Prediction of Early Symptom Remission in Two Independent Samples of First-Episode Psychosis Patients Using Machine Learning

Rigas F. Soldatos*[,1,2,3,9,©], Micah Cearns[2,4,9], Mette Ø. Nielsen[5,6], Costas Kollias[1], Lida-Alkisti Xenaki[1], Pentagiotissa Stefanatou[1], Irene Ralli[1], Stefanos Dimitrakopoulos[1], Alex Hatzimanolis[1], Ioannis Kosteletos[1], Ilias I. Vlachos[1,©], Mirjana Selakovic[1], Stefania Foteli[1], Nikolaos Nianiakas[1,©], Leonidas Mantonakis[1], Theoni F. Triantafyllou[1], Aggeliki Ntigridaki[1], Vanessa Ermiliou[1], Marina Voulgaraki[1], Evaggelia Psarra[1], Mikkel E. Sørensen[5], Kirsten B. Bojesen[5], Karen Tangmose[5,6], Anne M. Sigvard[5,6], Karen S. Ambrosen[5], Toni Meritt[2], Warda Syeda[2], Birte Y. Glenthøj[5,6], Nikolaos Koutsouleris[3,7], Christos Pantelis[2,3,10,©], Bjørn H. Ebdrup[2,5,6,10], and Nikos Stefanis[1,3,8,10]

[1]First Department of Psychiatry, National and Kapodistrian University of Athens Medical School, Eginition Hospital, Athens, Greece; [2]Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne and Melbourne Health, Carlton South, Victoria, Australia; [3]World Federation of Societies of Biological Psychiatry, First Episode Psychosis Task Force, Barsbüttel, Germany; [4]Discipline of Psychiatry, School of Medicine, University of Adelaide, Australia; [5]Center for Neuropsychiatric Schizophrenia Research (CNSR) & Center for Clinical Intervention and Neuropsychiatric Schizophrenia Research (CINS), Mental Health Centre Glostrup, University of Copenhagen, Glostrup, Denmark; [6]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; [7]Department of Psychiatry and Psychotherapy, Section for Neurodiagnostic Applications, Ludwig-Maximilian University, Munich, Germany; [8]University Mental Health, Neurosciences and Precision Medicine Research Institute, 2 Soranou Efesiou, 11527 Athens, Greece; [9]These authors contributed equally to this work and are equal first authors; [10]These authors contributed equally to this work and are equal senior authors.

*To whom correspondence should be addressed; First Department of Psychiatry, Eginition Hospital, 72-74 Vasilissis Sofias Avenue, Athens 11527, Greece, tel: +302107289400, e-mail: rsoldatos@gmail.com

Background: Validated clinical prediction models of short-term remission in psychosis are lacking. Our aim was to develop a clinical prediction model aimed at predicting 4−6-week remission following a first episode of psychosis. *Method:* Baseline clinical data from the Athens First Episode Research Study was used to develop a Support Vector Machine prediction model of 4-week symptom remission in first-episode psychosis patients using repeated nested cross-validation. This model was further tested to predict 6-week remission in a sample of two independent, consecutive Danish first-episode cohorts. *Results:* Of the 179 participants in Athens, 120 were male with an average age of 25.8 years and average duration of untreated psychosis of 32.8 weeks. 62.9% were antipsychotic-naïve. Fifty-seven percent attained remission after 4 weeks. In the Danish cohort, 31% attained remission. Eleven clinical scale items were selected in the Athens 4-week remission cohort. These included the Duration of Untreated Psychosis, Personal and Social Performance Scale, Global Assessment of Functioning and eight items from the Positive and Negative Syndrome Scale. This model significantly predicted 4-week remission status (area under the receiver operator characteristic curve (ROC-AUC) = 71.45, $P < .0001$). It also predicted 6-week remission status in the Danish cohort (ROC-AUC = 67.74, $P < .0001$), demonstrating reliability. *Conclusions:* Using items from common and validated clinical scales, our model significantly predicted early remission in patients with first-episode psychosis. Although replicated in an independent cohort, forward testing between machine learning models and clinicians' assessment should be undertaken to evaluate the possible utility as a routine clinical tool.

Key words: first-episode/psychosis/psychosis/schizophrenia/remission/prediction/psychopathology/machine learning

Introduction

Recovery, remission and relapse prevention in psychotic disorders have become important areas of interest in psychiatric research.[1] Despite observed symptom reduction, pharmacological treatments appear to have little impact on these broader outcomes, leading to high costs and increased mortality for patients.[2–4] Early intervention may lead to better patient outcomes and a greater chance of
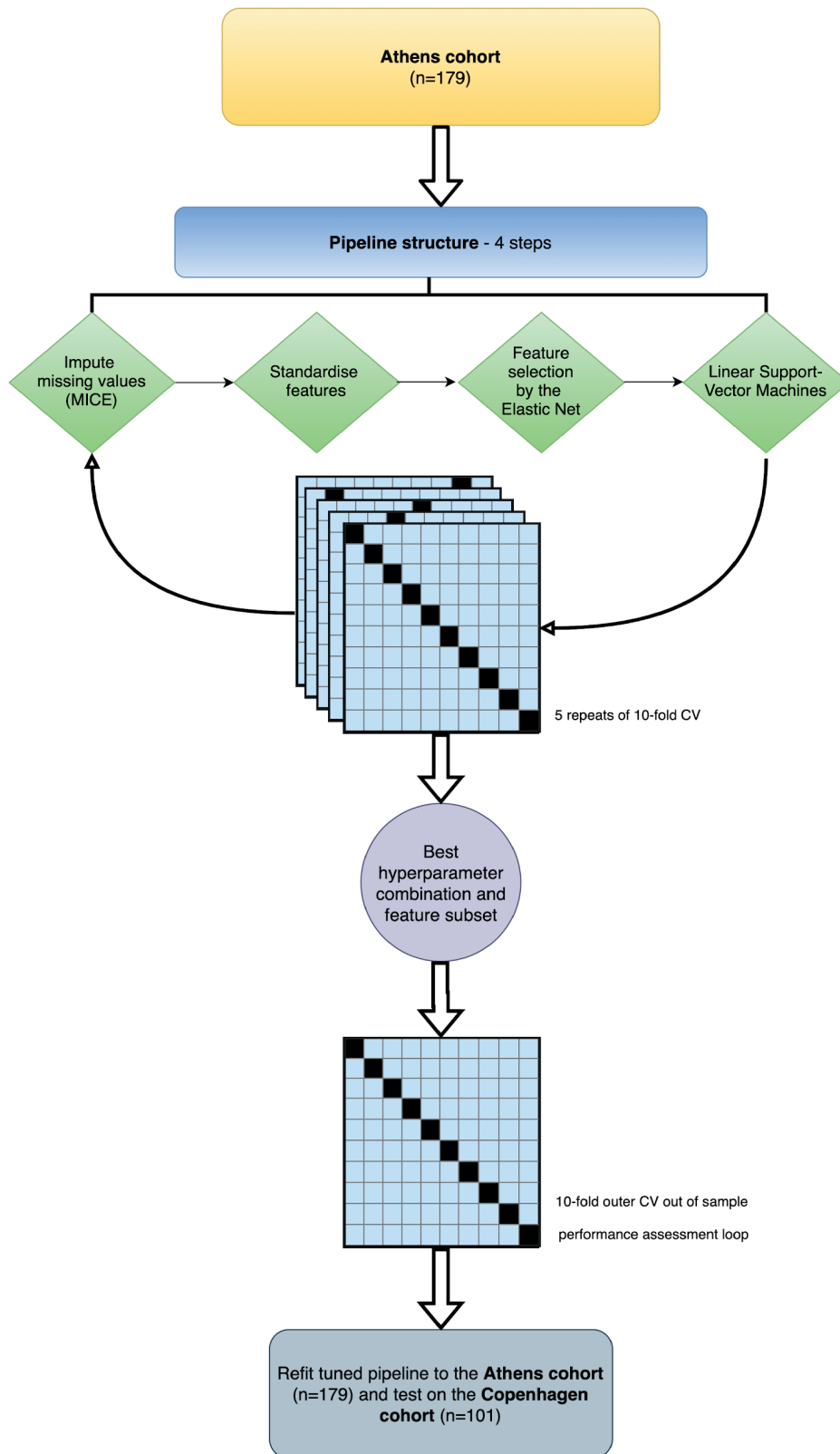
Fig. 1. First episode psychosis machine learning model pipeline. Athens first-episode psychosis patients were passed into the machine learning pipeline and used for model training and selection. The best hyperparameters and predictors were chosen using five repeats of 10-fold cross-validation (CV). This optimized pipeline was then refitted to the Athens first-episode psychosis cohort and used to make predictions on the Copenhagen cohort.

long-term recovery.[5–8] Currently, a longer duration of untreated psychosis (DUP), male gender, an earlier age of illness onset, an insidious onset, a longer duration of the prodromal phase, a higher degree of negative symptoms at baseline, comorbid substance use disorders, treatment delay, shorter time in treatment, and lack of insight in First Episode Psychosis (FEP) patients have all been recognized as risk factors for relapse in schizophrenia and predictors of poor outcome at 3 and 10-year follow-up intervals.[9–18] Some constitute modifiable risk factors and efforts to address them have been developed through Early Intervention (EI) services.[1,19–23] However, EI services are faced with the challenge of providing individual detection for patients who transition to first-onset psychosis, which is hampered by the small fraction of patients in contact with them before making the transition to a psychotic prevention.[7]

Evidence suggests that multivariate prediction models may be useful and even outperform clinician prognostication.[24] Recently, Koutsouleris et al[24] developed prediction models of functional outcomes in young individuals at risk of developing psychosis and depression. These models outperformed clinicians in predicting functional outcomes, highlighting their potential in clinical practice and early intervention. The breadth of application extends beyond early intervention and into optimizing allocation of health resources. To this end, machine learning (ML) based decision support systems that predict the likelihood of remission upon first presentation may assist in the allocation of these resources, leading to better patient outcomes at scale.

As previous studies focusing on remission at the one-year time point and improvement in functioning at both the 1 month and 1-year timepoints have yielded promising results,[25–27] our intention is to focus on the initial month of treatment in order to develop a prediction model of 4-week FEP remission. The first years following a FEP can provide a unique insight into a period characterized by a high degree of illness variability.[29,30] In addition, at this time clinicians commonly aim for patients to be reintegrated back into their social framework. Therefore, our aim is to create a clinically useful and applicable tool that can be used to identify patients' remission status in one month following a FEP. To achieve this, we have used clinical scales that are used in the first assessment of FEP, to predict remission at 4 weeks following initial patient presentation. Scales used in model training included the Positive and Negative Syndrome Scale (PANSS), the Hamilton Rating Scale for Depression (HAM-D 17 item), and the Young Mania Rating Scale.[30–32] Applying the PANSS assumes specific training and is not part of routine clinical assessment. However, due to its granularity, it can help identify clinical aspects that relate to remission in FEP. In order to create a robust outcome, we used the validated definition of remission as defined by Andreasen et al.[33] Finally, we trained, validated, and tested a linear Support Vector Machine classifier (SVM) on the Greek "Athens First Episode Psychosis Research Study" dataset using repeated nested cross-validation and then further validated our model in an independent first-episode dataset from Copenhagen, Denmark.

Methods

Athens Cohort

Our prediction model was developed, validated, and tested using patients from the ongoing Athens First Episode Psychosis Research Study. Methods detailing participants, recruitment, measures and overall construction of the Athens cohort have been described.[34] The study was approved by the Research Ethics committee of the Eginition hospital (protocol 644Υ46Ψ8Ν2-ΓΚΣ). Patients gave written informed consent. At time of analysis, 181 FEP patients were recruited over a 40-month period (study commencement: March 2015). Recruitment is ongoing and will end when 220 cases have been collected. See table 1A for final sample characteristics, supplementary methods 1.1 and supplementary figure S1.

Copenhagen Cohort

To further assess the generalizability of our model, we tested our model on the merged patient sample derived from two independent, consecutive, longitudinal, multimodal first-episode cohorts, PECANS I (Pan European Collaboration on Antipsychotic Naive Schizophrenia) (initiated December 2008) and PECANS II (initiated January 2014), for initial descriptions of the cohorts, see.[35,36] See table 1B for final sample characteristics and supplementary methods 1.2 for more information.
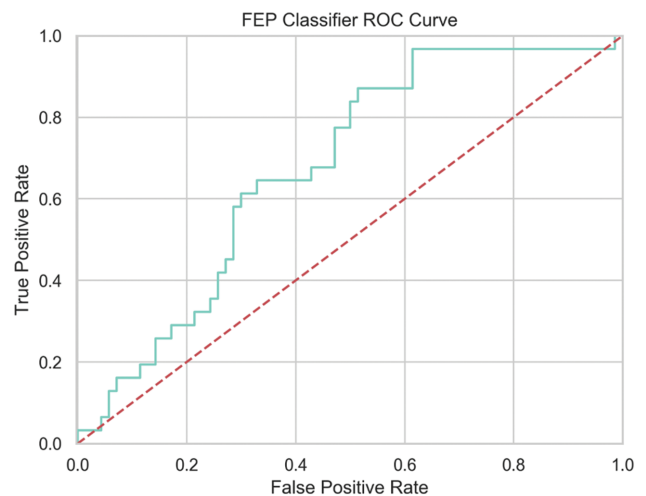

Fig. 2. Area under the receiver operator characteristic curve (ROC-AUC) for the First Episode Psychosis (FEP) classifier. As we change the decision threshold for the models underlying probabilities, trade-offs between true and false positives can be observed.

Table 1. Baseline Patient Characteristics for the Athens (Section A) and Copenhagen (Section B) Cohorts Stratified by Remission

A) Athens Baseline Characteristic	Descriptive Statistic	Remission N = 102	Nonremission N = 77	Remission vs Nonremission P-value*
Age (*n* = 176)	Mean (SD)	25.9 (7.5)	25.7 (7.7)	.92
Gender, male (*n* = 179)	*N* (%)	64 (63%)	56 (73%)	.60
Education (years) (*n* = 176)	Mean (SD)	13.6 (2.6)	13.6 (2.3)	.92
GAF baseline (*n* = 151)	Mean (SD)	45.1 (16.6)	37.8 (13.9)	**.01**
PSP baseline total score (*n* = 137)	Mean (SD)	40.5 35.7 (16.5)	35.7 (15.4)	.12
Cohabitants (*n* = 175)	Mean (SD)	2.5 (2.22)	2.6 (2)	.76
HAM total (*n* = 140)	Mean (SD)	15.2 (8.1)	19.1 (11.1)	.05
YMRS total (*n* = 139)	Mean (SD)	13.1 (11)	16.4 (11.3)	.13
DUP (*n* = 169)	Mean (SD)	19.1 (32.9)	52.6 (81.1)	**<.01**
DUI (*n* = 150)	Mean (SD)	37.8 (50.4)	82.5 (134.9)	**.03**
PANSS total (*n* = 179)	Mean (SD)	93.9 (21.2)	106.6 (25.2)	**<.01**
PANSS positive total (*n* = 179)	Mean (SD)	27.8 (7)	29.2 (6.6)	.21
PANSS negative total (*n* = 179)	Mean (SD)	18.1 (8.3)	25.1 (9.9)	**<.01**
PANSS general total (*n* = 179)	Mean (SD)	48.1 (12.5)	52.3 (14.7)	.08
Antipsychotic use prior to entry, yes (*n* = 178)	*N* (%)	36 (36%)	30 (39%)	.70
Currently living alone, yes (*n* = 161)	*N* (%)	16 (17%)	15 (22%)	.70
Currently unemployed, yes (*n* = 179)	*N* (%)	25 (25%)	30 (39%)	.25
Lifetime cannabis use, yes (*n* = 177)	*N* (%)	51 (50%)	42 (55%)	.70
Cannabis use in the past month, yes (*n* = 168)	*N* (%)	25 (27%)	17 (23%)	.70
Affective/nonaffective (*n* = 165)	*N* (%)	69 (73%)	64 (91%)	**.03**
Stimulant use last 12 months, yes (*n* = 178)	*N* (%)	11 (11%)	9 (12%)	.87
Opioid use last 12-months, yes (*n* = 168)	*N* (%)	0 (0%)	1 (1%)	.60
Hallucinogen use last 12-months, yes (*n* = 170)	*N* (%)	8 (8%)	3 (4%)	.60
Tobacco use last 12-months, yes (*n* = 179)	*N* (%)	57 (56%)	40 (52%)	.70
Alcohol use >14 units per week, yes (*n* = 179)	*N* (%)	2 (2%)	3 (4%)	.70
CGI Severity of illness: 3/4/5/6/7 (*n* = 172)	Median (IQR)	6 (5–7)	6 (5–7)	.60

B) Copenhagen Baseline Characteristic	Descriptive Statistic	Remission N = 31	Nonremission N =70	Remission vs Nonremission P-value*
Age (*n* = 101)	Mean (SD)	22.8 (3.7)	24.3 (5.9)	.19
Gender, male (*n* = 101)	*N* (%)	16 (52%)	37 (53%)	1
Education (years) (*n* = 58)	Mean (SD)	12.9 (2.7)	11.7 (2.1)	.18
GAF baseline (*n* = 98)	Mean (SD)	41 (8)	36 (6.9)	**.02**
PSP baseline total score (*n* = 52)	Mean (SD)	50.4 (11.8)	47.7 (13.8)	.55
DUP (*n* = 51)	Mean (SD)	102.7 (165.4)	123.84 (248)	.72
PANSS total (*n* = 101)	Mean (SD)	70 (15.9)	78.8 (14)	**.02**
PANSS negative total (*n* = 101)	Mean (SD)	17 (5.6)	20.4 (6.3)	**.02**
Lifetime cannabis ICD10 Dx, yes (*n* = 101)	*N* (%)	1 (3%)	0 (0%)	1
Lifetime opioids ICD10 Dx, yes (*n* = 101)	*N* (%)	0 (0%)	3 (4%)	1
Lifetime tobacco ICD10 Dx, yes (*n* = 101)	*N* (%)	4 (13%)	0 (0%)	1
Lifetime stimulants ICD10 Dx, yes (*n* = 101)	*N* (%)	0 (0%)	9 (13%)	1
Lifetime hallucinogens ICD10 Dx, yes (*n* = 101)	*N* (%)	0 (0%)	0 (0%)	1
Lifetime other recreational drugs ICD10 Dx, yes (*n* = 101)	*N* (%)	0 (0%)	0 (0%)	1
Lifetime benzodiazepines ICD10 Dx, yes (*n* = 101)	*N* (%)	0 (0%)	0 (0%)	1

*Calculated using Welches *t*-test for continuous variables and chi-square tests for categorical variables. All *P*-values were FDR corrected using the Benjamini and Hochberg method. Section A) For substance use, any use in the last 12 months regardless of frequency. Remission based on 4-weeks assessment. Section B) Lifetime substance use was assessed on a 5-point Likert scale (0 = never tried, 1 = tried a few times, 2 = use regularly, 3 = harmful use, 4 = dependency), with levels 3 and 4 corresponding to an ICD10 diagnoses of harmful use or dependence. Therefore, we scored lifetime use as ≥3. Remission based on 6 weeks assessment.
GAF: Global Assessment of Functioning, PSP: Personal and Social Performance Scale, PANSS: Positive And Negative Syndrome Scale, HAM: Hamilton Depression Rating Scale, YMRS: Young Mania Rating Scale, DUP: Duration of Untreated Psychosis, DUI: Duration of Untreated Illness.

Exclusion criteria differences between the Athens and Copenhagen cohorts can be found in table 4.

Sample Predictors and Outcome

To train our model we used a subset of sociodemographic and psychopathological predictors from the Athens cohort. The initial set of measures included clinical scales, environmental factors, demographic characteristics, cognitive tests and biological tests including routine blood tests, biochemical and immunological assays. The subset used to develop our model contains clinical and demographic variables that have been identified in the past to be associated with clinical remission.[9–18] The final variables included all items and total scores from the PANSS, the 17-item HAM-D, the Young Mania Rating Scale (YMRS), the Clinical Global Impression (CGI) scale, the Global Assessment of Functioning (GAF) scale and the Personal and Social Performance scale (PSP). All scales were completed upon first contact and at the 4-week/6-week follow-up assessments in both samples. In total, 43 clinical and demographic predictors were included in the analysis. See supplementary table S1 for all included predictors. Remission was defined according to the Andreasen PANSS remission criteria which has shown good validity across different studies.[33,37–41] However, as our outcomes were assessed at 4- and 6-week time points, we waived the 6-month criterion.

Training, Model Selection, and Testing – Athens Cohort

For model selection, training, and initial testing we conducted repeated nested cross-validation on the Athens Cohort. To derive the most predictive model of remission, we used an ML pipeline composed of the following transformations (see figure 1). First, we imputed missing values using multivariate imputation by chained equations (MICE).[42] Second, we standardized all predictors to have a mean of 0 and standard deviation of 1. Third, we fitted an Elastic Net model, a form of penalized logistic regression that removes highly correlated predictors from the model whilst retaining the most predictive groups of correlated predictors for model fitting. We fit two pipelines in this process, one with a set of hyperparameters that covered the full range of the mixing parameter (l1 ratio) and the regularization strength parameter (alpha), and another set of hyperparameters where the l1 ratio was closer to Ridge (eg, biased towards the inclusion of more predictors) and implemented less regularization (lower values of alpha). This was done to avoid selecting a model with too few predictors that may not generalize to an independent validation cohort as has been shown in previous studies.[43] Finally, we fit a linear Support Vector Machine (SVM) to the predictors selected by the Elastic Net to classify patients who were in remission at their 4-week follow-up compared to those who were not. All steps were completed within a pipeline to avoid the leakage of statistics across transformations. Hyperparameters for the Elastic Net and SVM were tuned using five repeats of 10-fold cross-validation in an inner cross-validation loop to maximize the area under the receiver operator characteristic curve (AUC). Details of the two sets of tuned hyperparameters and software used can be found in supplementary Appendix 1.5. Out of sample performance was then assessed in a 10-fold outer cross-validation loop.

As clinicians' predictions for patients' remission status were not taken at baseline assessment, we used the baseline clinical global impression severity (CGI-S) scale to assess whether clinicians rated patients as significantly different in their illness severity upon first presentation. See supplementary Methods 1.3 for more information.

Further Independent Testing – Copenhagen Cohort

To further test our model independently of the Athens cohort, the best combination of hyperparameters from the inner training loop were refitted without modification to the entire Athens cohort and used to make predictions on the Copenhagen cohort. As the remission rates were different between these datasets (Athens = 57%, Copenhagen = 30.7%) we used Bayes' adjusted positive predictive value (PPV) and negative predictive value (NPV) to correct for these differences in prevalence.[44] This method standardizes the prevalence rates between the two cohorts allowing for the proper interpretation of model performance (see supplementary Methods 1.4). We standardized the Copenhagen prevalence to the Athens cohort prevalence as this value closely represents the event rate seen on similar cohorts in the literature.[14,15,45–49] In addition, we re-calculated Bayes' adjusted PPV and NPV for all prevalence rates of FEP remission observed in the literature (30−60%)[15,50] allowing for estimates of how our model might perform across this prevalence distribution on different cohorts (see supplementary table S2). Finally, we derived the null distribution and statistical significance of this classifier using a permutation test of 10 000 permutations. See supplementary Methods 1.5 for the tuned and selected hyperparameters, for more information on this pipeline configuration see.[51]

Secondary Analysis − Diagnostic Status and Sample Size Effects

As the Athens sample was composed of patients with affective, nonaffective, and drug-induced psychosis components in their diagnoses, whereas the Copenhagen cohort was only composed of patients with a nonaffective diagnosis, we conducted further analysis to test the generalizability of our model across these three diagnostic categories. See supplementary Methods 1.6. In addition, we conducted a subsampling analysis to quantify the relationship between balanced accuracy and sample size.[52]

Table 2. Section A: Patient Characteristics Stratified by Cohort (Athens vs Copenhagen) and Section B: Athens Affective/Nonaffective Status

A) Athens vs Copenhagen	Descriptive Statistic	Athens N = 179	Copenhagen N =101	Athens vs Copenhagen P-value*
Age (*n* = 277)	Mean (SD)	25.8 (7.5)	23.8 (5.4)	**.01**
Gender (*n* = 280)	*N* (%)	120 (67%)	53 (52%	**.02**
Education (years) (*n* = 234)	Mean (SD)	13.6 (2.5)	12.1 (2.3)	**<.01**
GAF baseline (*n* = 249)	Mean (SD)	41.9 (15.8)	37.5 (7.5)	**.01**
PSP baseline total score (*n* = 189)	Mean (SD)	38.3 (16.1)	48.5 (13.1)	**<.01**
DUP (*n* = 220)	Mean (SD)	32.8 (59.8)	116 (219.3)	**.01**
PANSS total (*n* = 280)	Mean (SD)	99.4 (23.8)	76.1 (15.1)	**<.01**
PANSS negative total (*n* = 280)	Mean (SD)	21.1 (9.7)	19.4 (6.3)	.07
Remission (*n* = 280)	*N* (%)	102 (57%)	31 (31%)	**<.01**
Cannabis (*n* = 278)	*N* (%)	93 (53%)	4 (4%)	**<.01**
Stimulant (*n* = 279)	*N* (%)	20 (11%)	0 (0%)	**<.01**
Opioid (*n* = 269)	*N* (%)	1 (1%)	0 (0%)	.44
Hallucinogens (*n* = 271)	*N* (%)	11 (6%)	0 (0%)	**.01**
Tobacco (*n* = 280)	*N* (%)	97 (54%)	13 (13%)	**<.01**

B) Affective vs Nonaffective	Descriptive Statistic	Affective N =32	Nonaffective N =133	Affective vs Nonaffective P-value*
Age (*n* = 162)	Mean (SD)	25.2 (7.2)	25.9 (7.6)	.83
Gender, male (*n* = 165)	*N* (%)	14 (44%)	94 (71%)	.03
Education (years) (*n* = 162)	Mean (SD)	13.9 (2.5)	13.5 (2.5)	.75
GAF baseline (*n* = 138)	Mean (SD)	40.4 (15.8)	41.4 (15.6)	.90
PSP baseline total score (*n* = 124)	Mean (SD)	33.6 (13.6)	38.2 (15.6)	.35
DUP (*n* = 157)	Mean (SD)	19.7 (48.8)	37.1 (63.6)	.26
DUI (*n* = 137)	Mean (SD)	63.9 (181.4)	59.2 (80.6)	.90
PANSS total (*n* = 165)	Mean (SD)	89 (18.9)	101.4 (24.3)	.02
PANSS positive total (*n* = 165)	Mean (SD)	28.7 (7.6)	28.4 (6.8)	090
PANSS negative total (*n* = 165)	Mean (SD)	14.8 (6.5)	22.5 (9.7)	<.01
PANSS general total (*n* = 165)	Mean (SD)	45.5 (10.8)	50.6 (14.1)	.09
Remission (*n* = 165)	*N* (%)	26 (81%)	69 (52%)	.03
Cohabitants (*n* = 161)	Mean (SD)	2.2 (2)	2.5 (2)	.81
HAM total (*n* = 127)	Mean (SD)	16 (10.5)	17.2 (9.4)	.83
YMRS total (*n* = 126)	Mean (SD)	21.2 (14.7)	13.2 (9.6)	.08
Antipsychotic use prior to entry, yes (*n* = 165)	*N* (%)	10 (31%)	53 (40%)	.53
Currently living alone, yes (*n* = 147)	*N* (%)	7 (23%)	22 (19%)	.71
Currently unemployed, yes (*n* = 165)	*N* (%)	7 (22%)	43 (32%)	.46
Lifetime cannabis use, yes (*n* = 16)	*N* (%)	19 (61%)	63 (48%)	.38
Cannabis use in the past month, yes (*n* = 154)	*N* (%)	10 (36%)	28 (22%)	.35
Stimulant use last 12 months, yes (*n* = 164)	*N* (%)	5 (16%)	13 (10%)	.53
Opioid use last 12 months, yes (*n* = 154)	*N* (%)	0 (0%)	1 (1%)	.71
Hallucinogen use last 12 months, yes (*n* = 156)	*N* (%)	5 (16%)	6 (5%)	.12
Tobacco use last 12 months, yes (*n* = 165)	*N* (%)	18 (56%)	71 (53%)	.77
Alcohol use >14 units per week, yes (*n* = 165)	*N* (%)	2 (6%)	2 (2%)	.35
CGI severity of illness: 3/4/5/6/7 (*n* = 158)	Median (IQR)	6 (5–7)	6 (5–7)	.71

*Calculated using Welches *t*-test for continuous variables and chi-square tests for categorical variables. All *P*-values were FDR corrected using the Benjamini and Hochberg method. For substance use, any use in the last 12 months regardless of frequency.

This was necessary to discern whether any changes in model accuracy were attributable to diagnostic status alone or were confounded by training sample size (as these stratified analyses drastically reduced model training and testing sample sizes).

Results

Sample Characteristics

The Athens Cohort comprised 179 participants, 120 (67%) were male with an average age of 25.8 years (SD = 7.54) and an average DUP of 32.8 weeks (median = 10, SD = 59.8). Of these subjects, 102 (57%) attained remission according to the Andreasen criteria, whilst 77 did not. See tables 1 and 2 for details. Of these patients, 112 (62.9%) were antipsychotic-naïve at their baseline clinical assessment. For the remaining 66 patients, the average length of antipsychotic treatment was 3.7 days prior to assessment (SD = 2.96).

The Copenhagen cohort consisted of 101 participants, 52% were male, the average age was 23.8 (SD = 5.36) years and the average DUP was 116 weeks (median = 26,

Table 3. Performance Metrics for the FEP Classifier Trained on Athens and Tested on Copenhagen

Athens	AUC	Model Aelection (Inner Loop Train and Validate) – Athens: 5 Repeats of 10-fold CV								
		BAC	Acc	Sens	Spec	PPV	NPV	PLR	NLR	DOR
	78.52	59.33	58.1	41.95	76.7	75.64	53.76	1.80	0.76	2.38
Athens	AUC	Testing – outer loop 10-Fold CV								
		BAC	Acc	Sens	Spec	PPV	NPV	PLR	NLR	DOR
	71.45	60.39	60.05	55.09	65.00	62.86	50.26	-0.86	-0.83	1.03
Copenhagen	AUC	Independent test –Copenhagen (full sample)								
		BAC	Acc	Sens	Spec	BPPV	BNPV	PLR	NLR	DOR
	67.74	63.53	59.41	74.19	52.86	67.60	60.71	1.57	0.49	3.22

Metric abbreviations: AUC: area-under-the-curve, BAC: balanced accuracy, Acc: accuracy, Sens: sensitivity, Spec: specificity, PPV: positive predictive value, NPV: negative predictive value, BPPV: Bayes positive predictive value, NPPV: Bayes negative predictive value, PLR: positive likelihood ratio, NLR: negative likelihood ratio, DOR: diagnostic odds ratio.

SD = 219.3). Remission criteria were fulfilled by 31 subjects in this group when the Andreasen criteria were applied, six weeks following initial assessment.

Remission Prediction

For predicting remission 4 weeks after a FEP in the Athens cohort using repeated nested cross-validation we attained an average inner loop training AUC = 78.52, BAC = 59.33, PPV = 75.64, and NPV = 53.76 and an average outer loop test set AUC = 71.45 (SD = 11.23, *P* = .0001), BAC = 60.39, PPV = 62.86, and NPV = 50.26. Refitting this model to the entire Athens cohort and testing on the Copenhagen cohort we attained a test AUC = 67.74 (*P* = .0001), BAC = 63.53, BPPV = 67.6, and BNPV = 60.71, further demonstrating the clinical generalizability of the model (see figure 2 and table 3 for all performance metrics). For the null distribution of the classifier, see supplementary figure S2. In this externally validated model, the final selected predictors listed in order of regularized coefficients from the Elastic Net were: PANSS-g12 (lack of judgement and insight), PANSS-n6 (lack of spontaneity and flow of conversation), DUP, PANSS-n1 (blunted affect), PANSS-n2 (emotional withdrawal), PANSS-n3 (poor rapport), PANSS-p1 (delusions), PANSS-g16 (active social avoidance), PANSS-n4 (passive/apathetic social withdrawal), PSP total score, and the GAF score. For the variability of these features under selection with the Elastic Net, see supplementary figure S3. No significant differences were found between baseline CGI scores for remission/ nonremission groups (*P* = .60). For confusion matrices of model performance by age range and gender, see supplementary figures S4–S7. For the sample size and subsample analysis results, see supplementary Methods 1.7. For the results using the second set of hyperparameters, see supplementary Appendix 1.8.

Discussion

The aim of this project was to develop a prediction model of 4-week FEP clinical remission according to the modified criterion defined by Andreasen et al.[33] Independent sample validation showed that the model developed was generalizable to patients in the Copenhagen cohort assessed for remission at 6 weeks. As in previous studies that developed cross-site predictive models, a drop in performance was observed between training and test sites.[43] Possible explanations for this variance include patient cohort variability arising from moderators such as different inclusion and exclusion criteria, geographic variability, rater variability across clinical scales, varying remission rates across the two studies, and finally, the temporal difference in the assessment of remission across studies (4 vs 6-weeks).

From a clinical perspective, the set of variables identified by the model may have utility in delineating a clinical signature that predicts remission. In previous studies, bilateral associations have been found linking remission to a shorter DUP, better social integration and less severe psychotic symptoms.[53,54] Remission has also been associated with treatment delay, symptom intensity at baseline, comorbid substance use, and time in treatment.[15] Conversely, relapse has been linked to poorer premorbid functioning, more severe negative symptoms[53] and a longer DUP.[10,54–61] A shorter DUP has been associated with a greater response to antipsychotic treatment.[58,62–64] Higher negative symptoms, in turn, have been linked to negative short-term treatment response. Regarding insight, higher scores of cognitive insight on the self-reflective scale have been found to correlate with less severe psychopathological symptoms four years after the FEP. This suggests a prospective relationship between self-reflection and patient symptom experience.[17] Therefore, given that remission, relapse, response to treatment, DUP, negative symptoms, social functioning and insight are associated, a multivariate model could be used to

Table 4. Summary of Exclusion Criteria of Athens and Copenhagen Cohorts

Criteria	Athens	Copenhagen
Age (years)	<16 or >45	<18 or >45
Period of antipsychotic treatment	>2 weeks	Not antipsychotic naïve
Physical illness, neurological illness, head trauma	Identified	Identified
Intelligence quotient	<65	<70
Current diagnosis of drug dependence	Not examined	Presence
Antidepressant, mood stabilizer use, current or in the past 1 month	Not examined	Presence
Methylphenidate exposure	Not examined	Presence
Diagnosis	Not used as an exclusion criterion	PECANS 1: Any diagnosis other than schizophrenia or schizoaffective disorder PECANS II: Any diagnosis other than schizophrenia, schizoaffective or nonorganic psychotic disorder

further the understanding of the inter-relationships between these factors.

An important consideration is the selected multivariate pattern of clinical variables that were prognostic of individual patient remission/nonremission in our model. We identified that greater severity on the following measures, namely, lack of insight (PANSS item G12), lack of spontaneity and flow of conversation (PANSS item N6), and blunted affect (PANSS item N1), together with a longer DUP significantly predicted patient nonremission. This relates to these features being considered in unison, rather than being assessed individually, where weak correlations have already been found. As mentioned above, previous studies have shown weak to moderate associations with remission,[1,15,17,25,55,61,65,66] but have not previously been studied in a multivariate classification context. Exploring the entire set of selected features, we observed that of the eight items used to operationalize remission by Andreasen et al, only four were included by our baseline feature selection. Interestingly, these items (PANSS-p1, PANSS-n1, PANSS-n4, and PANSS-n6) mainly pertain to negative symptoms, supporting previous findings.[26] The remaining features included social avoidance (PANSS-g16), social withdrawal (PANSS-n4) and two features pertaining to function (PSP, GAF). When viewed as a whole, these features indicate that negative symptoms and functioning affect remission. In addition, on statistical comparison of the remission and the nonremission groups in the Athens cohort we find that they differ with significance in the following variables. Functioning as measured by the GAF scale is higher in the remission group. The PANSS total score is higher in the nonremission group, as is the total PANSS negative score. Moreover, both the DUP and the Duration of Untreated Illness (DUI) are longer in the nonremission group. Finally, diagnosis, indicates that the remission group is more likely to receive an affective psychosis diagnosis. Considering these features for the construction of remission status in future works may be

worthwhile, as they could provide a clinical symptom signature of remission in FEP.

While this multivariate model is clinically useful as a tool within its own right, considering the broader clinical profile may be of use for personalized interventions following patient stratification. Since the Athens study was not framed around the development of a predictive remission tool, clinician prognostication was not measured *a priori*. However, as has been demonstrated in previous studies,[67,68] we used the CGI scale as a proxy indicator of remission. When stratified by remission/nonremission status in the Athens cohort, there were no statistically significant differences between the two groups. This indicates that clinicians did not distinguish between the two groups upon first presentation regarding differences in illness severity. It could therefore be argued that in accordance with previous tests of clinician's prognostic abilities,[43] that they may be no better than chance at prognosticating patient outcomes. However, future comparator trials are necessary before this conclusion can be safely reached.

Regarding the clinical usefulness of our classification model, classifying one-month remission status of FEP patients on their baseline symptomatic presentation could provide clinicians with a valuable tool to guide support and treatment interventions and assist patients to work towards social reintegration and recovery following their first psychotic episode. A negative prognostic evaluation would aid the clinician towards more focused clinical follow-ups and more assertive treatment plans. After correcting for the difference in prevalence rates between the Athens and Copenhagen cohorts, we observed that our predictive model trained on the Athens cohort generalized successfully to the Copenhagen cohort (Athens PPV = 62.86, NPV = 50.26, Copenhagen PPV = 67.6, NPV = 60.71). Considering these, we conclude that remission is more reliably predicted than nonremission. However, given the moderate PPV of the model, it is likely that the model would be best suited to support

clinical decision making in cases of fluctuating or unstable clinical presentation rather than in routine clinical assessment. However, forward tested clinical trials will first be required to support this use case.

There were a number of factors affecting the performance of our model. Firstly, the temporal shifts in the first available assessments of remission (Athens = 4-weeks, Copenhagen = 6-weeks) and the mixed diagnostic status of the Athens cohort compared to the Copenhagen cohort. Whilst these factors likely also played a role, they reflect natural heterogeneity, providing a robust validation of the model's ability to generalize beyond the scope of the patient characteristics observed in the Athens cohort. However, to understand the effect of these diagnostic categories on overall model performance, we conducted analyses consisting of sub-samples that were made up of all possible train/test combinations of diagnoses. However, no model performed better than chance and only learnt the dominant class proportion (the no-information rate). Given that when these diagnostic categories were combined our model performed significantly better than chance, it was possible that we were statistically underpowered for these sub-analyses. Therefore, we conducted a second sub-analysis exploring the effects of training sample size on balanced accuracy. Here, we found that our model had still not ceiled in balanced accuracy even when using the full Athens sample, and was no better than chance when using samples equivalent in size to those in each diagnostic category (see supplementary figure S4). Therefore, analyses stratified by these diagnostic groups lacked the statistical power to delineate these effects and should instead be investigated in future works on larger samples. Finally, it is important to add that treatment effects were not controlled for in the development of the prediction model. This was not possible due to the TAU conditions and the wide variety of treatments used in the Athens cohort. It is worth mentioning that despite this limitation, the derived model performed well in a cohort with differing treatment conditions.

Regarding study strengths, the predictive model was developed on data from a naturalistic study in treatment as usual (TAU) conditions. Remission was operationalized according to standardized criteria.[33] Finally, the generalizability of the derived model was investigated through external validation on an independent sample, showing good generalizability. Whilst promising, several limitations still exist in our study. Increasing sample size will plausibly improve the generalizability of the model by identifying a wider range of clinical presentations given the established clinical and biological heterogeneity of psychosis. Moreover, this will afford the opportunity to test higher complexity nonlinear models that have demonstrated their efficacy and superiority in previous works.[26,43] In addition, benchmarking clinician prognostication against this model will be required before we can consider incremental utility and clinical deployment.

Finally, selection effects may be prevalent in our sample as FEP patients may differ in their psychopathological characteristics compared to those who did not consent to participation.[25]

Regarding future considerations, the use of similarly developed algorithms could inform clinical guidelines by offering a unique insight into symptom progression on an individual basis. Further, other clinical targets could be pursued, facilitating multilevel predictors and models that could help personalize both patient prognosis and subsequent treatment. Finally, a multimodal model containing biomarker predictors could be considered in order to enhance the predictive power of clinical models.[51,69–72]

Supplementary Material

Supplementary material is available at *Schizophrenia Bulletin* online.

Acknowledgments

References

1. Fusar-Poli P, McGorry PD, Kane JM. Improving outcomes of first-episode psychosis: an overview. *World Psychiatry*. 2017;16(3):251–265.

2. Chong HY, Teoh SL, Wu DB, Kotirum S, Chiou CF, Chaiyakunapruk N. Global economic burden of schizophrenia: a systematic review. *Neuropsychiatr Dis Treat*. 2016;12:357–373.

3. Walker ER, McGee RE, Druss BG. Mortality in mental disorders and global disease burden implications a

systematic review and meta-analysis. *JAMA Psychiatry*. 2015;72(4):334–341.

4. Millan MJ, Andrieux A, Bartzokis G, et al. Altering the course of schizophrenia: progress and perspectives. *Nat Rev Drug Discov*. 2016;15(7):485–515.

5. Fusar-Poli P, Díaz-Caneja CM, Patel R, et al. Services for people at high risk improve outcomes in patients with first episode psychosis. *Acta Psychiatr Scand*. 2016;133(1):76–85.

6. Kane JM, Robinson DG, Schooler NR, et al. Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the NIMH RAISE early treatment program. *Am J Psychiatry*. 2016;173(4):362–372.

7. Ajnakina O, Morgan C, Gayer-Anderson C, et al. Only a small proportion of patients with first episode psychosis come via prodromal services: a retrospective survey of a large UK mental health programme. *BMC Psychiatry*. 2017;17(1):308.

8. Cheng SC, Schepp KG. Early intervention in schizophrenia: a literature review. *Arch Psychiatr Nurs*. 2016;30(6):774–781.

9. Chang WC, Lau ESK, Chiu SS, et al. Three-year clinical and functional outcome comparison between first-episode mania with psychotic features and first-episode schizophrenia. *J Affect Disord*. 2016;200:1–5.

10. Friis S, Melle I, Johannessen JO, et al. Early predictors of ten-year course in first-episode psychosis. *Psychiatric Services*. 2016;67(4):438–443.

11. Austin SF, Mors O, Budtz-Jørgensen E, et al. Long-term trajectories of positive and negative symptoms in first episode psychosis: a 10-year follow-up study in the OPUS cohort. *Schizophr Res*. 2015;168(1-2):84–91.

12. Alvarez-Jimenez M, O'Donoghue B, Thompson A, et al. Beyond clinical remission in first episode psychosis: thoughts on antipsychotic maintenance vs. guided discontinuation in the functional recovery era. *CNS Drugs*. 2016;30(5):357–368.

13. Di Capite S, Upthegrove R, Mallikarjun P. The relapse rate and predictors of relapse in patients with first-episode psychosis following discontinuation of antipsychotic medication. *Early Interv Psychiatry*. 2018;12(5):893–899.

14. Petersen L, Thorup A, Øqhlenschlaeger J, et al. Predictors of remission and recovery in a first-episode schizophrenia spectrum disorder sample: 2-year follow-up of the OPUS trial. *Can J Psychiatry*. 2008;53(10):660–670.

15. Conus P, Cotton S, Schimmelmann BG, McGorry PD, Lambert M. Rates and predictors of 18-months remission in an epidemiological cohort of 661 patients with first-episode psychosis. *Soc Psychiatry Psychiatr Epidemiol*. 2017;52(9):1089–1099.

16. David AS, Bedford N, Wiffen B, Gilleen J. Failures of metacognition and lack of insight in neuropsychiatric disorders. *Philos Trans R Soc Lond B Biol Sci*. 2012;367(1594):1379–1390.

17. O'Connor JA, Ellett L, Ajnakina O, et al. Can cognitive insight predict symptom remission in a first episode psychosis cohort? *BMC Psychiatry*. 2017;17(1):54.

18. Radua J, Ramella-Cravaro V, Ioannidis JPA, et al. What causes psychosis? An umbrella review of risk and protective factors. *World Psychiatry*. 2018;17(1):49–66.

19. Birchwood M, Todd P, Jackson C. Early intervention in psychosis: the critical-period hypothesis. *Int Clin Psychopharmacol*. 1998;172(33):53–59.

20. Agius M, Hadjinicolaou AV, Ramkisson R, et al. Does early intervention for psychosis work? An analysis of outcomes of early intervention in psychosis based on the critical period hypothesis, measured by number of admissions and bed days used over a period of six years, the first three in an early intervention service, the second three in a community mental health team. *Psychiatr Danub*. 2010;22 Suppl 1:S72–S84.

21. Correll CU, Galling B, Pawar A, et al. Comparison of early intervention services vs treatment as usual for early-phase psychosis: a systematic review, meta-analysis, and meta-regression. *JAMA Psychiatry*. 2018;75(6):555–565.

22. Lieberman JA, Small SA, Girgis RR. Early detection and preventive intervention in schizophrenia: from fantasy to reality. *Am J Psychiatry*. 2019;176(10):794–810.

23. McGorry PD. Early intervention in psychosis: obvious, effective, overdue. *J Nerv Ment Dis*. 2015;203(5):310–318.

24. Koutsouleris N, Kambeitz-Ilankovic L, Ruhrmann S, et al.; PRONIA Consortium. 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.

25. Leighton SP, Upthegrove R, Krishnadas R, et al. Development and validation of multivariable prediction models of remission, recovery, and quality of life outcomes in people with first episode psychosis: a machine learning approach. *Lancet Digit Health*. 2019;1(6):e261–e270.

26. 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.

27. Leighton SP, Krishnadas R, Chung K, et al. Predicting one-year outcome in first episode psychosis using machine learning. *PLoS One*. 2019;14(3):e0212846.

28. Crespo-Facorro B, Pelayo-Teran JM, Mayoral-van Son J. Current data on and clinical insights into the treatment of first episode nonaffective psychosis: a comprehensive review. *Neurol Ther*. 2016;5(2):105–130.

29. Talpalaru A, Bhagwat N, Devenyi GA, Lepage M, Chakravarty MM. Identifying schizophrenia subgroups using clustering and supervised learning. *Schizophr Res*. 2019;214:51–59.

30. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr Bull*. 1987;13(2):261–276.

31. Hamilton M. Rating depressive patients. *J Clin Psychiatry*. 1980;41(12 Pt 2):21–24.

32. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry*. 1978;133:429–435.

33. Andreasen NC, Carpenter WT Jr, Kane JM, Lasser RA, Marder SR, Weinberger DR. Remission in schizophrenia: proposed criteria and rationale for consensus. *Am J Psychiatry*. 2005;162(3):441–449.

34. Xenaki LA, Kollias CT, Stefanatou P, et al. Organization framework and preliminary findings from the Athens First-Episode Psychosis Research Study. *Early Interv Psychiatry*. 2020;14(3):343–355.

35. Nielsen MØ, Rostrup E, Wulff S, et al. Alterations of the brain reward system in antipsychotic naïve schizophrenia patients. *Biol Psychiatry*. 2012;71(10):898–905.

36. Bojesen KB, Ebdrup BH, Jessen K, et al. Treatment response after 6 and 26 weeks is related to baseline glutamate and GABA levels in antipsychotic-naïve patients with psychosis. *Psychol Med*. 2020;50(13):2182–2193.

37. Docherty JP, Bossie CA, Lachaux B, et al. Patient-based and clinician-based support for the remission criteria in schizophrenia. *Int Clin Psychopharmacol*. 2007;22(1):51–55.

38. Oorschot M, Lataster T, Thewissen V, et al. Symptomatic remission in psychosis and real-life functioning. *Br J Psychiatry*. 2012;201(3):215–220.

39. Lambert M, Karow A, Leucht S, Schimmelmann BG, Naber D. Remission in schizophrenia: validity, frequency, predictors, and patients' perspective 5 years later. *Dialogues Clin Neurosci*. 2010;12(3):393–407.

40. Emsley R, Rabinowitz J, Medori R; Early Psychosis Global Working Group. Remission in early psychosis: rates, predictors, and clinical and functional outcome correlates. *Schizophr Res*. 2007;89(1-3):129–139.

41. De Hert M, van Winkel R, Wampers M, Kane J, van Os J, Peuskens J. Remission criteria for schizophrenia: evaluation in a large naturalistic cohort. *Schizophr Res*. 2007;92(1–3):68–73. doi:10.1016/j.schres.2007.01.010

42. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: what is it and how does it work? *Int J Methods Psychiatr Res*. 2011;20(1):40–49.

43. 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.

44. Abu-Akel A, Bousman C, Skafidas E, Pantelis C. Mind the prevalence rate: overestimating the clinical utility of psychiatric diagnostic classifiers. *Psychol Med*. 2018;48(8):1225–1227.

45. Simonsen E, Friis S, Haahr U, et al. Clinical epidemiologic first-episode psychosis: 1-year outcome and predictors. *Acta Psychiatr Scand*. 2007;116(1):54–61.

46. Menezes NM, Malla AM, Norman RM, Archie S, Roy P, Zipursky RB. A multi-site Canadian perspective: examining the functional outcome from first-episode psychosis. *Acta Psychiatr Scand*. 2009;120(2):138–146.

47. Addington J, Addington D. Symptom remission in first episode patients. *Schizophr Res*. 2008;106(2–3):281–285.

48. Wunderink L, Sytema S, Nienhuis FJ, Wiersma D. Clinical recovery in first-episode psychosis. *Schizophr Bull*. 2009;35(2):362–369.

49. Turner MA, Boden JM, Smith-Hamel C, Mulder RT. Outcomes for 236 patients from a 2-year early intervention in psychosis service. *Acta Psychiatr Scand*. 2009;120(2):129–137.

50. Heston TF. Standardizing predictive values in diagnostic imaging research. *J Magn Reson Imaging*. 2011;33(2):505; author reply 506–505; author reply 507.

51. Cearns M, Opel N, Clark S, et al. Predicting rehospitalization within 2 years of initial patient admission for a major depressive episode: a multimodal machine learning approach. *Transl Psychiatry*. 2019;9(1):285.

52. Flint C, Cearns M, Opel N, et al. Systematic overestimation of machine learning performance in neuroimaging studies of depression. *arXiv Prepr arXiv191206686*. 2019.

53. Chen EYH, Hui CLM, Lam MML, et al. Maintenance treatment with quetiapine versus discontinuation after one year of treatment in patients with remitted first episode psychosis: randomised controlled trial. *BMJ*. 2010;341:c4024.

54. Kane JM, Rifkin A, Quitkin F, Nayak D, Ramos-Lorenzi J. Fluphenazine vs placebo in patients with remitted, acute first-episode schizophrenia. *Arch Gen Psychiatry*. 1982;39(1):70–73.

55. Bowtell M, Ratheesh A, McGorry P, Killackey E, O'Donoghue B. Clinical and demographic predictors of continuing remission or relapse following discontinuation of antipsychotic medication after a first episode of psychosis. A systematic review. *Schizophr Res*. 2018;197:9–18.

56. Marshall M, Lewis S, Lockwood A, Drake R, Jones P, Croudace T. Association between duration of untreated psychosis and outcome in cohorts of first-episode patients: a systematic review. *Arch Gen Psychiatry*. 2005;62(9):975–983.

57. Compton MT, Gordon TL, Weiss PS, Walker EF. The "doses" of initial, untreated hallucinations and delusions: a proof-of-concept study of enhanced predictors of first-episode symptomatology and functioning relative to duration of untreated psychosis. *J Clin Psychiatry*. 2011;72(11):1487–1493.

58. Perkins DO, Gu H, Boteva K, Lieberman JA. Relationship between duration of untreated psychosis and outcome in first-episode schizophrenia: a critical review and meta-analysis. *Am J Psychiatry*. 2005;162(10):1785–1804.

59. Boonstra N, Klaassen R, Sytema S, et al. Duration of untreated psychosis and negative symptoms–a systematic review and meta-analysis of individual patient data. *Schizophr Res*. 2012;142(1-3):12–19.

60. Penttilä M, Jääskéläinen E, Hirvonen N, Isohanni M, Miettunen J. Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis. *Br J Psychiatry*. 2014;205(2):88–94.

61. Allott K, Fraguas D, Bartholomeusz CF, et al. Duration of untreated psychosis and neurocognitive functioning in first-episode psychosis: a systematic review and meta-analysis. *Psychol Med*. 2018;48(10):1592–1607.

62. Nordon C, Rouillon F, Azorin JM, Barry C, Urbach M, Falissard B. Trajectories of antipsychotic response in drug-naive schizophrenia patients: results from the 6-month ESPASS follow-up study. *Acta Psychiatr Scand*. 2014;129(2):116–125.

63. Díaz I, Pelayo-Terán JM, Pérez-Iglesias R, et al. Predictors of clinical remission following a first episode of non-affective psychosis: sociodemographics, premorbid and clinical variables. *Psychiatry Res*. 2013;206(2–3):181–187. doi:10.1016/j.psychres.2012.10.011

64. Crespo-Facorro B, de la Foz VO, Ayesa-Arriola R, et al. Prediction of acute clinical response following a first episode of non affective psychosis: results of a cohort of 375 patients from the Spanish PAFIP study. *Prog Neuropsychopharmacol Biol Psychiatry*. 2013;44:162–167.

65. Schennach R, Riedel M, Musil R, Möller HJ. Treatment response in first-episode schizophrenia. *Clin Psychopharmacol Neurosci*. 2012;10(2):78–87.

66. Murru A, Carpiniello B. Duration of untreated illness as a key to early intervention in schizophrenia: a review. *Neurosci Lett*. 2018;669:59–67.

67. Pinna F, Deriu L, Diana E, et al.; Cagliari Recovery Study Group. Clinical Global Impression-severity score as a reliable measure for routine evaluation of remission in schizophrenia and schizoaffective disorders. *Ann Gen Psychiatry*. 2015;14:6.

68. Masand P, O'Gorman C, Mandel FS. Clinical Global Impression of Improvement (CGI-I) as a valid proxy measure for remission in schizophrenia: analyses of ziprasidone clinical study data. *Schizophr Res*. 2011;126(1-3):174–183.

69. Ebdrup BH, Axelsen MC, Bak N, et al. Accuracy of diagnostic classification algorithms using cognitive-, electrophysiological-, and neuroanatomical data in antipsychotic-naïve schizophrenia patients. *Psychol Med*. 2019;49(16):2754–2763.

70. Walter M, Alizadeh S, Jamalabadi H, et al. Translational machine learning for psychiatric neuroimaging. *Prog Neuropsychopharmacol Biol Psychiatry*. 2019;91:113–121.

71. Dinga R, Marquand AF, Veltman DJ, et al. Predicting the naturalistic course of depression from a wide range of clinical, psychological, and biological data: a machine learning approach. *Transl Psychiatry*. 2018; 8(1):241.

72. Ambrosen KS, Skjerbæk MW, Foldager J, et al. A machine-learning framework for robust and reliable prediction of short- and long-term treatment response in initially antipsychotic-naïve schizophrenia patients based on multimodal neuropsychiatric data. *Transl Psychiatry*. 2020;10(1):276.