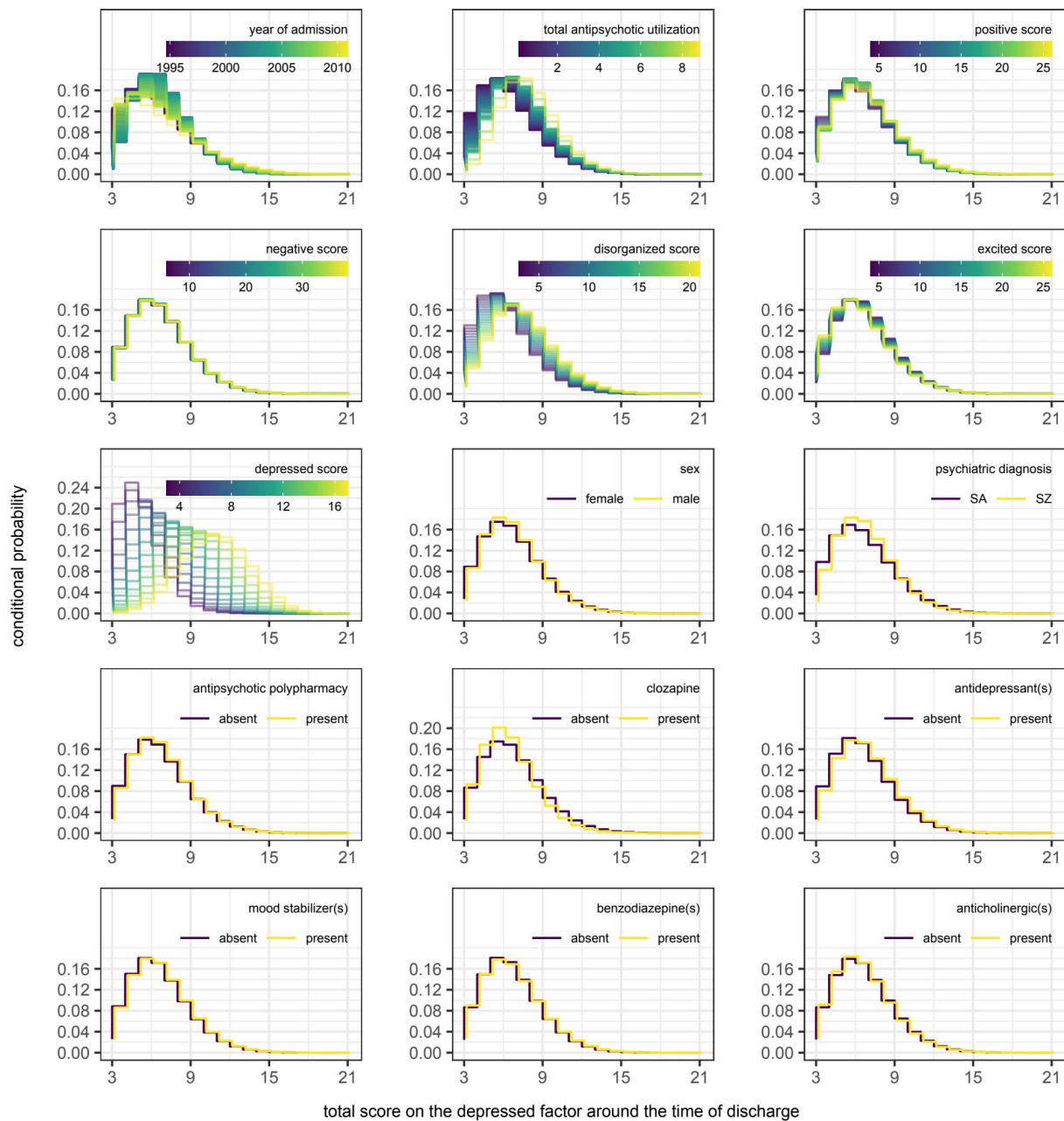
FIGURE 11
Partial dependence plots for variables included in the boosted beta regression model predicting the total score on the depressed factor of the Positive and Negative Syndrome Scale (PANSS) around the time of discharge. The boosted beta regression model was trained on all available data (*N*=320). The plotting of conditional probabilities required discretization of the predicted beta distributions. Individual items of the PANSS assigned to the depressed factor include anxiety (G2), guilt feelings (G3), and depression (G6). The minimum and maximum theoretical scores were therefore 3 and 21, respectively.

design, REB approval, and supervision. WH was responsible for securing project funds. All authors contributed to the article and approved the submitted version.

Conflict of interest

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Ramos-Miguel A, Barr AM, Honer WG. Spines, synapses, and schizophrenia. *Biol Psychiatry*. (2015) 78:741–3. doi: 10.1016/j.biopsych.2015.08.035

2. Ramos-Miguel A, Beasley CL, Dwork AJ, Mann JJ, Rosoklija G, Barr AM, et al. Increased SNARE protein–protein interactions in orbitofrontal and anterior cingulate cortices in schizophrenia. *Biol Psychiatry*. (2015) 78:361–73. doi: 10.1016/j.biopsych.2014.12.012

3. Avram M, Rogg H, Korda A, Andreou C, Müller F, Borgwardt S. Bridging the gap? Altered thalamocortical connectivity in psychotic and psychedelic states. *Front Psych*. (2021) 12:706017. doi: 10.3389/fpsyt.2021.706017

4. Berdenis van Berlekom A, Muflihah CH, Snijders G, MacGillavry HD, Middeldorp J, Hol EM, et al. Synapse pathology in schizophrenia: a meta-analysis of postsynaptic elements in postmortem brain studies. *Schizophr Bull*. (2020) 46:374–86. doi: 10.1093/schbul/sbz060

5. Canu E, Agosta F, Filippi M. A selective review of structural connectivity abnormalities of schizophrenic patients at different stages of the disease. *Schizophr Res*. (2015) 161:19–28. doi: 10.1016/j.schres.2014.05.020

6. Siskind D, Orr S, Sinha S, Yu O, Brijball B, Warren N, et al. Rates of treatment-resistant schizophrenia from first-episode cohorts: systematic review and meta-analysis. *Br J Psychiatry*. (2022) 220:115–20. doi: 10.1192/bjp.2021.61

7. Lian L, Kim DD, Procyshyn RM, Cázares D, Honer WG, Barr AM. Long-acting injectable antipsychotics for early psychosis: a comprehensive systematic review. *PLoS One*. (2022) 17:e0267808. doi: 10.1371/journal.pone.0267808

8. Lian L, Kim DD, Procyshyn RM, Fredrikson DH, Cázares D, Honer WG, et al. Efficacy of long-acting injectable versus oral antipsychotic drugs in early psychosis: a systematic review and meta-analysis. *Early Interv Psychiatry*. (2021) 16:589–99. doi: 10.1111/eip.13202

9. Chiliza B, Asmal L, Kilian S, Phahladira L, Emsley R. Rate and predictors of non-response to first-line antipsychotic treatment in first-episode schizophrenia. *Hum Psychopharmacol*. (2015) 30:173–82. doi: 10.1002/hup.2469

10. Smart SE, Kępińska AP, Murray RM, MacCabe JH. Predictors of treatment resistant schizophrenia: a systematic review of prospective observational studies. *Psychol Med*. (2021) 51:44–53. doi: 10.1017/s0033291719002083

11. Howes OD, McCutcheon R, Agid O, de Bartolomeis A, van Beveren NJ, Birnbaum ML, et al. Treatment-resistant schizophrenia: treatment response and resistance in psychosis (TRRIP) working group consensus guidelines on diagnosis and terminology. *Am J Psychiatry*. (2017) 174:216–29. doi: 10.1176/appi.ajp.2016.16050503

12. Siskind D, McCartney L, Goldschlager R, Kisely S. Clozapine v. first-and second-generation antipsychotics in treatment-refractory schizophrenia: systematic review and meta-analysis. *Br J Psychiatry*. (2016) 209:385–92. doi: 10.1192/bjp.bp.115.177261

13. Sarkar SN, Tracy DK, Fernandez M-JM, Nalesnik N, Dhillon G, Onwumere J, et al. Unheard voices: outcomes of tertiary care for treatment-refractory psychosis. *Psychiatr Bull*. (2014) 38:71–4. doi: 10.1192/pb.bp.112.042598

14. Sims S, Hepsipa Omega Juliet S, Joseph J, Gopal S, Raghavan V, Venkatraman L, et al. Acceptability of peer support for people with schizophrenia in Chennai, India: a cross sectional study amongst people with lived experience, caregivers, and mental health professionals. *Front Psychiatry*. (2022) 13:797427. doi: 10.3389/fpsyt.2022.797427

15. Lindenmayer JP, Khan A, Wance D, Maccabee N, Kaushik S, Kaushik S. Outcome evaluation of a structured educational wellness program in patients with severe mental illness. *J Clin Psychiatry*. (2009) 70:1385–96. doi: 10.4088/JCP.08m04740yel

16. White C, Nash L, Karageorge A, van der Pol R, Hunt GE, Hamilton B, et al. Bridging the gap: a new integrated early intervention service for young people with complex mental health issues. *Early Interv Psychiatry*. (2022) 16:186–94. doi: 10.1111/eip.13152

17. Lee LHN, Procyshyn RM, White RF, Woodward TS, Honer WG, Barr AM. Antipsychotic prescribing patterns on admission to and at discharge from a tertiary care program for treatment-resistant psychosis. *PLoS One*. (2018) 13:e0199758. doi: 10.1371/journal.pone.0199758

18. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. (1987) 13:261–76. doi: 10.1093/schbul/13.2.261

19. Woodward TS, Jung K, Smith GN, Hwang H, Barr AM, Procyshyn RM, et al. Symptom changes in five dimensions of the positive and negative syndrome scale in refractory psychosis. *Eur Arch Psychiatry Clin Neurosci*. (2014) 264:673–82. doi: 10.1007/s00406-013-0460-x

20. Tse L, Schwarz SK, Bowering JB, Moore RL, Barr AM. Incidence of and risk factors for delirium after cardiac surgery at a quaternary care center: a retrospective cohort study. *J Cardiothorac Vasc Anesth*. (2015) 29:1472–9. doi: 10.1053/j.jvca.2015.06.018

21. Tse L, Bowering JB, Schwarz SK, Moore RL, Burns KD, Barr AM. Postoperative delirium following transcatheter aortic valve implantation: a historical cohort study. *Can J Anaesth*. (2015) 62:22–30. doi: 10.1007/s12630-014-0254-2

22. Wickham H, Averick M, Bryan J, Chang WDML, Francois R, et al. Welcome to the Tidyverse. *J Open Source Softw*. (2019) 4:1686. doi: 10.21105/joss.01686

23. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. *J R Stat Soc C*. (2005) 54:507–54. doi: 10.1111/j.1467-9876.2005.00510.x

24. Hofner B, Mayr A, Robinzonov N, Schmid M. Model-based boosting in R: a hands-on tutorial using the R package mboost. *Comput Stat*. (2014) 29:3–35. doi: 10.1007/s00180-012-0382-5

25. Kuhn M, Wickham H. *Rsample: general resampling infrastructure. R package version 0.0.4*. (2019)

26. Wickham H. *Ggplot 2: elegant graphics for data analysis*. New York, NY: Springer (2016).

27. Garnier S, Ross N, Rudis B, Sciaini M, Scherer C. *Viridis: default color maps from 'matplotlib'. R package version* (2018)

28. Wilke CO. *Cowplot: streamlined plot theme and plot annotations for "ggplot2". R package version* (2019)

29. Chang W. Extrafont: tools for using fonts. R package version 0.17, Available at: https://CRAN.R-projectorg/package=extrafont; (2014)

30. Schmid M, Wickler F, Maloney KO, Mitchell R, Fenske N, Mayr A. Boosted Beta regression. *PLoS One*. (2013) 8:e61623. doi: 10.1371/journal.pone.0061623

31. Hofner B, Mayr A, Fenske N, Thomas J, Schmid M. gamboostLSS: Boosting methods for 'GAMLSS' (2020) (Accessed April, 2 2020). Available at: https://cran.r-project.org/web/packages/gamboostLSS/index.html.

32. Wallwork R, Fortgang R, Hashimoto R, Weinberger D, Dickinson D. Searching for a consensus five-factor model of the positive and negative syndrome scale for schizophrenia. *Schizophr Res*. (2012) 137:246–50. doi: 10.1016/j.schres.2012.01.031

33. Kieschnick R, McCullough BD. Regression analysis of variates observed on (0,1): percentages, proportions and fractions. *Stat Model*. (2003) 3:193–213. doi: 10.1191/1471082X03st053oa

34. Ferrari S, Cribari-Neto F. Beta regression for modelling rates and proportions. *J Appl Stat*. (2004) 31:799–815. doi: 10.1080/0266476042000214501

35. Taverne C, Lambert P. Inflated discrete beta regression models for Likert and discrete rating scale outcomes. arXiv arXiv: 14054637 (2014)

36. Schmid M, Hothorn T. Boosting additive models using component-wise P-splines. *Comput Stat Data Anal*. (2008) 53:298–311. doi: 10.1016/j.csda.2008.09.009

37. Thomas J, Mayr A, Bischl B, Schmid M, Smith A, Hofner B. Gradient boosting for distributional regression: faster tuning and improved variable selection via noncyclical updates. *Stat Comput*. (2018) 28:673–87. doi: 10.1007/s11222-017-9754-6

38. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. (2018) 14:91–118. doi: 10.1146/annurev-clinpsy-032816-045037

39. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinf*. (2006) 7:91. doi: 10.1186/1471-2105-7-91

40. World Health Organization. ATC/DDD Index (2020) (Accessed December 16, 2019, September 2020). Available at: https://www.whocc.no/atc_ddd_index/.

41. Leising J, Barr AM, Procyshyn RM, Ainsworth NJ, White RF, Honer W, et al. High-dose fluvoxamine augmentation to clozapine in treatment-resistant psychosis. *J Clin Psychopharmacol*. (2021) 41:186–90. doi: 10.1097/jcp.0000000000001342

42. Hofner B, Hothorn T, Kneib T, Schmid M. A framework for unbiased model selection based on boosting. *J Comput Graph Stat*. (2011) 20:956–71. doi: 10.1198/jcgs.2011.09220

43. Mayr A, Weinhold L, Hofner B, Titze S, Gefeller O, Schmid M. The betaboost package—a software tool for modelling bounded outcome variables in potentially high-dimensional epidemiological data. *Int J Epidemiol*. (2018) 47:1383–8. doi: 10.1093/ije/dyy093

44. Gneiting T, Balabdaoui F, Raftery AE. Probabilistic forecasts, calibration, and sharpness. *J R Stat Soc B*. (2007) 69:243–68. doi: 10.1111/j.1467-9868.2007.00587.x

45. Gneiting T, Katzfuss M. Probabilistic forecasting. *Ann Rev Stat Appl*. (2014) 1:125–51. doi: 10.1146/annurev-statistics-062713-085831

46. Greenwell BM. PDP: an R package for constructing partial dependence plots. *R J*. (2017) 9:421. doi: 10.32614/RJ-2017-016

47. Semiz UB, Cetin M, Basoglu C, Ebrinc S, Uzun O, Herken H, et al. Clinical predictors of therapeutic response to clozapine in a sample of Turkish patients with treatment-resistant schizophrenia. *Prog Neuro-Psychopharmacol Biol Psychiatry*. (2007) 31:1330–6. doi: 10.1016/j.pnpbp.2007.06.002

48. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. (2015) 102:148–58. doi: 10.1002/bjs.9736

49. Bates S, Hastie T, Tibshirani R. *Cross-validation: what does it estimate and how well does it do it?*: 210400673 (2021) doi: 10.48550/arXiv.2104.00673

50. Siskind D, Siskind V, Kisely S. *Clozapine response rates among people with treatment-resistant schizophrenia: data from a systematic review and meta-analysis*. SAGE Publications Sage CA: Los Angeles, CA; (2017), 62, 772–777

51. Vasiliu O. Third-generation antipsychotics in patients with schizophrenia and non-responsivity or intolerance to clozapine regimen: what is the evidence? *Front Psych*. (2022) 13:1069432. doi: 10.3389/fpsyt.2022.1069432

52. Correll CU, Howes OD. Treatment-resistant schizophrenia: definition, predictors, and therapy options. *J Clin Psychiatry*. (2021) 82:MY20096AH1C. doi: 10.4088/jcp.my20096ah1c

53. Seppälä A, Pylvänäinen J, Lehtiniemi H, Hirvonen N, Corripio I, Koponen H, et al. Predictors of response to pharmacological treatments in treatment-resistant schizophrenia – a systematic review and meta-analysis. *Schizophr Res*. (2021) 236:123–34. doi: 10.1016/j.schres.2021.08.005

54. Yuen JWY, Kim DD, Procyshyn RM, Panenka WJ, Honer WG, Barr AM. A focused review of the metabolic side-effects of clozapine. *Front Endocrinol*. (2021) 12:609240. doi: 10.3389/fendo.2021.609240

55. Yuen JWY, Kim DD, Procyshyn RM, White RF, Honer WG, Barr AM. Clozapine-induced cardiovascular side effects and autonomic dysfunction: a systematic review. *Front Neurosci*. (2018) 12:203. doi: 10.3389/fnins.2018.00203

56. Tse L, Procyshyn RM, Fredrikson DH, Boyda HN, Honer WG, Barr AM. Pharmacological treatment of antipsychotic-induced dyslipidemia and hypertension. *Int Clin Psychopharmacol*. (2014) 29:125–37. doi: 10.1097/YIC.0000000000000014

57. Tanzer TD, Brouard T, Pra SD, Warren N, Barras M, Kisely S, et al. Treatment strategies for clozapine-induced hypotension: a systematic review. *Therapeut Adv Psychopharmacol*. (2022) 12:20451253221092931. doi: 10.1177/20451253221092931

58. Samanaite R, Gillespie A, Sendt K-V, McQueen G, MacCabe JH, Egerton A. Biological predictors of clozapine response: a systematic review. *Front Psych*. (2018) 9:327. doi: 10.3389/fpsyt.2018.00327

59. Shah P, Iwata Y, Brown EE, Kim J, Sanches M, Takeuchi H, et al. Clozapine response trajectories and predictors of non-response in treatment-resistant schizophrenia: a chart review study. *Eur Arch Psychiatry Clin Neurosci*. (2020) 270:11–22. doi: 10.1007/s00406-019-01053-6

60. Shah P, Iwata Y, Plitman E, Brown EE, Caravaggio F, Kim J, et al. The impact of delay in clozapine initiation on treatment outcomes in patients with treatment-resistant schizophrenia: a systematic review. *Psychiatry Res*. (2018) 268:114–22. doi: 10.1016/j.psychres.2018.06.070

61. Wagner E, Kane JM, Correll CU, Howes O, Siskind D, Honer WG, et al. Clozapine combination and augmentation strategies in patients with schizophrenia—recommendations from an international expert survey among the treatment response and resistance in psychosis (TRRIP) working group. *Schizophr Bull*. (2020) 46:1459–70. doi: 10.1093/schbul/sbaa060

62. Marques TR, Levine SZ, Reichenberg A, Kahn R, Derks EM, Fleischhacker WW, et al. How antipsychotics impact the different dimensions of schizophrenia: a test of competing hypotheses. *Eur Neuropsychopharmacol*. (2014) 24:1279–88. doi: 10.1016/j.euroneuro.2014.04.001

63. Okhuijsen-Pfeifer C, Sterk A, Horn I, Terstappen J, Kahn R, Luykx J. Demographic and clinical features as predictors of clozapine response in patients with schizophrenia spectrum disorders: a systematic review and meta-analysis. *Neurosci Biobehav Rev*. (2020) 111:246–52. doi: 10.1016/j.neubiorev.2020.01.017

64. Leucht S, Kane JM, Kissling W, Hamann J, Etschel E, Engel RR. What does the PANSS mean? *Schizophr Res*. (2005) 79:231–8. doi: 10.1016/j.schres.2005.04.008

65. Esfahlani FZ, Visser K, Strauss GP, Sayama H. A network-based classification framework for predicting treatment response of schizophrenia patients. *Expert Syst Appl*. (2018) 109:152–61. doi: 10.1016/j.eswa.2018.05.005

66. Breiman L. Statistical modeling: the two cultures (with comments and a rejoinder by the author). *Stat Sci*. (2001) 16:199–231. doi: 10.1214/ss/1009213726

67. Del Giudice M. *The prediction-explanation fallacy: a pervasive problem in scientific applications of machine learning Psy ArXiv* (2021) doi: 10.31234/osf.io/4vq8f

68. Del Fabro L, Bondi E, Serio F, Maggioni E, D'Agostino A, Brambilla P. Machine learning methods to predict outcomes of pharmacological treatment in psychosis. *Transl Psychiatry*. (2023) 13:75. doi: 10.1038/s41398-023-02371-z

69. Opler MGA, Yavorsky C, Daniel DG. Positive and negative syndrome scale (PANSS) training: challenges, solutions, and future directions. *Innovat Clin Neurosci*. (2017) 14:77–81.

70. Østergaard SD, Opler MGA, Correll CU. Bridging the measurement gap between research and clinical Care in Schizophrenia: positive and negative syndrome Scale-6 (PANSS-6) and other assessments based on the simplified negative and positive symptoms interview (SNAPSI). *Innovat Clin Neurosci*. (2017) 14:68–72.