Title

Individualized Prediction of Prodromal Symptom Remission for Youth at Clinical High Risk for Psychosis.

Permalink

Journal

ISSN

Authors

Worthington, Michelle A
Addington, Jean
Bearden, Carrie E
et al.

Publication Date

DOI

Peer reviewed

Individualized Prediction of Prodromal Symptom Remission for Youth at Clinical High Risk for Psychosis

Michelle A. Worthington[1,], **Jean Addington[2]**, **Carrie E. Bearden[3,]**, **Kristin S. Cadenhead[4,]**, **Barbara A. Cornblatt[5]**, **Matcheri Keshavan[6]**, **Daniel H. Mathalon[7]**, **Thomas H. McGlashan[8]**, **Diana O. Perkins[9]**, **William S. Stone[6,]**, **Ming T. Tsuang[4]**, **Elaine F. Walker[10,]**, **Scott W. Woods[8]**, and **Tyrone D. Cannon*[,1,8]**

[1]Department of Psychology, Yale University, New Haven, CT, USA; [2]Department of Psychiatry, Hotchkiss Brain Institute, Calgary, AB, Canada; [3]Department of Psychiatry and Biobehavioral Sciences and Psychology, UCLA, Los Angeles, CA, USA; [4]Department of Psychiatry, UCSD, San Diego, CA, USA; [5]Department of Psychiatry, Zucker Hillside Hospital, Long Island, NY, USA; [6]Department of Psychiatry, Harvard Medical School at Beth Israel Deaconess Medical Center, Boston, MA, USA; [7]Department of Psychiatry, UCSF, and SFVA Medical Center, San Francisco, CA, USA; [8]Department of Psychiatry, Yale University, New Haven, CT, USA; [9]Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA; [10]Department of Psychology and Psychiatry, Emory University, Atlanta, GA, USA

*To whom correspondence should be addressed; 2 Hillhouse Avenue, PO Box 208205, New Haven, CT 06511, USA; tel: 203-436-1545, fax: 203-432-5281, e-mail: tyrone.cannon@yale.edu

The clinical high-risk period before a first episode of psychosis (CHR-P) has been widely studied with the goal of understanding the development of psychosis; however, less attention has been paid to the 75%–80% of CHR-P individuals who do not transition to psychosis. It is an open question whether multivariable models could be developed to predict remission outcomes at the same level of performance and generalizability as those that predict conversion to psychosis. Participants were drawn from the North American Prodrome Longitudinal Study (NAPLS3). An empirically derived set of clinical and demographic predictor variables were selected with elastic net regularization and were included in a gradient boosting machine algorithm to predict prodromal symptom remission. The predictive model was tested in a comparably sized independent sample (NAPLS2). The classification algorithm developed in NAPLS3 achieved an area under the curve of 0.66 (0.60–0.72) with a sensitivity of 0.68 and specificity of 0.53 when tested in an independent external sample (NAPLS2). Overall, future remitters had lower baseline prodromal symptoms than nonremitters. This study is the first to use a data-driven machine-learning approach to assess clinical and demographic predictors of symptomatic remission in individuals who do not convert to psychosis. The predictive power of the models in this study suggest that remission represents a unique clinical phenomenon. Further study is warranted to best understand factors contributing to resilience and recovery from the CHR-P state.

Key words: remission/clinical high risk/schizophrenia/psychosis/risk prediction/machine learning

Introduction

The clinical high risk for psychosis (CHR-P) paradigm is widely used to study predictors of psychosis. Most CHR-P individuals are ascertained due to a recent onset or worsening of attenuated psychotic symptoms; these individuals also show moderate to severe impairments in social and role functioning.[1] Well performing and generalizable multivariable "risk calculator" prediction models for outcomes of psychosis among CHR-P samples have been developed and validated.[2–7] However, less attention has been paid to the approximately 75%–80% of CHR-P individuals who do not transition to psychosis.[8,9] This group experiences heterogeneous outcomes, not simply representing the inverse of conversion.[10–12] Specifically, while most nonconverting CHR-P individuals continue to experience some positive symptoms or remain socially, cognitively, or functionally impaired, some appear to show a remission in symptoms, wherein all positive symptoms are rated below the prodromal threshold.[13] It is an open question whether multivariable models could be developed to predict remission outcomes as distinct from other outcomes in nonconverters[14] at the same level of performance and generalizability as those for prediction of conversion to psychosis.

Only a handful of studies have examined the clinical characteristics of CHR-P remitters. In the North American Prodrome Longitudinal Study (NAPLS2), group-based multitrajectory modeling identified a group that exhibited improvement in all symptom domains and functioning and a high likelihood of symptomatic and/or

functional remission and 2 other groups that exhibited either moderate or no improvement in symptoms and functioning and a lower likelihood of remission.[15] In other studies, remitters had better neurocognitive functioning at baseline than nonremitters in domains of attention, verbal memory, verbal fluency, and immediate visual memory, and nonremitters may even show continued deterioration in domains of semantic fluency and speed of processing as compared to remitters.[13,16,17] Overall, those who do not convert to psychosis still seem to be more impaired and symptomatic than healthy controls, though some do show a trajectory consistent with a reduction of symptom severity below the prodromal threshold.[13,18]

Ultimately, a better understanding of the characteristics and outcomes of individuals who remit from the CHR-P syndrome could inform treatment programs, as remission would be a more proximal goal for interventions administered during the CHR-P state than prevention of conversion to psychosis.[19] Further, the ability to predict whether an individual might remit spontaneously or with usual and customary treatment could be useful for randomization in clinical trials of stratified interventions; such cases might be excluded or filtered into a different intervention group to improve sensitivity of the trial to differentiate effects of interventions.

The present study takes an empirical machine-learning approach to predicting symptom-defined remission as an outcome of the CHR-P syndrome. While this approach has had some success when applied to predicting conversion to psychosis,[4,5] negative symptoms,[20] and psychiatric treatment response,[21–23] it is novel in the context of predicting remission from the CHR-P state. We developed predictive models using a machine-learning classification algorithm and tested the validity/generalizability of models discovered in NAPLS3 in a comparably sized and fully independent sample (NAPLS2).

Methods

Participants

NAPLS3 The discovery sample included CHR-P participants from a 9-site observational consortium study that aims to identify predictors and mechanisms related to conversion to psychosis.[24] Participants were individuals aged 12–30 meeting criteria for a psychosis risk syndrome as determined by the Criteria of Prodromal States[25] and as assessed by the Structured Interview for Psychosis-risk Syndromes (SIPS).[26,27] Exclusion criteria included any current or lifetime DSM-IV[28] diagnosis of a psychotic disorder, IQ <70, the presence of a neurological disorder, or psychosis-risk symptoms caused by another Axis I disorder. Study visits occurred every 2 months for the first 8 months of the study, and at 12, 18, and 24 months. After 24 months, the longer-term follow-up period included phone-based checkups at 6-month intervals up to 48 months.

NAPLS2 As a test of replication, the models developed in the discovery sample (NAPLS3) were tested in a completely independent sample that had been gathered in the prior iteration of the NAPLS study (NAPLS2).[29] This wave includes data collected at 8 study sites—all sites in NAPLS3 except UCSF—between 2008 and 2012, and the samples are independent and nonoverlapping with respect to NAPLS3. NAPLS2 included participants aged 12–35. All other inclusion and exclusion criteria were the same. Study visits in NAPLS2 occurred every 6 months for the duration of the 2-year follow-up period. We took a complete cases approach to the present analysis, as the methods used are not robust to missing data, and participants were excluded from either sample if baseline measures were incomplete. Those excluded did not differ from those included in regards to total baseline positive symptoms and remission status (see eTable 8 for a full comparison of demographic variables between excluded and included participants). All participants in both NAPLS2 and NAPLS3 provided written informed consent after receiving a complete description of the study.

Participants from both NAPLS2 and NAPLS3 provided written informed consent for the study. The protocol and consent forms for each study were approved by the institutional review boards at each site.

Clinical Assessments

Clinical assessments were administered every 6 months or at the time of conversion to psychosis. Measures included the SIPS, the Scale of Psychosis-Risk Symptoms (SOPS),[30] the Calgary Depression Scale for Schizophrenia (CDSS),[31] the Brief Assessment of Cognition in Schizophrenia (BACS) Symbol Coding (SC),[32] the Hopkins Verbal Learning Test-Revised (HVLT-R),[33] the Global Assessment of Functioning scale (GAF),[34] the Global Functioning Social scale (GFS), and the Global Functioning Role scale (GFR).[35] Demographic information including age, sex, race, ethnicity, years of education, and parental educational attainment were collected at baseline.

Remission Participants met criteria for remission if they scored below the "prodromal risk" threshold on each of the positive symptom subscale items on the SOPS.[36] Participants achieving symptom ratings below this threshold may still experience mild forms of positive symptoms (eg, hypervigilance in the absence of danger, unusual superstitions) but do not experience significant distress or impairment due to these experiences.[25] For the present study, we calculated "sustained remission" according to whether participants met criteria for remission at either 6, 12, or 18 months and continued to meet remission criteria at the next consecutive 6-month study visit. Nonremitters included those who never met criteria for sustained remission and those who transitioned to psychosis.

Statistical Analysis

All statistical analyses were performed in R version 4.4.0.[37] Packages used in this analysis include glmnet,[38] smotefamily,[39] and caret.[40]

Feature Selection   Although a large number of clinical and demographic variables were available to include in the predictive model, it was unlikely that all of these variables would contribute significantly in the final model. Thus, as a first step, we performed a feature selection method to identify a smaller set of predictor variables that may be important in predicting remission. An elastic net regularization method was used to select no more than 10 variables from a set of 40 available baseline clinical and demographic variables representing the primary clinical dimensions of interest (eg, attenuated psychotic symptoms, affective symptoms, neurocognitive functioning, global/social/role functioning, and demographics). These included the individual items from all 4 subscales on the SOPS, change in GAF, GFR, and GFS over the past year, BACS-SC score, HVLT-R score, all individual items from the CDSS, age, sex, race, ethnicity, educational attainment, and parental educational attainment. This set of variables comprises a large portion of the clinical battery administered in the NAPLS studies and measures were included if they were administered in both studies and if most of the participants completed them in both studies. The final feature set size was limited to a maximum of 10 variables to maintain degrees of freedom in the final model given the relatively small number of remitters in the NAPLS3 discovery sample (N = 75/568). The elastic net method iteratively fits logistic regression models predicting remission status while applying a shrinkage penalty to the coefficients of highly correlated or unimportant predictor variables. During this process, the elastic net tests values ranging from 0.0001 to 0.1 for the penalization parameter *lambda* to identify the optimal coefficient shrinkage penalty for the final set of predictor variables. The elastic net also tests values ranging from 0 to 1 for the parameter *alpha* which selects the contributing weight of both the ridge (*alpha* = 0) and lasso (*alpha* = 1) penalties and ultimately helps to determine the number of predictor variables to include through the coefficient shrinkage process.[41] To limit the final feature set size, we specified in the parameters that the resulting feature set should consist of no more than 10 variables with nonzero coefficients, which we then used as the final feature set. The final parameter values included were *alpha* = 0.5 and *lambda* = 0.03. Prediction models were then built with the final set of selected predictor variables. For all predictor variables selected during the feature selection process, *t* tests comparing mean differences between remitters and nonremitters were performed for descriptive purposes and to aid in interpretation (table 2).

Classification   The predictive classification model was built using a gradient boosting machine (GBM) algorithm using the caret[40] package in R. GBM is a powerful and commonly used machine-learning algorithm which performs classification through an ensemble of trees approach. Through this approach, the algorithm builds shallow decision tree models wherein each successive tree built learns from the error of the previous tree.[42] During the training phase, the GBM model was tuned across 2 main parameters using default values included in the caret package: number of trees built (between 50 and 150) and the number of splits performed within each tree (between 1 and 3). The final parameter values used were number of trees = 50 and number of splits performed = 2. In addition, 10-fold cross-validation was used in the training process wherein the NAPLS3 data were randomly split into 10 subsets, then the model was iteratively trained on 9/10 of the folds together and tested on each left-out fold exactly once. Through this process, the model hyperparameters were tuned within each fold using a grid search technique which was optimized for classification performance as evaluated by the area under the receiver operating characteristic curve (ROC) metric of discriminability between remitters and nonremitters. Additional classification algorithms were tested for comparison in performance to GBM, the results of which are presented in eTable 7.

Balanced Sampling   Due to the class imbalance in remission outcomes, we employed a synthetic minority oversampling technique (SMOTE) in the training phase. This method creates a new dataset in which the minority class in the outcome variable is oversampled based on information from neighboring data points, thus reducing bias induced by class imbalance during model learning.[43] The SMOTE method was applied within each of the 10 training folds so as to avoid information leakage across folds in the training set.

Training and Testing Samples   GBM classification models were developed in the NAPLS3 dataset using SMOTE balanced sampling within a 10-fold cross-validation framework. These models were then tested in the NAPLS2 dataset to assess the model's generalizability to a completely independent sample which contains imbalanced outcome classes. For comparison, we also tested the model in a version of the NAPLS2 dataset which underwent the SMOTE balanced sampling technique on the remission outcome variable. This comparison provides information on the model performance where the outcome classes are balanced and are more similar to the outcome classes within the discovery sample and served to ensure replicability. Model performance was evaluated using area under the curve (AUC), sensitivity, specificity, and balanced accuracy (BAC). Beyond the AUC metric of discriminability, sensitivity assesses the proportion of actual remitters in the NAPLS2 sample who were correctly classified and specificity assesses the proportion of actual nonremitters in the NAPLS2 sample who were correctly classified. Sensitivity, specificity, and BAC were assessed using the median predicted likelihood of remitting

as the threshold for distinguishing between predicted remission or nonremission. As a secondary analysis, all models were retrained in the NAPLS2 dataset using the same training methods described above and tested in the original NAPLS3 dataset. Figure 1 provides an illustration of the entire analysis pipeline. Finally, to account for potential site differences, a leave-site-out cross-validation procedure was performed to ensure stability across recruitment sites (see eAppendix 1 in Supplement for results). All R-code developed for analysis is available upon request.

Results

A total of 568 participants from the NAPLS3 sample who had complete baseline data and who completed any follow-up assessments were included in the analysis. Of these, 75 participants (13.2%) achieved remission. From NAPLS2, 553 participants had complete baseline data and completed any follow-up assessments and were included in the replication sample. Of these, 94 participants (17%) achieved remission. There were no differences between remitters and nonremitters in either the NAPLS3 or NAPLS2 samples in terms of age, sex, race, ethnicity, years of education, or baseline GAF score (table 1). Remitters in both samples had lower levels of positive symptoms at baseline as compared to nonremitters (NAPLS3: remitters mean (SD) = 11.7 (3.63), nonremitters mean (SD) = 13.0 (3.23), P value = .003; NAPLS2: remitters mean (SD) = 9.6 (4.28), nonremitters mean (SD) = 12.3 (3.63), P value < .001). This pattern was similar even when converters were excluded from the nonremitter group (eTable 2).

The final variables selected based on the elastic net feature selection process described above were suspiciousness/persecutory ideas, trouble with focus and attention, unusual thought content/delusional ideas, impaired tolerance to normal distress, guilty ideas of reference, decreased ideational richness, motor disturbances, early wakening, and impairment in personal hygiene. The coefficients of these variables derived during elastic net regularization can be found in eTable 3 and the relative importance of these predictor variables in the final classification model predicting remission are presented in eFigure 1. This figure gives an indication of the relative decrease in accuracy of the model that would occur if the variable were removed from the model. When compared between remitters and nonremitters, all predictor variables were found to be lower in remitters as compared to nonremitters in the NAPLS3 sample, and significantly so except for unusual thought content/delusional ideas (table 2). When converters were removed from the nonremitter group, decreased ideational richness was also no longer significantly different (eTable 4).

In the original (eg, non-SMOTE-sampled) NAPLS3 discovery sample, the final GBM model achieved an AUC of 0.69 (95% DeLong[44] confidence interval [CI] = 0.63–0.75) and using the median remission likelihood of 0.4 achieved a sensitivity of 0.65 and specificity of 0.68, indicating good performance within the training data. The primary validation test of interest, however, was the model's performance in the external validation
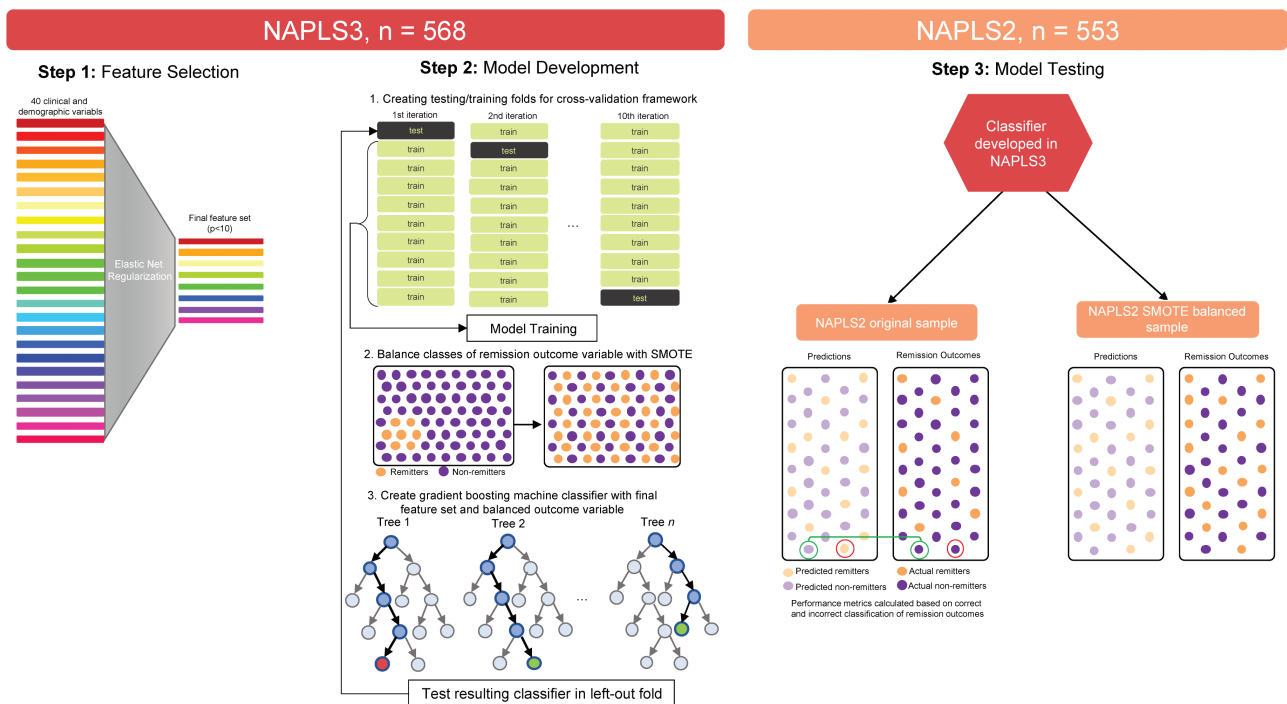

Fig. 1. Analysis pipeline for the development and validation of prediction models with remission as the outcome.

Table 1. Demographics of Symptom-Based Remitters and Nonremitters in Both the NAPLS3[a] and NAPLS2 Samples

| | NAPLS3 | | | | | | NAPLS2 | | | | | |
| | Remitters | | Nonremitters | | Test Statistic | P Value | Remitters | | Nonremitters | | Test Statistic | P Value |
	N	Mean (SD)	N	Mean (SD)	t/χ² (df)		N	Mean (SD)	N	Mean (SD)	t/χ² (df)	
Age	75	18.4 (4.06)	493	18.4 (4.02)	$t(97.3) = -0.06$	.94	94	18.3 (4.35)	459	18.6 (4.25)	$t(131.9) = 0.54$	.59
Sex, no. (%) female	75	34 (45.3%)	493	225 (45.6%)	$\chi^2(2) = 0.0025$	.96	94	34 (36%)	459	202 (44%)	$\chi^2(2) = 1.96$	.16
Race, no. (%) nonwhite	75	28 (37.3%)	493	221 (44.8%)	$\chi^2(2) = 1.49$	.22	94	42 (44.7%)	459	184 (40.1%)	$\chi^2(2) = 0.68$	.41
Ethnicity, no. (%) Hispanic	75	15 (20%)	493	98 (19.9%)	$\chi^2(2) = 0.0006$	.98	94	23 (24.5%)	459	74 (16.1%)	$\chi^2(2) = 3.76$	.052
Years of education	75	11.5 (3.16)	493	11.6 (3.09)	$t(96.7) = 0.20$	.84	94	11.4 (3.23)	459	11.3 (2.75)	$t(122.1) = -0.09$	.92
Baseline positive symptoms	75	11.7 (3.63)	493	13.0 (3.23)	$t(92.6) = 3.08$	**.003**	94	9.6 (4.28)	459	12.3 (3.63)	$t(121.9) = 5.75$	**<.001**
Baseline functioning	75	52.3 (11.8)	493	51.3 (12.29)	$t(99.9) = -0.63$	.53	94	49.5 (11.12)	459	48.4 (10.74)	$t(130.9) = -0.88$	.38

Bold values indicate significant differences between remitters and non-remitters at the significance level of $p = 0.05$.
[a]NAPLS = North American Prodrome Longitudinal Study.

Table 2. Descriptive Statistics of Predictor Variables in the NAPLS3[a] and NAPLS2 Samples That Were Found to Be Associated With Remission During the Elastic Net Regularization Feature Selection Process

| | NAPLS3 | | | | | | NAPLS2 | | | | | |
| | Remitters | | Nonremitters | | Test Statistic | P Value | | Remitters | | Nonremitters | | Test Statistic | P Value |
Variable	N	Mean (SD)	N	Mean (SD)	t/χ^2 (df)		N	Mean (SD)	N	Mean (SD)	t/χ^2 (df)	
Suspiciousness/persecutory ideas	75	2.68 (1.47)	493	3.16 (1.33)	$t(93.4) = 2.7$	**.009**	94	2.11 (1.48)	459	2.86 (1.49)	$t(134.5) = 4.5$	**<.001**
Trouble with focus/attention	75	2.23 (1.29)	493	2.64 (1.3)	$t(98.2) = 2.6$	**.01**	94	2.33 (1.28)	459	2.75 (1.24)	$t(131.4) = 2.9$	**.004**
Unusual thought content/delusional ideas	75	3.56 (1.08)	493	3.77 (1.02)	$t(95.1) = 1.6$	.11[b]	94	2.78 (1.54)	459	3.46 (1.25)	$t(119.3) = 4.1$	**<.001**
Impaired tolerance to normal stress	75	2.28 (1.83)	493	2.98 (1.89)	$t(99.3) = 3.1$	**.003**	94	2.41 (1.95)	459	2.81 (1.86)	$t(129.9) = 1.8$	.07
Guilty ideas of reference	75	0.24 (0.54)	493	0.41 (0.71)	$t(117.2) = 2.4$	**.02**	94	0.3 (0.62)	459	0.41 (0.72)	$t(148.7) = 1.6$	.12
Decreased ideational richness	75	0.88 (1.17)	493	1.23 (1.28)	$t(102.6) = 2.3$	**.02**	94	0.96 (1.24)	459	1.21 (1.33)	$t(140.3) = 1.8$	.07
Motor disturbances	75	0.52 (0.89)	493	0.97 (1.07)	$t(109.5) = 3.9$	**<.001**	94	0.67 (1.08)	459	0.9 (1.05)	$t(131.4) = 1.8$	.07
Early wakening	75	0.31 (0.72)	493	0.51 (0.88)	$t(110.6) = 2.2$	**.03**	94	0.37 (0.82)	459	0.38 (0.8)	$t(132.2) = 0.05$	.96
Impairment in personal hygiene	75	0.48 (0.91)	493	0.87 (1.27)	$t(123.2) = 3.3$	**.001**	94	0.62 (1.17)	459	0.81 (1.22)	$t(137.5) = 1.4$	.16

Bold values indicate significant differences between remitters and non-remitters at the significance level of $p = 0.05$.

[a]NAPLS = North American Prodrome Longitudinal Study.

[b]Although "unusual thought content/delusional ideas" was selected as a significant predictor of remission in the feature selection process, the group means did not differ between remitters and nonremitters. The elastic net model fits multiple variables simultaneously and may select features that account for variance in a multivariable context but do not differ in a univariate context.

Table 3. Performance Metrics of Predictive Classification Models During Testing

Model	AUC[a] (95% CI)	Sensitivity	Specificity	BAC
Discovery stage				
N3$_{SMOTE}$ → N3$_{ORIGINAL}$	0.69 (0.63–0.75)	0.65	0.68	0.66
N3$_{SMOTE}$ → N3$_{SMOTE}$	0.85 (0.82–0.87)	0.76	0.80	0.78
Validation stage				
N3$_{SMOTE}$ → N2$_{ORIGINAL}$	**0.66 (0.60–0.72)**	**0.64**	**0.59**	**0.62**
N3$_{SMOTE}$ → N2$_{SMOTE}$	0.86 (0.84–0.89)	0.84	0.59	0.72

Primary validation step of interest is bolded.
[a]AUC = area under the curve; BAC = balanced accuracy; CI = confidence interval; N2 = NAPLS2/North American Prodrome Longitudinal Study-2; N3 = NAPLS3/North American Prodrome Longitudinal Study-3; SMOTE = synthetic minority oversampling technique.
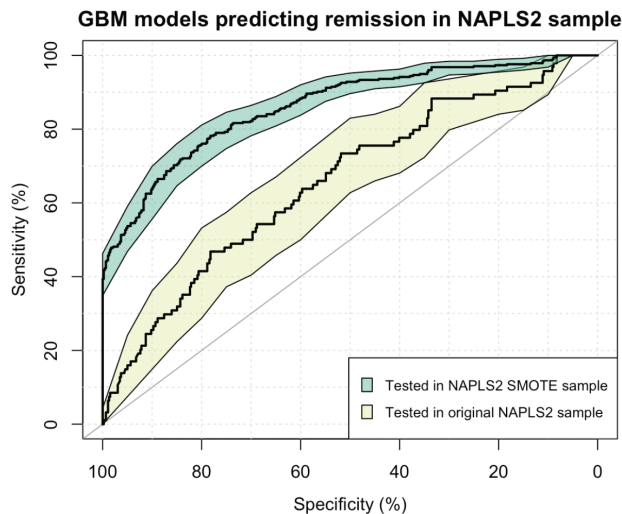

Fig. 2. Receiver operating characteristic (ROC) curves for performance classification models predicting remission in the NAPLS2 testing samples.

sample (NAPSL2) which provides a more robust and "real-world" estimate of the generalizability and utility of the prediction model. When tested in the original NAPLS2 sample, the model achieved an AUC of 0.66 (95% CI = 0.60–0.72) and with a median remission likelihood cutoff of 0.4, achieved a sensitivity of 0.64 and specificity of 0.59. When tested in the SMOTE-sampled NAPLS2 dataset, the model achieved an AUC of 0.86 (95% CI = 0.84–0.89) and with the same remission likelihood cutoff of 0.4 achieved a sensitivity of 0.84 and specificity of 0.59 (table 3). The ROC curves of the models predicting remission can be found in figure 2. As a final test of robustness, when derived in the NAPLS2 training sample and tested in the original NAPLS3 sample and a SMOTE-sampled NAPLS3 sample (ie, reversing the roles of the discovery and validation datasets), the models achieved similar performances to the models trained in the NAPLS3 sample.

Discussion

This study is the first of its kind to use a discovery-based, data-driven approach to develop generalizable models

predicting remission as a distinct clinical outcome among CHR-P individuals. The performance of the multivariable models was generally comparable to the performance of existing models predicting conversion to psychosis.[2–4] A major strength of this study is the inclusion of a completely independent validation sample with which to test the generalizability of the models. The model performed well when internally tested on the NAPLS3 discovery sample, and also performed adequately well in the validation sample when accounting for expected statistical shrinkage across samples and the potential for overfitting in the discovery sample by excluding the elastic net feature selection from the cross-validation. The importance of accounting for the class imbalance during the classification model development stage is evident in the fact that when not doing so, the NAPLS3-derived model is only able to predict nonremission status and does not correctly detect any remitters in the NAPLS2 sample (AUC [95% CI] = 0.66 (0.60–0.72); sensitivity = 0.01, specificity = 1.0).

Within the predictor variables used in the final model, unusual thought content/delusional ideas was the only item that was not significantly different between remitters and nonremitters at baseline, though remitters scored lower than nonremitters on this item as well. Further, this item was one of the top 3 variables according to the variable importance plot output from the GBM model (eFigure 1), indicating a greater reduction in accuracy when this variable is removed from the final model. This suggests that the change in this particular symptom may be an important feature of remitters. Greater severity of unusual thought content and suspiciousness have been shown to be strongly associated with conversion to psychosis[1,2,45] and may represent the primary positive symptoms driving individuals to seek specialized clinical services. Thus, it would be expected that 1 or both of these items would be elevated across CHR-P individuals at baseline despite future remission status. Nevertheless, this item seems to play an important role in predicting remission as well. In fact, while there is a significant change in positive symptoms between baseline and the time of remission for remitters, unusual thought content shows the greatest decline in severity as compared to the other positive symptoms (eTable 5 and eFigure 2).

The clinical applications of a model predicting an individual's likelihood for remission could include informing treatment decisions based on an in-person clinical assessment. Understanding the likelihood of an individual's outcome (eg, conversion to psychosis, remission, or neither) could help clinicians decide how to assign interventions such that those at highest risk receive the most intensive interventions while those at lowest risk or who are more likely to remit with usual and customary treatment receive less intensive interventions. Further, understanding protective factors characterizing remitters could inform how to tailor components of existing and novel psychosocial treatments.

Limitations and Future Directions

A limitation of the models used in the present study is the ability to draw descriptive conclusions about remitters. We included mean comparisons between remitters and nonremitters, and between remitters and nonremitters with converters excluded (see Supplement), for all predictor variables. These patterns are consistent with the interpretation that future remitters are less symptomatic at baseline than nonremitters. Future work to validate the construct of remission should include the longitudinal analysis of symptom clusters in addition to the evaluation of functional and structural neuroimaging patterns. Whereas converters tend to show deterioration over time in measures of brain structure and functional analysis,[46,47] it is yet to be determined what type of pattern might be present in remitters, eg, whether there is an improvement in functional connectivity or lack of structural deterioration, and whether these patterns are distinct or look similar to health controls.

The definition of remission used in this study may not adequately capture the long-term trajectory of CHR-P individuals whose symptoms substantially improve (eg, cognition may remain impaired relative to healthy controls while prodromal symptoms or functioning remit[13,16]). "Remission" has not yet been clearly defined or consistently used in the CHR-P field. While criteria have been set based on the clinical measures, it is still unknown what remission looks like in practice. Our ability to detect whether individuals stay remitted for longer than 1 year is not possible in the present study. Observational studies with longer follow-up periods would be required to address this. The fact that a specific subset of prodromal symptoms is predictive of remission indicates that there may be a more precise clinical picture associated with remission, ie, detectable at a baseline visit that warrants further study.

Another limitation of this analysis is that the feature selection process was performed outside of the cross-validation procedure, which may potentially result in overfitting our model in the discovery sample. Given that the model was tested in a completely independent test sample (NAPLS2) and still performed well, the potential for a biased model is lower. However, future studies should be sure to implement all model training procedures within the cross-validation to prevent data leakage, especially when a robust external validation dataset is not available.

In addition, it is unknown how effective current interventions may be in ameliorating attenuated psychotic symptoms. In the NAPLS3 sample, approximately half of the full sample received psychosocial interventions (eg, cognitive-behavioral therapy, supportive therapy, case management, etc.) from a community provider either prior to, or during, the course of the study and a small proportion received antipsychotics. Baseline antipsychotic use, the types of psychosocial interventions, and average number of sessions received did not differ between remitters and nonremitters (eTable 6). Due to the naturalistic/observational nature of the NAPLS3 study and absence of randomized clinical trial structure, it is not possible to draw inferences regarding the potential effects of psychosocial or pharmacological treatment on remission outcomes. Randomized control trials with remission as the primary outcome would be needed to better understand which interventions are most associated with sustained symptomatic remission.

Future analyses may also incorporate biological data such as neuroimaging data, EEG data, and cortisol. In the present study, we chose not to include biomarkers, as the subset of participants who completed these assessments is substantially smaller, and variance in biomarker collection methods between samples limits model generalizability during external validation. Further, the focus of the present study was to determine whether it was possible to build individualized models that accurately and reliably predict sustained remission; future studies focusing more specifically on the subsamples of participants who completed biomarker assessments will seek to determine whether models involving these biomarkers outperform clinical models. Clinical measures may be preferred as the sampling of biomarkers is often more time-consuming, costly, and sometimes invasive. Nevertheless, such measures may improve prediction of remission outcomes and shed light on mechanisms associated with remission. Many studies have shown neuroanatomical structural differences between converters and nonconverters[48]; however, to date, no studies have examined potential structural differences between remitters and nonremitters. Functional differences have been shown in the NAPLS2 data for the Target P300 event-related potential component elicited during an auditory oddball task wherein remitters had P300 amplitudes similar to healthy controls at baseline, in addition to significantly higher P300 amplitudes than both nonremitting CHR-P nonconverters and CHR-P converters when considered separately.[49] In addition, baseline cortisol levels for remitters may resemble the levels of healthy controls

more than any other diagnostic outcome from the CHR-P state.[50] Other studies have implicated intact auditory novelty P300[51] and mismatch negativity[52] as characteristic of CHR cases who later remit.[53]

The present study is a starting point in a nascent area of research within the CHR-P framework. Using remission as an outcome of interest will hopefully shift the field's focus toward recovery-oriented outcomes, which could inform more effective interventions through a better understanding of protective factors and mechanisms.

Supplementary Material

Supplementary material is available at *Schizophrenia Bulletin* online.

Funding

Acknowledgments

References

1. Fusar-Poli P, Borgwardt S, Bechdolf A, et al. The psychosis high-risk state: a comprehensive state-of-the-art review. *JAMA Psychiatry.* 2013;70(1):107–120.

2. Cannon TD, Yu C, Addington J, et al. An individualized risk calculator for research in prodromal psychosis. *Am J Psychiatry.* 2016;173(10):980–988.

3. Zhang T, Xu L, Tang Y, et al. Prediction of psychosis in prodrome: development and validation of a simple, personalized risk calculator. *Psychol Med.* 2018;49(12):1–9.

4. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al. Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: a multimodal, multisite machine learning analysis. *JAMA Psychiatry.* 2018;75(11):1156–1172.

5. Mechelli A, Lin A, Wood S, et al. Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. *Schizophr Res.* 2017;184:32–38.

6. Rosen M, Betz LT, Schultze-Lutter F, et al. Towards clinical application of prediction models for transition to psychosis: a systematic review and external validation study in the PRONIA sample. *Neurosci Biobehav Rev.* 2021;125:478–492.

7. Salazar de Pablo G, Studerus E, Vaquerizo-Serrano J, et al. Implementing precision psychiatry: a systematic review of individualized prediction models for clinical practice. *Schizophr Bull.* 2021;47(2):284–297.

8. Polari A, Lavoie S, Yuen HP, et al. Clinical trajectories in the ultra-high risk for psychosis population. *Schizophr Res.* 2018;197:550–556.

9. Polari A, Yuen HP, Amminger P, et al. Prediction of clinical outcomes beyond psychosis in the ultra-high risk for psychosis population. *Early Interv Psychiatry.* 2021;15(3):642–651.

10. Salazar de Pablo G, Besana F, Arienti V, et al. Longitudinal outcome of attenuated positive symptoms, negative symptoms, functioning and remission in people at clinical high risk for psychosis: a meta-analysis. *EClinicalMedicine.* 2021;36:1–10.

11. Catalan A, Salazar de Pablo G, Vaquerizo Serrano J, et al. Annual Research Review: Prevention of psychosis in adolescents—systematic review and meta-analysis of advances in detection, prognosis and intervention. *J Child Psychol Psychiatry.* 2021;62(5):657–673.

12. Hartmann JA, Schmidt SJ, McGorry PD, et al. Trajectories of symptom severity and functioning over a three-year period in a psychosis high-risk sample: a secondary analysis of the Neurapro trial. *Behav Res Ther.* 2020;124:1–9.

13. Addington J, Stowkowy J, Liu L, et al. Clinical and functional characteristics of youth at clinical high-risk for psychosis who do not transition to psychosis. *Psychol Med.* 2019;49(10):1670–1677.

14. Carrión RE, McLaughlin D, Goldberg TE, et al. Prediction of functional outcome in individuals at clinical high risk for psychosis. *JAMA Psychiatry.* 2013;70(11):1133–1142.

15. Allswede DM, Addington J, Bearden CE, et al. Characterizing covariant trajectories of individuals at clinical high risk for psychosis across symptomatic and functional domains. *Am J Psychiatry.* 2020;177(2):164–171.

16. Lee TY, Shin YS, Shin NY, et al. Neurocognitive function as a possible marker for remission from clinical high risk for psychosis. *Schizophr Res.* 2014;153(1–3):48–53.

17. Glenthøj LB, Kristensen TD, Wenneberg C, Hjorthøj C, Nordentoft M. Predictors of remission from the ultra-high risk state for psychosis. *Early Interv Psychiatry.* 2020;15(1):104–112.

18. Ziermans TB, Schothorst PF, Sprong M, van Engeland H. Transition and remission in adolescents at ultra-high risk for psychosis. *Schizophr Res.* 2011;126(1–3):58–64.

19. Ferrarelli F, Mathalon D. The prodromal phase: time to broaden the scope beyond transition to psychosis? *Schizophr Res.* 2020;216:5–6.

20. Hauke DJ, Schmidt A, Studerus E, et al. Multimodal prognosis of negative symptom severity in individuals at increased risk of developing psychosis. *Transl Psychiatry.* 2021;11(1):1–11.

21. Chekroud AM, Zotti RJ, Shehzad Z, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry.* 2016;3(3):243–250.

22. Koutsouleris N, Kahn RS, Chekroud AM, et al. Multisite prediction of 4-week and 52-week treatment outcomes in patients with first-episode psychosis: a machine learning approach. *Lancet Psychiatry.* 2016;3(10):935–946.

23. Kambeitz J, Goerigk S, Gattaz W, et al. Clinical patterns differentially predict response to transcranial direct current stimulation (tDCS) and escitalopram in major depression: a machine learning analysis of the ELECT-TDCS study. *J Affect Disord.* 2020;265:460–467.

24. Addington J, Liu L, Brummitt K, et al. North American Prodrome Longitudinal Study (NAPLS 3): methods and baseline description. *Schizophr Res.* 2020. doi:10.1016/j.schres.2020.04.010

25. McGlashan TH, Walsh B, Woods S. *The Psychosis-Risk Syndrome: Handbook for Diagnosis and Follow-up*. New York, NY: Oxford University Press; 2010.

26. McGlashan TH, Walsh BC, Woods SW. *Structured Interview for Psychosis-Risk Syndromes*. New Haven, CT: Yale School of Medicine; 2001.

27. Addington J, Cadenhead KS, Cannon TD, et al. North American Prodrome Longitudinal Study: a collaborative multisite approach to prodromal schizophrenia research. *Schizophr Bull.* 2007;33(3):665–672.

28. Castillo R, Carlat D, Millon T, et al. *Diagnostic and Statistical Manual of Mental Disorders*. Washington, DC: American Psychiatric Association Press; 2007.

29. Addington J, Cadenhead KS, Cornblatt BA, et al. North American Prodrome Longitudinal Study (NAPLS 2): overview and recruitment. *Schizophr Res.* 2012;142(1–3):77–82.

30. Hawkins KA, McGlashan TH, Quinlan D, et al. Factorial structure of the Scale of Prodromal Symptoms. *Schizophr Res.* 2004;68(2–3):339–347.

31. Addington D, Addington J, Maticka-Tyndale E, Joyce J. Reliability and validity of a depression rating scale for schizophrenics. *Schizophr Res.* 1992;6(3):201–208.

32. Keefe RS, Harvey PD, Goldberg TE, et al. Norms and standardization of the Brief Assessment of Cognition in Schizophrenia (BACS). *Schizophr Res.* 2008;102(1–3):108–115.

33. Benedict RH, Schretlen D, Groninger L, Brandt J. Hopkins Verbal Learning Test—Revised: normative data and analysis of inter-form and test-retest reliability. *Clin Neuropsychol.* 1998;12(1):43–55.

34. Hall RC. Global assessment of functioning. A modified scale. *Psychosomatics.* 1995;36(3):267–275.

35. Cornblatt BA, Auther AM, Niendam T, et al. Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophr Bull.* 2007;33(3):688–702.

36. Miller TJ, McGlashan TH, Rosen JL, et al. Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: predictive validity, interrater reliability, and training to reliability. *Schizophr Bull.* 2003;29(4):703–715.

37. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2021. https://www.R-project.org/. Accessed January 1, 2021.

38. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 2010;33(1):1.

39. Siriseriwan W. *Smotefamily: A Collection of Oversampling Techniques for Class Imbalance Problem Based on SMOTE*. 2018. http://cranr-projectorg/package=smotefamily. Accessed January 1, 2021.

40. Kuhn M. *Caret: Classification and Regression Training*. 2015. https://cran.r-project.org/web/packages/caret/. Accessed January 1, 2021.

41. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc B (Stat Methodol).* 2005;67(2):301–320.

42. Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal.* 2002;38(4):367–378.

43. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002;16:321–357.

44. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics.* 1988;44(3):837–845.

45. Cannon TD, Cadenhead K, Cornblatt B, et al. Prediction of psychosis in youth at high clinical risk: a multisite longitudinal study in North America. *Arch Gen Psychiatry.* 2008;65(1):28–37.

46. Chung Y, Allswede D, Addington J, et al. Cortical abnormalities in youth at clinical high-risk for psychosis: findings from the NAPLS2 cohort. *Neuroimage Clin.* 2019;23:101862.

47. Cao H, Chén OY, Chung Y, et al. Cerebello-thalamo-cortical hyperconnectivity as a state-independent functional neural signature for psychosis prediction and characterization. *Nat Commun.* 2018;9(1):1–9.

48. Worthington MA, Cao H, Cannon TD. Discovery and validation of prediction algorithms for psychosis in youths at clinical high risk. *Biol Psychiatry Cogn Neurosci Neuroimaging.* 2020;5(8):738–747.

49. Hamilton HK, Roach BJ, Bachman PM, et al. Association between P300 responses to auditory oddball stimuli and clinical outcomes in the psychosis risk syndrome. *JAMA Psychiatry.* 2019;76(11):1187–1197.

50. Walker EF, Trotman HD, Pearce BD, et al. Cortisol levels and risk for psychosis: initial findings from the North American prodrome longitudinal study. *Biol Psychiatry.* 2013;74(6):410–417.

51. Tang Y, Wang J, Zhang T, et al. P300 as an index of transition to psychosis and of remission: data from a clinical high risk for psychosis study and review of literature. *Schizophr Res.* 2020;226:74–83.

52. Kim M, Lee TH, Yoon YB, Lee TY, Kwon JS. Predicting remission in subjects at clinical high risk for psychosis using mismatch negativity. *Schizophr Bull.* 2018;44(3):575–583.

53. Hamilton HK, Boos AK, Mathalon DH. Electroencephalography and event-related potential biomarkers in individuals at clinical high risk for psychosis. *Biol Psychiatry.* 2020;88(4):294–303.