JAMA Psychiatry | Original Investigation

Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression

Nikolaos Koutsouleris, MD; Dominic B. Dwyer, PhD; Franziska Degenhardt, MD; Carlo Maj, PhD; Maria Fernanda Urquijo-Castro, MSc; Rachele Sanfelici, MSc; David Popovic, MD; Oemer Oeztuerk, MD; Shalaila S. Haas, PhD; Johanna Weiske, MSc; Anne Ruef, MSc; Lana Kambeitz-Ilankovic, PhD; Linda A. Antonucci, PhD; Susanne Neufang, PhD; Christian Schmidt-Kraepelin, MD; Stephan Ruhrmann, MD; Nora Penzel, MSc; Joseph Kambeitz, MD; Theresa K. Haidl, MD; Marlene Rosen, PhD; Katharine Chisholm, PhD; Anita Riecher-Rössler, MD; Laura Egloff, PhD; André Schmidt, PhD; Christina Andreou, MD, PhD; Jarmo Hietala, MD, PhD; Timo Schirmer, PhD; Georg Romer, MD; Petra Walger, MD; Maurizia Franscini, MD; Nina Traber-Walker, PhD; Benno G. Schimmelmann, MD; Rahel Flückiger, PhD; Chantal Michel, PhD; Wulf Rössler, MD; Oleg Borisov, MD; Peter M. Krawitz, MD; Karsten Heekeren, MD; Roman Buechler, MD; Christos Pantelis, PhD, MD; Peter Falkai, MD; Raimo K. R. Salokangas, MD, PhD, MSc; Rebekka Lencer, MD; Alessandro Bertolino, MD, PhD; Stefan Borgwardt, MD; Markus Noethen, MD; Paolo Brambilla, MD; Stephen J. Wood, PhD; Rachel Upthegrove, MD, PhD; Frauke Schultze-Lutter, PhD; Anastasia Theodoridou, MD; Eva Meisenzahl, MD; for the PRONIA Consortium

 Supplemental content

IMPORTANCE Diverse models have been developed to predict psychosis in patients with clinical high-risk (CHR) states. Whether prediction can be improved by efficiently combining clinical and biological models and by broadening the risk spectrum to young patients with depressive syndromes remains unclear.

OBJECTIVES To evaluate whether psychosis transition can be predicted in patients with CHR or recent-onset depression (ROD) using multimodal machine learning that optimally integrates clinical and neurocognitive data, structural magnetic resonance imaging (sMRI), and polygenic risk scores (PRS) for schizophrenia; to assess models' geographic generalizability; to test and integrate clinicians' predictions; and to maximize clinical utility by building a sequential prognostic system.

DESIGN, SETTING, AND PARTICIPANTS This multisite, longitudinal prognostic study performed in 7 academic early recognition services in 5 European countries followed up patients with CHR syndromes or ROD and healthy volunteers. The referred sample of 167 patients with CHR syndromes and 167 with ROD was recruited from February 1, 2014, to May 31, 2017, of whom 26 (23 with CHR syndromes and 3 with ROD) developed psychosis. Patients with 18-month follow-up (n = 246) were used for model training and leave-one-site-out cross-validation. The remaining 88 patients with nontransition served as the validation of model specificity. Three hundred thirty-four healthy volunteers provided a normative sample for prognostic signature evaluation. Three independent Swiss projects contributed a further 45 cases with psychosis transition and 600 with nontransition for the external validation of clinical-neurocognitive, sMRI-based, and combined models. Data were analyzed from January 1, 2019, to March 31, 2020.

MAIN OUTCOMES AND MEASURES Accuracy and generalizability of prognostic systems.

RESULTS A total of 668 individuals (334 patients and 334 controls) were included in the analysis (mean [SD] age, 25.1 [5.8] years; 354 [53.0%] female and 314 [47.0%] male). Clinicians attained a balanced accuracy of 73.2% by effectively ruling out (specificity, 84.9%) but ineffectively ruling in (sensitivity, 61.5%) psychosis transition. In contrast, algorithms showed high sensitivity (76.0%-88.0%) but low specificity (53.5%-66.8%). A cybernetic risk calculator combining all algorithmic and human components predicted psychosis with a balanced accuracy of 85.5% (sensitivity, 84.6%; specificity, 86.4%). In comparison, an optimal prognostic workflow produced a balanced accuracy of 85.9% (sensitivity, 84.6%; specificity, 87.3%) at a much lower diagnostic burden by sequentially integrating clinical-neurocognitive, expert-based, PRS-based, and sMRI-based risk estimates as needed for the given patient. Findings were supported by good external validation results.

CONCLUSIONS AND RELEVANCE These findings suggest that psychosis transition can be predicted in a broader risk spectrum by sequentially integrating algorithms' and clinicians' risk estimates. For clinical translation, the proposed workflow should undergo large-scale international validation.

Author Affiliations: Author affiliations are listed at the end of this article.

Group Information: A complete list of the PRONIA (Personalised Prognostic Tools for Early Psychosis Management) Consortium members appears at the end of article.

Corresponding Author: Nikolaos Koutsouleris, MD, Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Nussbaumstrasse 7, D-80336 Munich, Germany (nikolaos.koutsouleris@med.uni-muenchen.de).

The clinical high-risk (CHR) criteria for psychosis have been established to detect vulnerable individuals as early as possible to intercept disease development.[1] These criteria identify a patient population with increased incidence compared with the general population,[2] yet only 22% of patients with CHR as detected by ultra–high-risk criteria show a psychosis transition during a 3-year period.[2] The clinical utility of the CHR designation may be further limited because its ascertainment is laborious and confined to specialized, well-equipped health care services that do not sufficiently cover the vulnerable population.[3,4] Hence, improved prognostic accuracy and clinical scalability are needed to accurately identify patients truly at risk for psychosis.

Prognostic accuracy may be increased using psychosis risk calculators for populations with CHR based on conventional statistics[3,5-7] or machine learning,[8-13] with studies finding that a first episode can be predicted with clinical data,[3,14] combinations of clinical and cognitive data,[6,15] neuroimaging,[8] and, recently, with polygenic risk scores (PRS) for schizophrenia,[16] among other measures.[17] However, reviews[18-20] have also highlighted methodological shortcomings, such as inadequate sample sizes and model validation strategies,[21,22] that may have inflated accuracy. Moreover, studies suggested that psychosis does not only emerge from CHR states[23,24] but occurs and can be predicted across a broader spectrum of comorbid conditions commencing in late adolescence and early adulthood.[3,14] Hence, generalizable risk prediction models may require transdiagnostic discovery and validation populations, encompassing patients with CHR states and early-onset affective syndromes that share environmental, clinical, and neurobiological features.[25-30] In addition, the growing diversity of risk prediction models originating from different data modalities has led to uncertainty about the minimum number of modalities needed to increase prognostic accuracy to a level justifying clinical implementation.[2,31] Finally, algorithms should be compared and integrated with clinicians' predictions of psychosis transition to determine their potential utility from public health and service provision perspectives.[13,32]

Addressing these challenges, the European Union–funded PRONIA study (Personalised Prognostic Tools for Early Psychosis Management [https://www.pronia.eu]) collected multimodal longitudinal data from adolescents and young adults in CHR states, those with recent-onset depression (ROD), and healthy control individuals. We evaluated clinical, neuroanatomical, and genetic machine learning models trained to identify patients with CHR syndromes and ROD who undergo psychosis transition. We compared our models' performance with our clinical raters' ability to predict psychosis transition and explored whether the sequential integration of algorithmic and expert-based prognoses produces clinically efficient cybernetic workflows,[33,34] that is, structured interactions between humans and machines that maximize prognostic accuracy while minimizing the examination burden in the given patient. We assessed potential confounders and moderators of prognostic performance and tested whether our models' and raters' estimates predicted not only psychosis transition but also distinct CHR syndromes and nonpsychotic disease

Key Points

Question Can a transition to psychosis be predicted in patients with clinical high-risk states or recent-onset depression by optimally integrating clinical, neurocognitive, neuroimaging, and genetic information with clinicians' prognostic estimates?

Findings In this prognostic study of 334 patients and 334 control individuals, machine learning models sequentially combining clinical and biological data with clinicians' estimates correctly predicted disease transitions in 85.9% of cases across geographically distinct patient populations. The clinicians' lack of prognostic sensitivity, as measured by a false-negative rate of 38.5%, was reduced to 15.4% by the sequential prognostic model.

Meaning These findings suggest that an individualized prognostic workflow integrating artificial and human intelligence may facilitate the personalized prevention of psychosis in young patients with clinical high-risk syndromes or recent-onset depression.

trajectories. Finally, we explored whether a considerably condensed, and hence less burdensome, clinical model could be generalized to 3 independent patient cohorts with CHR syndromes and other mental conditions.[35-37] This external validation step also benchmarked the neuroanatomical and combined models derived from the PRONIA cohort.

Methods

Study Design and Population

The eMethods section in the Supplement details all of the methods for this prognostic study, which followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) guideline.[38] In summary, our analysis included 334 patients with CHR states (n = 167) or ROD (n = 167) recruited across 7 sites in Finland, Germany, Italy, Switzerland, and the United Kingdom from February 1, 2014, to May 31, 2017, using standardized inclusion and exclusion criteria (eTables 1 and 2 and CONSORT diagram in eFigure 1 in the Supplement). Follow-up for all patients ranged from 9 to 36 months,[13] with visits every 3 months to the 18-month point and every 9 months thereafter. Furthermore, 334 healthy controls matched for age, sex, and site were included to evaluate prognostic patterns. Adult participants gave informed consent before study inclusion. Participants younger than 18 years and their guardians provided their written informed assent and consent. The PRONIA observational study was registered at the German Clinical Trials Register (DRKS00005042) and approved by all local research ethics committees.

Sociodemographic and clinical variables were compared between diagnostic groups (eTable 3 in the Supplement), patients with psychosis transition and nontransition, and patients with 18-month or later follow-up data (discovery sample [n = 246]; PRONIA plus 18M) or earlier attrition (validation sample [n = 88]; PRONIA minus 18M) (Table 1). Psychosis transition was defined when at least 1 of the 5 positive symptom items in the Structured Interview for Psychosis–Risk

Table 1. Study-Related, Sociodemographic, Physical, Functional, and Clinical Differences in Patients With vs Without Transition to Psychosis During the Follow-up Period

Variable	Psychosis transition cohort (n = 26)	Nontransition cohort (n = 308)	Statistical analysis	P value (FDR)[a]	PRONIA plus 18M cohort (n = 246)	PRONIA minus 18M cohort (n = 88)	Statistical analysis	P value (FDR)[b]
Samples and study variables								
Sample site, No. of patients								
Munich	10	99			86	23		
Milan	2	23			19	6		
Basel	2	32			22	12		
Cologne	4	59	$\chi^2_6 = 6.78$	.56	50	13	$\chi^2_6 = 6.53$	.83
Birmingham	2	33			25	11		
Turku	6	32			24	13		
Udine	0	30			20	10		
Follow-up interval, mean (SD), d	628.3 (316.7)	727.6 (310.7)	$t_{328} = -1.45$	.38	842.7 (272.3)	390.6 (99.6)	$t_{328} = 22.10$	<.001
Sociodemographic data								
Age, mean (SD), y	23.8 (5.6)	24.8 (5.8)	$t_{332} = -0.81$	.65	24.6 (5.6)	25.0 (6.1)	$t_{332} = -0.59$	.86
Female, No (%)	10 (38.5)	160 (51.9)	$\chi^2_1 = 1.75$	.38	124 (51.6)	46 (52.3)	$\chi^2_1 = 0.09$	.93
Race/ethnicity, No. (%)								
White	24 (92.3)	268 (87.0)			214 (87.0)	78 (88.6)		
Asian	1 (3.9)	23 (7.5)			20 (8.1)	4 (4.5)		
African	1 (3.9)	3 (1.0)	$\chi^2_4 = 3.35$	.65	3 (1.2)	1 (1.1)	$\chi^2_4 = 4.18$	.83
Mixed	0	5 (1.6)			2 (0.8)	3 (3.4)		
Other	0	9 (2.9)			7 (2.8)	2 (2.3)		
BMI, mean (SD)	23.5 (4.7)	23.7 (4.6)	$t_{331} = -0.17$	.92	23.6 (4.6)	24.0 (4.6)	$t_{331} = -0.72$	.83
Edinburgh Handedness score, mean (SD)[c]	56.7 (66.6)	70.4 (50.6)	$t_{303} = -1.28$	.38	68.0 (53.1)	72.9 (49.4)	$t_{303} = -0.70$	.83
Education, mean (SD), y	13.3 (2.5)	14.3 (2.9)	$t_{331} = -1.65$	.28	14.3 (3.0)	14.1 (2.7)	$t_{331} = 0.52$	.86
Educational problems, mean (SD), y repeated	0.67 (0.88)	0.26 (0.61)	$t_{323} = 3.20$	.02	0.30 (0.64)	0.27 (0.66)	$t_{323} = -0.34$	.93
Having a partnership most of the time in the year before study inclusion, No. (%)	8 (30.8)	88 (28.6)	$\chi^2_1 = 0.57$	.90	76 (30.9)	20 (22.7)	$\chi^2_1 = 2.11$	.59
CHR criteria met, No. (%)								
Schizotypal personality disorder present	2 (7.7)	9 (2.9)	$\chi^2_1 = 1.71$	.38	10 (4.1)	1 (1.1)	$\chi^2_1 = 1.75$	.67
First-degree relatives with psychosis	3 (11.5)	25 (8.1)	$\chi^2_1 = 0.37$	.83	20 (8.1)	8 (9.1)	$\chi^2_1 = 0.08$	.93
30% Loss of global functioning compared with highest levels in the year before study inclusion	9 (34.6)	88 (28.6)	$\chi^2_1 = 0.43$	.81	72 (29.3)	25 (28.4)	$\chi^2_1 = 0.02$	.93
Criteria								
GRDP	2 (7.7)	23 (7.5)	$\chi^2_1 = 0.002$	1.00	18 (7.3)	7 (8.0)	$\chi^2_1 = 0.04$	.93
COGDIS	11 (42.3)	83 (26.9)	$\chi^2_1 = 2.80$	.29	69 (28.0)	25 (28.4)	$\chi^2_1 = 0.004$	.95
APS	18 (69.2)	88 (28.6)	$\chi^2_1 = 18.30$	<.001	85 (34.6)	21 (23.9)	$\chi^2_1 = 3.42$	.58
BLIPS	2 (7.7)	5 (1.6)	$\chi^2_1 = 4.30$	.28	4 (1.6)	3 (3.4)	$\chi^2_1 = 1.00$	.83
CHR	23 (88.5)	144 (46.8)	$\chi^2_1 = 16.68$	<.001	126 (51.2)	41 (46.6)	$\chi^2_1 = 0.55$	.83
Functioning, mean (SD)								
GF:S[d]								
Highest lifetime	7.85 (0.78)	7.96 (0.88)	$z = -0.78$	.65	7.91 (0.86)	8.06 (0.91)	$z = -1.45$	.59
Baseline	5.92 (1.72)	6.44 (1.33)	$z = -1.32$	.38	6.34 (1.38)	6.57 (1.34)	$z = -1.22$	.70
GF:R[d]								
Highest lifetime	7.92 (0.74)	8.11 (0.83)	$z = -1.27$	.38	8.15 (0.78)	7.97 (0.94)	$z = -11.73$	.59
Baseline	5.85 (1.78)	6.08 (1.66)	$z = -0.57$	.75	6.07 (1.70)	6.06 (1.56)	$z = -0.16$	.93
High-risk symptoms, mean (SD)								
SIPS[e]								
Positive symptoms	1.77 (0.81)	0.83 (0.80)	$t_{332} = -5.06$	<.001	0.91 (0.82)	0.89 (0.90)	$t_{332} = 0.25$	.93
Negative symptoms	1.90 (1.57)	1.55 (1.02)	$t_{332} = -0.54$	.75	1.58 (1.04)	1.48 (0.99)	$t_{332} = 1.20$	.70
Disorganized symptoms	1.07 (1.15)	0.66 (0.60)	$t_{332} = -1.84$	.26	0.69 (0.62)	0.60 (0.55)	$t_{332} = 1.54$	.59
General psychopathology	2.29 (1.19)	1.86 (0.95)	$t_{332} = -1.65$	.28	1.92 (0.94)	1.71 (0.95)	$t_{332} = 2.24$	.47

(continued)

Table 1. Study-Related, Sociodemographic, Physical, Functional, and Clinical Differences in Patients With vs Without Transition to Psychosis During the Follow-up Period (continued)

Variable	Psychosis transition cohort (n = 26)	Nontransition cohort (n = 308)	Statistical analysis	P value (FDR)[a]	PRONIA plus 18M cohort (n = 246)	PRONIA minus 18M cohort (n = 88)	Statistical analysis	P value (FDR)[b]
History of *DSM-IV* comorbid disorders at study inclusion								
Any affective, substance, anxiety, or eating disorders at study inclusion, diagnosis, No. (%)								
None	5 (19.2)	135 (44.1)			102 (41.8)	39 (44.8)		
1	13 (50.0)	93 (30.4)	$\chi^2_3 = 8.20$	.21	78 (32.0)	27 (31.0)	$\chi^2_3 = 1.12$	.93
2	6 (23.1)	42 (13.7)			38 (15.6)	10 (11.5)		
≥3	2 (7.7)	36 (11.8)			26 (10.7)	11 (12.6)		
Major depressive disorder, No. (%)								
No	9 (34.6)	58 (19.0)	$\chi^2_1 = 3.65$	.26	47 (19.2)	20 (23.3)	$\chi^2_1 = 0.58$	.83
Yes	17 (65.4)	248 (81.0)			198 (80.8)	67 (77.0)		
Affective disorder, No. (%)[f]								
None	12 (46.2)	200 (65.4)			154 (62.9)	59 (67.8)		
1	14 (53.8)	98 (32.0)	$\chi^2_2 = 5.45$	.26	86 (35.1)	25 (28.7)	$\chi^2_2 = 1.55$	.83
2	0	8 (2.6)			5 (2.0)	3 (3.4)		
Substance use disorder, No. (%)								
No	25 (96.2)	302 (98.7)	$\chi^2_2 = 1.64$	>.99	241 (98.8)	86 (98.9)	$\chi^2_1 = 0.003$	.95
Yes	1 (3.8)	3 (1.0)			3 (1.2)	1 (1.1)		
Anxiety disorders, No. (%)								
None	17 (65.4)	212 (69.3)			172 (70.2)	58 (66.7)		
1	7 (26.9)	59 (19.3)	$\chi^2_3 = 1.35$	.86	45 (18.4)	21 (24.1)	$\chi^2_3 = 1.92$	.86
2	1 (3.8)	25 (8.2)			21 (8.6)	5 (5.7)		
≥3	1 (3.8)	10 (3.3)			7 (2.9)	3 (3.4)		
Eating disorders, No. (%)								
None	24 (92.3)	290 (94.8)			230 (93.9)	84 (96.6)		
1	2 (7.7)	13 (4.2)	$\chi^2_2 = 0.90$	.83	14 (5.7)	1 (1.1)	$\chi^2_2 = 5.55$	.58
2	0	3 (1.0)			1 (0.4)	2 (2.3)		
Treatments, No. (%)								
Antipsychotics	10 (38.5)	53 (17.2)	$\chi^2_1 = 7.08$	.09	40 (20.2)	14 (15.9)	$\chi^2_1 = 0.76$	.83
Antidepressants	13 (50.0)	176 (57.1)	$\chi^2_1 = 0.50$	.74	140 (57.6)	48 (54.5)	$\chi^2_1 = 0.25$	.86
Inpatient	15 (60.0)	144 (46.8)	$\chi^2_1 = 2.10$	.38	122 (50.2)	36 (40.9)	$\chi^2_1 = 2.24$	.59
Psychotherapy	14 (56.0)	229 (74.4)	$\chi^2_1 = 5.09$	.21	179 (73.7)	62 (70.5)	$\chi^2_1 = 0.34$	.86

Abbreviations: APS, attenuated positive symptom; BLIPS, brief limited intermittent psychotic symptoms; BMI, body mass index (calculated as weight in kilograms divided by square of height in meters); CHR, clinical high-risk; COGDIS, Schizophrenia Proneness Instrument: Cognitive Disturbances criteria; FDR, false discovery rate; GF:R Global Functioning Scale: Role; GF:S, Global Functioning Scale: Social; GRDP, genetic risk and deterioration psychosis risk syndrome; PRONIA plus 18M, Personalised Prognostic Tools for Early Psychosis Management (PRONIA) with follow-up of at least 18 months; PRONIA minus 18M, PRONIA with follow-up of less than 18 months; SIPS, Structured Interview for Psychosis–Risk Syndromes.

[a] Calculated as psychosis transition vs nontransition groups.

[b] Calculated as PRONIA plus 18M vs PRONIA minus 18M samples.

[c] Scores range from −100 to 100, with higher scores indicating more pronounced right-handedness.

[d] Scores range from 0 to 10, with higher scores indicating better social functioning.

[e] Scores range from 0 to 6, with higher scores indicating more severe symptoms.

[f] Excludes major depressive disorder.

Syndromes[39] reached psychotic intensity daily for at least 7 days.[40] Diagnoses of cases with psychosis transition are listed in eTable 4 in the Supplement.

Prognostic Modeling Strategy

Data were analyzed from January 1, 2019, to March 31, 2020. Using the machine learning software NeuroMiner, version 1.05 (GitHub [https://github.com/neurominer-git/NeuroMiner-1]), we constructed and tested unimodal, multimodal, and clinically scalable sequential risk calculators for transition prediction in the PRONIA plus 18M cohort using leave-one-site-out

cross-validation (LOSOCV)[21,41] (eMethods and eFigures 2-4 in the Supplement). We evaluated the risk calculators using baseline and longitudinal data and validated their specificity in the PRONIA minus 18M sample, which did not include cases with psychosis transition but provided all data modalities. In addition, 3 external data sets consisting of cases with psychosis transition and nontransition were available to test selected models.[35-37]

First, unimodal risk calculators were trained with literature-based baseline predictors of psychosis transition,[2,6,16] including prodromal symptoms,[39,42] functioning,[43,44] child-

hood adversity,[45,46] and neurocognitive measures[6,47] in the clinical-neurocognitive domain (eTable 5 in the Supplement); PRS for schizophrenia[48] in the genetic domain[16]; and gray matter volume maps in the structural magnetic resonance imaging (sMRI) domain[8,9] (eTable 6 in the Supplement). In addition, we evaluated our raters' predictions, which, at the conclusion of baseline assessments, were provided as yes or no replies to the question, "Do you think the patient will likely transition to psychosis?" Then we assessed whether combining unimodal algorithms using stacked generalization[49] improved prognostic accuracy (eFigures 2 and 3 in the Supplement and **Table 2**).[13] Following the concept of expert-based machine learning,[50] we integrated our raters' estimates as additional predictors to produce a cybernetic model[33] (eFigure 2 in the Supplement). Models' predictive signatures were visualized in **Figures 1** and **2** and eFigure 5 in the Supplement using measures of pattern element stability (cross-validation ratio) and pattern element significance (sign-based consistency; eMethods in the Supplement). In addition, the prognostic models were assessed using random-label permutations (Table 2). Raters' and models' performances were compared statistically at the omnibus level using the Quade test,[51] an extension of the nonparametric Wilcoxon signed rank test, followed by post hoc pairwise mean differences tests using the *t* distribution.[52] Statistical significance was determined at α = .05. Obtained *P* values were 2-sided; *P* values computed in the pairwise classifier comparisons were corrected using the false discovery rate (FDR) (**Figure 3**). Classifiers were visually compared in eFigure 6 in the Supplement).[51,52]

In-Depth Model Evaluation

We tested differences between the prognostic assignment groups and the matched healthy control group to determine whether the models' predictive patterns represented a deviation from normality (eFigures 7-9 in the Supplement). Potential model confounders and moderators were systematically tested, including image quality (eFigure 10 in the Supplement), treatments (eTable 7 in the Supplement), follow-up frequency and duration (eFigure 11 and eTable 8 in the Supplement), site-related effects (eTables 9-11 in the Supplement), baseline study group membership (eTables 12 and 13 in the Supplement), and, specifically, the inclusion of patients with ROD (eTable 14 and eFigure 12 in the Supplement). Using the patients' longitudinal data, we evaluated whether model predictions were not specific for the binary transition vs nontransition outcome, but we also separated transitions from nonremitting/de novo CHR symptom courses (P-CHR) and asymptomatic/nonpersisting trajectories (NP-CHR) (eTable 15 in the Supplement). To further explore a prognostic generalization effect,[13] we used unsupervised machine learning (eMethods in the Supplement) to construct fine-grained CHR syndrome trajectories (eFigure 13A in the Supplement) and used linear mixed-effects modeling to compare trajectories between predicted and observed outcome groups (eFigure 13B [part 1] in the Supplement). Finally, we investigated whether assignments generalized to the prediction of nonpsychotic outcomes (eTable 16 in the Supplement).

Optimization for Clinical Scalability

To facilitate clinical implementation, we developed a sequential prediction method that optimizes the ordering and number of data modalities as well as the prognostic uncertainty thresholds to decide whether a patient needs further testing (eTable 17 and eMethods in the Supplement). We analyzed the identified optimal prognostic workflow (eFigure 14 in the Supplement) and tested whether it achieved similar performance as the fully stacked models at lower diagnostic burden for the patients. To further enhance clinical scalability, we condensed the clinical-neurocognitive model, which was the workflow's entry point, from 141 to 7 (5.0%) variables (Figure 2) using sign-based consistency mapping (eMethods in the Supplement).[53] We tested the condensed model and the respective workflow's specificity in the PRONIA minus 18M sample (Table 2). Finally, we explored whether the prognostic sequence could be further trimmed using diagnostic parsimony regularization (eMethods in the Supplement and Table 2). Nonregularized and regularized workflows including the full or condensed clinical-neurocognitive models were compared in eFigure 15 in the Supplement.

External Validation Experiments

We validated the condensed clinical-neurocognitive model in 2 external cohorts: 146 patients with CHR (aged 15-35 years; 16 [11.0%] transitions) provided by the Zurich Early Recognition Program (ZInEP) (eTable 18 in the Supplement)[36] and 462 patients with diverse mental conditions (aged 8-17 years; 13 [2.8%] transitions) drawn from the Bi-national Evaluation of At-Risk Symptoms in Children and Adolescents study (eTable 19 in the Supplement).[37] Second, we validated the sMRI-based model in ZInEP and in 37 patients with CHR (16 [43.2%] transitions) from the Früherkennung von Psychosen study (Table 2).[35,54] To validate the increased performance of multimodal risk calculators, we trained a stacked model using the condensed clinical-neurocognitive and sMRI-based models, tested it in the ZInEP data (Table 2), and used the Quade test[51] to compare the 2 unimodal prediction models with the stacked classifier (Figure 3). Finally, we made our models available in the NeuroMiner Model Library (http://www.proniapredictors.eu) to facilitate their independent external validation.

Results

Group-Level Differences

A total of 668 patients and controls were included in the analysis (mean [SD] age, 25.1 [5.8] years; 354 [53.0%] female and 314 [47.0%] male). Patients in the PRONIA plus 18M and PRONIA minus 18M groups were followed up for a mean (SD) of 842.7 (272.3) and 390.6 (99.6) days, respectively (Table 1). They did not differ in any examined variable (Table 1). Psychosis transition occurred after a mean (SD) of 246.9 (244.5) days in 26 cases and developed into schizophrenia in 8 (30.8%) (eTable 4 in the Supplement). Follow-up durations differed between sites but not time to psychosis transition (eTable 9 in the Supplement). Compared with nontransition, individuals with psychosis transition had more repeated school years (mean [SD],

Table 2. Prediction Performance of Clinical Raters; Unimodal, Stacked, and Cybernetic Risk Calculators; and Prognostic Workflows[a]

Model by cohort	No. of findings				Sensitivity, %	Specificity, %	BAC, %	PPV	NPV	PSI	Positive LR	AUC	P value for FDR
	True positive	True negative	False positive	False negative									
Rater-based estimates													
PRONIA plus 18M	16	185	33	10	61.5	84.9	73.2	32.7	94.9	27.5	4.1	0.73	NA
PRONIA minus 18M	NA	75	11	NA	NA	87.2	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	16	260	44	10	61.5	85.5	73.5	26.7	96.3	23.0	4.3	0.74	NA
Clin-NC risk calculator													
PRONIA plus 18M	22	147	73	4	84.6	66.8	75.7	23.2	97.4	20.5	2.6	0.83	NA
PRONIA minus 18M[b]	NA	58	30	NA	NA	65.9	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	21	201	107	5	80.8	65.3	73.0	16.4	97.6	14.0	2.3	0.79	<.001
Condensed Clin-CN risk calculator using 7 significant predictors													
PRONIA minus 18M[b]	NA	58	30	NA	NA	65.9	NA	NA	NA	NA	NA	NA	NA
ZInEP[b]	14	74	56	2	87.5	43.1	65.3	15.9	96.6	12.5	1.5	0.67	NA
BEARS-Kid[b]	10	287	162	3	76.9	63.9	70.4	5.8	99.0	4.8	2.1	0.68	NA
PRS-based risk calculator													
PRONIA plus 18M	19	113	88	6	76.0	56.2	66.1	17.8	95.0	12.7	1.7	0.74	NA
PRONIA minus 18M[b]	NA	35	37	NA	NA	48.6	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	22	157	116	3	88.0	57.5	72.8	15.9	98.1	14.1	2.1	0.74	<.001
sMRI-based risk calculator													
PRONIA plus 18M	22	116	101	3	88.0	53.5	70.7	17.9	97.5	15.4	1.9	0.70	NA
PRONIA minus 18M[b]	NA	40	48	NA	NA	45.5	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	22	164	137	3	88.0	54.5	71.2	13.8	98.2	12.0	1.9	0.73	<.001
FePsy[b]	12	13	8	4	75.0	61.9	68.5	60.0	76.5	36.5	2.0	0.71	NA
ZInEP[b]	12	79	51	4	75.0	60.8	67.9	19.0	95.2	14.2	1.9	0.71	NA
Stacked risk calculator analyzing the predictions of the Clin-NC and sMRI classifiers													
PRONIA minus 18M[b]	NA	70	18	NA	NA	79.5	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	21	265	43	5	80.8	86.0	83.4	32.8	98.1	31.0	5.8	0.89	NA
ZInEP[b]	12	88	42	4	75.0	67.7	71.3	22.2	95.7	17.9	2.3	0.74	NA
Stacked Clin-NC, PRS, and sMRI risk calculator													
PRONIA plus 18M	21	187	33	5	80.8	85.0	82.9	38.9	97.4	36.3	5.4	0.88	NA
PRONIA minus 18M[b]	NA	71	17	NA	NA	80.7	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	18	263	45	8	69.2	85.4	77.3	28.6	97.0	25.6	4.7	0.86	<.001
Cybernetic risk calculator including raters, Clin-NC, PRS, and sMRI													
PRONIA plus 18M	22	190	30	4	84.6	86.4	85.5	42.3	97.9	40.2	6.2	0.90	NA
PRONIA minus 18M[b]	NA	73	15	NA	NA	83.0	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	21	266	42	5	80.8	86.4	83.6	33.3	98.2	31.5	5.9	0.90	<.001
Optimal prognostic workflow[c]													
PRONIA plus 18M	22	192	28	4	84.6	87.3	85.9	44.0	98.0	42.0	6.6	0.90	NA
PRONIA minus 18M[b]	NA	74	14	NA	NA	84.1	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	21	267	41	5	80.8	86.7	83.7	33.9	98.2	32.0	6.1	0.89	<.001
Prognostic workflows optimized for clinical scalability													
PRONIA minus 18M cohort[b,c]	NA	74	14	NA	NA	84.1	NA	NA	NA	NA	NA	NA	NA
Workflow optimized for light examination sparsity[d]													
PRONIA plus 18M	21	189	31	5	80.8	85.9	83.3	40.4	97.4	37.8	5.7	0.87	NA
PRONIA minus 18M[b]	NA	74	14	NA	NA	84.1	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	19	266	42	7	73.1	86.4	79.7	31.1	97.4	28.6	5.4	0.85	<.001

(continued)

Table 2. Prediction Performance of Clinical Raters; Unimodal, Stacked, and Cybernetic Risk Calculators; and Prognostic Workflows[a] (continued)

Model by cohort	No. of findings				Sensitivity, %	Specificity, %	BAC, %	PPV	NPV	PSI	Positive LR	AUC	P value for FDR
	True positive	True negative	False positive	False negative									
PRONIA minus 18M cohort[b,d]	NA	75	13	NA	NA	85.2	NA	NA	NA	NA	NA	NA	NA
Workflow optimized for strong examination sparsity[e]													
PRONIA plus 18M	22	164	56	4	84.6	74.5	79.6	28.2	97.6	25.8	3.3	0.84	NA
PRONIA minus 18M[b]	NA	67	21	NA	NA	76.1	NA	NA	NA	NA	NA	NA	NA
Complete PRONIA	21	228	80	5	80.8	74.0	77.4	20.8	97.9	18.6	3.1	0.81	<.001
PRONIA minus 18M cohort[b,e]	NA	63	25	NA	NA	71.6	NA	NA	NA	NA	NA	NA	NA

Abbreviations: AUC, area under the curve; BAC, balanced accuracy; BEARS-Kid, Bi-national Evaluation of At-Risk Symptoms in Children and Adolescents; Clin-NC, clinical-neurocognitive; FDR, false discovery rate; LR, likelihood ratio; NA, not applicable; NPV, negative predictive value; PPV positive predictive value; PRONIA plus 18M, Personalised Prognostic Tools for Early Psychosis Management (PRONIA) with follow-up of at least 18 months; PRONIA minus 18M, PRONIA with follow-up of less than 18 months; PRS, polygenic risk score; PSI, Prognostic Summary Index; sMRI, structural magnetic resonance imaging; ZInEP, Zurich Early Recognition Program.

[a] Risk calculators were first trained and cross-validated in the PRONIA plus 18M cohort and then validated in the PRONIA minus 18M sample. To estimate the models' significance, they were retrained and cross-validated using the complete PRONIA cohort. Model significance was computed for each risk calculator using 1000 label permutations in the complete PRONIA cohort (eMethods in the Supplement), and P values were corrected for multiple comparisons using the FDR. External validation was conducted for the condensed Clin-NC model, the sMRI-based risk calculator, and the stacked

model. These models were retrained in the complete PRONIA cohort before external validation. In addition, prognostic workflows that included the condensed Clin-NC model were validated in the PRONIA minus 18M sample. eFigure 2 and the eMethods in the Supplement give a detailed description of the entire analysis process leading from unimodal to workflow models.

[b] Because variable extraction for the condensed Clin-NC model was performed in the PRONIA plus 18M sample, we report its performance and the respective metrics of its dependent stacked model only for the PRONIA minus 18M, the ZInEP, and BEARS-Kid samples.

[c] Γ = 0.0 (case propagation cutoffs: 25.0% and 100%). Sequence is Clin-NC, Clin-NC plus raters, Clin-NC plus PRS, and Clin-NC plus sMRI.

[d] Γ = 0.5 (case propagation cutoffs: 37.5% and 100%). Sequence is Clin-NC, Clin-NC plus PRS, Clin-NC plus raters, and Clin-NC plus sMRI.

[e] Γ = 1.0 (case propagation cutoffs: 37.5% and 75.0%). Sequence is Clin-NC, Clin-NC plus PRS, Clin-NC plus raters, and Clin-NC plus sMRI.

0.67 [0.88] vs 0.26 [0.61] years) and more prevalent attenuated positive symptoms (APSs) at baseline (18 of 26 [69.2%] vs 88 of 308 [28.6%]) (Table 1). Of 167 patients with CHR, 23 (13.8%) developed psychosis, whereas 53 (31.7%) had remitted from CHR criteria at the 9-month visit. Major depression affected 103 patients with CHR syndromes (61.7%) but was not differentially associated with psychosis transition (Table 1). Nonremitting mood and anxiety disorders were present during follow-up in cases with psychosis transition and nontransition (eTable 16 in the Supplement). Compared with healthy controls, ROD was associated with low but significantly elevated CHR symptom scores (eg, mean [SD] Structured Interview for Psychosis–Risk Syndromes positive symptoms, 0.43 [0.44] vs 0.10 [0.21]; $P < .001$; mean [SD] Schizophrenia Proneness Instrument: Cognitive Disturbances symptoms, 0.24 [0.32] vs 0.02 [0.08]; $P < .001$) (eTable 3 in the Supplement). Functional-cognitive and interpersonal abnormalities were comparable between ROD and CHR groups. Of 167 patients with ROD, 32 (19.2%) developed psychosis-related outcomes, including CHR states in 29 (17.4%) and psychosis transitions in 3 (1.8%).

Machine Learning Analyses
The full clinical-neurocognitive model predicted psychosis transition with a balanced accuracy (BAC-LOSOCV) of 75.7% (sensitivity, 84.6%; specificity, 66.8%; $P < .001$) (Table 2). Significant predictors as determined by sign-based consistency mapping ($z > 3.28$; $P < .05$ for FDR) were APS and motor disturbances, a nonsupportive family environment during child-

hood, and reduced facial emotion recognition (Figure 1B). Compared with healthy controls, those assigned to psychosis transition had elevated abnormality scores in these variables, whereas those with nontransition assignments showed an abnormality pattern focused on unusual thought content, suspiciousness, perceptual abnormalities, and childhood adversity, with higher visual working memory and semantic verbal fluency performance ($P < .05$ for FDR) (eFigure 7 in the Supplement).

The PRS-based model achieved a BAC-LOSOCV of 66.1% (sensitivity, 76.0%; specificity, 56.2%; $P < .001$). Among the 10 tested genome-wide significance thresholds, only $P = 1.0$ reached significance (Figure 2). Compared with healthy controls, those with psychosis transition assignments had elevated PRS across all whole-genome P thresholds, whereas those with nontransition assignments expressed reduced PRS at $P \geq 5.7 \times 10^{-4}$ (eFigure 8A in the Supplement). Patients with observed nontransition and those with ROD did not show reduced PRS (eFigure 8B-C in the Supplement).

The sMRI-based model attained a BAC-LOSOCV of 70.7% (sensitivity, 88.0%; specificity, 53.5%; $P < .001$). At a stability threshold (cross-validation ratio) of at least 3, the brain pattern predicting psychosis transition involved reduced gray matter volume in the superior temporal, supramarginal, angular, orbitofrontal, inferior frontal, dorsomedial prefrontal, and occipital cortices. The predictive pattern also included areas of increased gray matter volume covering the dorsolateral prefrontal, precuneal, insular, hippocampal, and cerebellar brain regions (eFigure 5 in the Supplement). This brain signature

Figure 1. Predictive Signatures Underlying the Clinical-Neurocognitive Models
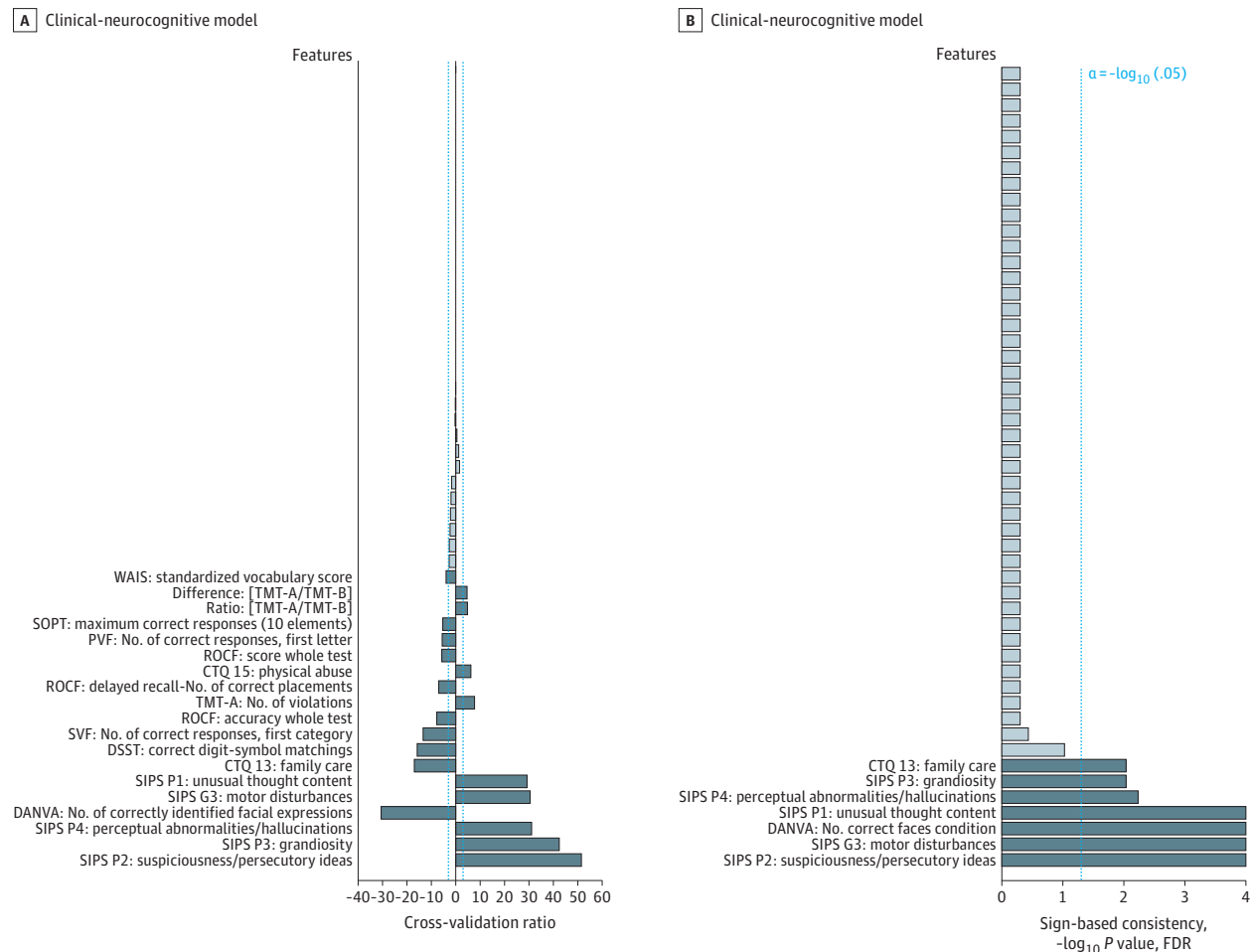
The reliability of predictive pattern elements was evaluated using cross-validation ratio mapping (A). In addition, the significance of predictive features used by the clinical-neurocognitive model was assessed by means of sign-based consistency mapping (B). Both visualization methods are detailed in the eMethods in the Supplement. CTQ indicates Childhood Trauma Questionnaire; DANVA, Diagnostic Analysis of Non-Verbal Accuracy;

DSST, Digit-Symbol Substitution Test; FDR, false discovery rate; PVF, Phonetic Verbal Fluency; ROCF, Rey-Osterreith Figure; SIPS, Structured Interview for Psychosis-Risk Syndromes; SOPT, Self-Ordered Pointing Task; SVF, Semantic Verbal Fluency; TMT, Trail Making Test; and WAIS, Wechsler Adult Intelligence Scale.

differentiated psychosis transition-assigned patients from healthy controls, whereas nontransition-assigned patients showed a partial pattern inversion with increased temporo-occipital gray matter volume compared with healthy controls (see threshold-free cluster enhancement statistics thresholded at $P < .05$ for FDR) (eFigure 9 in the Supplement).

Clinical raters achieved a BAC of 73.2% (sensitivity, 61.5%; specificity, 84.9%), which was independent of the length of their early recognition experience (mean [SD] length for correct predictions: 31.8 [46.8] months; mean [SD] length for wrong predictions, 29.4 [38.1] months; unpaired 2-tailed $t_{326} = 0.35$; $P = .72$). The stacked model combining unimodal algorithms produced a BAC-LOSOCV of 82.9% (sensitivity, 80.8%; specificity, 85.0%). Integration of raters' prognoses into the stacked model increased BAC-LOSOCV to 85.5% (sensitivity, 84.6%; specificity, 86.4%), and they were the cybernetic model's most relevant predictor (Figure 2C).

Baseline Moderators of Prediction Performance

Image quality (eFigure 10 in the Supplement), baseline treatments or previous hospitalizations (eTable 7 in the Supplement), follow-up duration and frequency (eTable 8 and eFigure 11 in the Supplement), and site effects (eTable 10 in the Supplement) did not influence model performance. Study group (CHR vs ROD) could be classified with a BAC-LOSOCV of 82.3% (sensitivity, 73.7%; specificity, 91.0%) using clinical-neurocognitive data and with a BAC of 55.7% (sensitivity, 49.7%; specificity, 61.7%) using PRS (eTable 12 in the Supplement). These diagnostic classifiers explained 49.7% and 18.3%, respectively, of the variance of the respective prognostic counterparts (both $P < .001$ for FDR) (eTable 13 in the Supplement). Raters' prognoses were also significantly informed by baseline study group (BAC, 62.7%; sensitivity, 30.9%; specificity, 94.5%) (eTable 12 in the Supplement). The removal of the patients with ROD from the training samples or their sub-

stitution with healthy controls significantly reduced the balanced accuracy of all risk calculators by −2.8% to −11.7% in the CHR group (eTable 14 and eFigure 12 in the Supplement).

Prognostic Generalization

Prognostic assignments also delineated psychosis transition and P-CHR and NP-CHR courses irrespective of model type (eTable 15 in the Supplement). The separability of P-CHR from NP-CHR courses was lower and only significant for clinical-neurocognitive models and raters. The nonnegative matrix factorization and linear mixed-model analysis showed that CHR syndrome trajectories were stratified by the predictions of the clinical-neurocognitive classifier (factor F1 paranoid-perceptual disturbances, $F_{1,832} = 136.35$ [$P < .001$ for FDR]; factor F2 disturbances of volition and affect, $F_{1,832} = 12.76$ [$P = .001$ for FDR]; factor F3 functional disturbances, $F_{1,832} = 24.34$ [$P < .001$ for FDR]; factor F4 cognitive disturbances, $F_{1,832} = 44.05$ [$P = .007$ for FDR]) as well as by raters' outcome estimates (factor F1 paranoid-perceptual disturbances, $F_{1,825} = 30.64$ [$P < .001$ for FDR]; factor F2 disturbances of volition and affect, $F_{1,825} = 6.15$ [$P = .03$ for FDR]; factor F3 functional disturbances, $F_{1,825} = 5.80$ [$P < .03$ for FDR]; factor F4 cognitive disturbances, $F_{1,825} = 9.00$ [$P = .007$ for FDR]) (eFigure 13 in the Supplement). Nonpsychotic disease courses were not associated with prognostic assignments (eTable 16 in the Supplement).

Clinical Scalability and External Validation of Predictive Models

We identified a prognostic sequence that produced a BAC-LOSOCV of 85.9% (sensitivity, 84.6%; specificity, 87.3%) (Table 2) and started with the clinical-neurocognitive model, added raters, and finally integrated PRS- and sMRI-based models (eFigure 14 in the Supplement). Across this sequence, the positive likelihood ratio increased to 6.6, whereas the population requiring all prognostic assessments decreased to 41.1%. Regularization for diagnostic parsimony further reduced this population to 23.2% (regularization strength Γ = 0.5) and 0 (regularization strength Γ = 1.0) (PRONIA plus 18M sample, left panel of eFigure 15B in the Supplement), with the latter parsimony level significantly reducing BAC by −6.4% (PRONIA plus 18M sample, left panel of eFigure 15C, part 2 in the Supplement). Highly similar findings were obtained when analyzing the complete PRONIA cohort (right panels in eFigure 15B and C in the Supplement).

Further clinical scalability experiments showed that the condensed clinical-neurocognitive model matched the full model in correctly predicting PRONIA minus 18M cases with nontransition (specificity, 65.9%). Furthermore, its performance in the external ZInEP-CHR cohort (BAC, 65.3%; sensitivity, 87.5%; specificity, 43.1%) and the Bi-national Evaluation of At-Risk Symptoms in Children and Adolescents sample (BAC, 70.4%; sensitivity, 76.9%; specificity, 63.9%) was similar to the full model in the PRONIA-CHR sample (BAC, 63.3%; sensitivity, 87.0%; specificity, 39.6%) (eTable 14 in the Supplement) and the complete PRONIA cohort (BAC, 72.8%; sensitivity, 88.0%; specificity, 57.5%) (Table 2). Replacing the full model with its condensed counterpart in the nonregularized

Figure 2. Predictive Signatures Underlying the Polygenic Risk Score (PRS)–Based and Cybernetic Risk Calculator Models
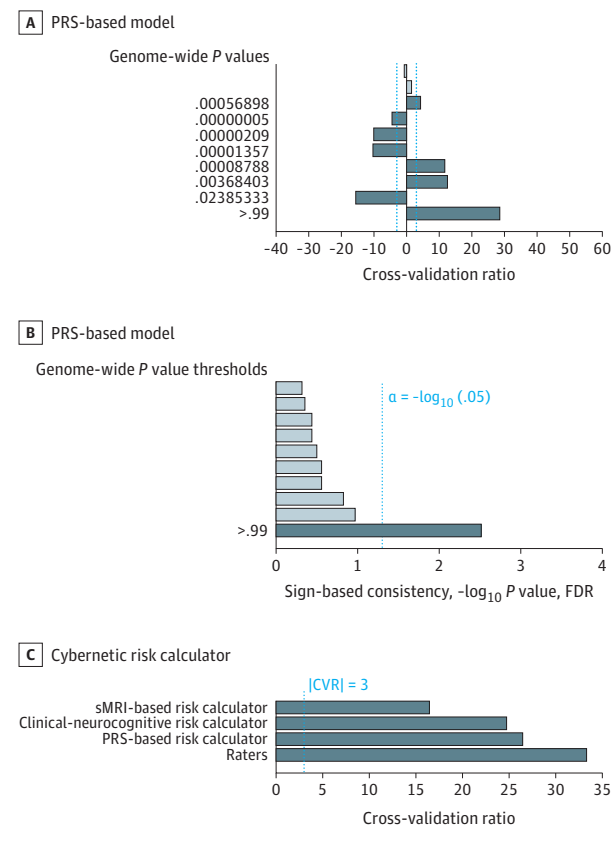
The reliability of predictive pattern elements was evaluated using cross-validation ratio (CVR) mapping (A). In addition, the significance of predictive features used by the PRS-based model was assessed by means of sign-based consistency mapping (B) (as described in the eMethods in the Supplement). The cybernetic model combines all algorithmic and human components (C). FDR indicates false discovery rate; and sMRI, structural magnetic resonance imaging.

or regularized (Γ = 0.5) workflows did not increase the false-positive rate in the PRONIA minus 18M sample or the complete PRONIA cohort (eFigure 14C in the Supplement).

The validation of the unimodal sMRI-based workflow component in the ZInEP (BAC, 67.9%; sensitivity, 75.0%; specificity, 60.8%) and Früherkennung von Psychosen (BAC, 68.5%; sensitivity, 75.0%; specificity, 61.9%) samples approximated the PRONIA-CHR results (BAC, 70.8%; sensitivity, 86.4%; specificity, 55.3%) (eTable 14 in the Supplement). Finally, the stacked risk calculator composed of the condensed clinical-neurocognitive and sMRI-based models significantly outperformed these models in the ZInEP data (BAC, 71.3%; sensitivity, 75.0%; specificity, 67.7%) (Figure 3 and Table 2).

Comparisons of Risk Calculators and Clinical Raters

All risk calculators were significant in the permutation analysis (mean [SD] BAC, 77.3 [4.8]; mean [SD] sensitivity, 80.2% [6.5%]; mean [SD] specificity, 74.5% [13.8%]; $P < .001$ for FDR for all models) Table 2), but differences in BAC emerged

Figure 3. Statistical Comparison of Prognostic Models
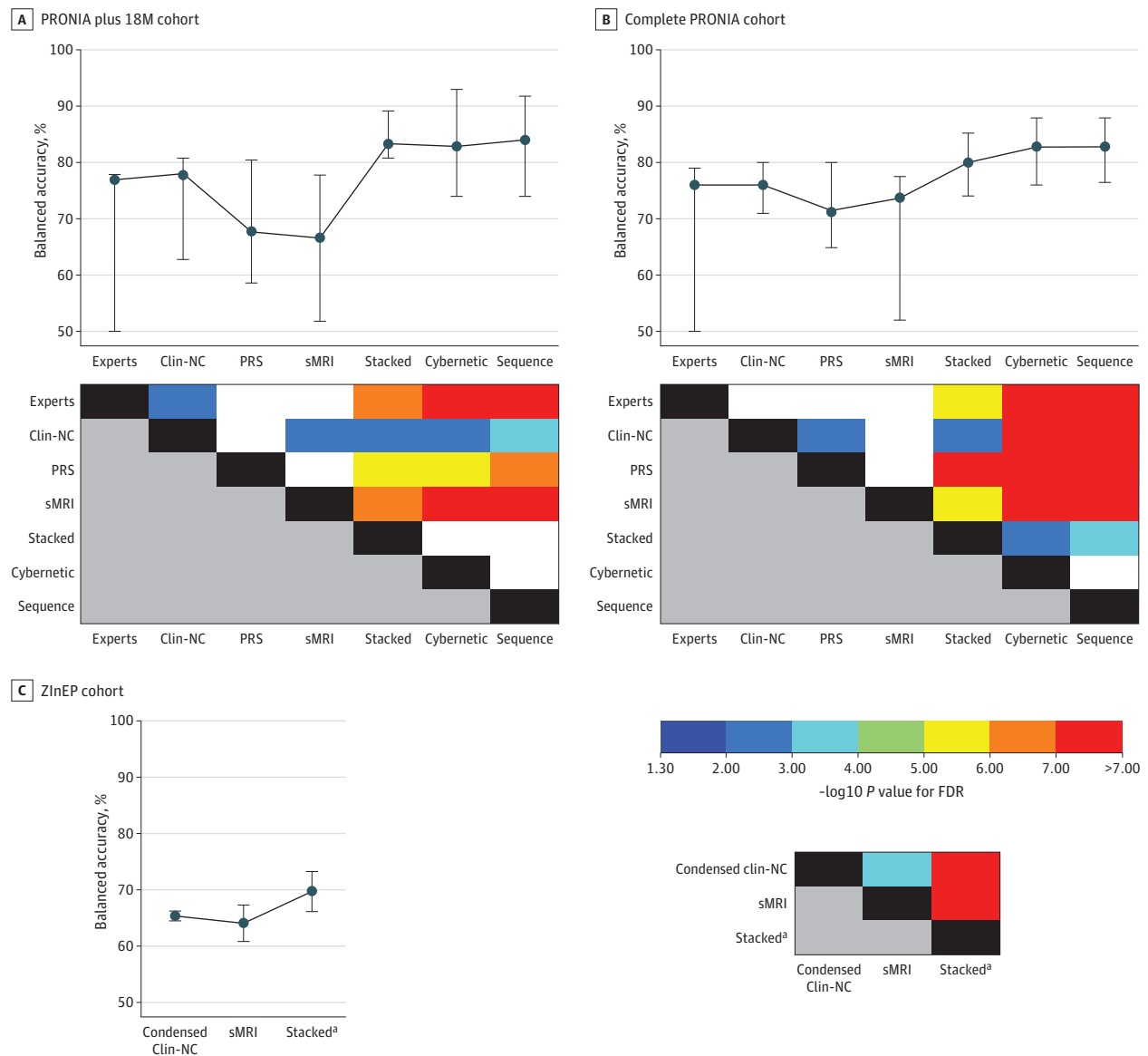
Cohorts include patients with follow-up of 18 months or longer (PRONIA plus 18M), the complete PRONIA cohort, and the Zurich Early Recognition Program (ZInEP). Data points indicate median. The Quade test[51] was used to compare the models' median balanced accuracy (BAC) computed across the cross-validation cycle (CV_2) test data partitions. The BAC measures obtained for the ZInEP cohort (C) were produced by applying the condensed clinical-neurocognitive (Clin-NC), structural magnetic resonance imaging (sMRI)–based, and respective stacked risk calculators of the complete PRONIA sample (B) to this external sample (eFigure 2 in the Supplement). Post hoc comparisons were performed using the t distribution approximation described by Heckert and Filliben.[52] P values were corrected for multiple comparisons using the false discovery rate (FDR). The upper graphs represent the median BAC for each risk calculator in analyses A, B, and C along with the lower and

upper quartiles of the BAC distributions (whiskers of the error bars). The lower figures show the logarithmized, FDR-corrected P matrix for the pairwise post hoc classifier comparisons. For an in-depth analysis of the prognostic sequence included in the risk classifier comparison, see eFigures 14 and 15 in the Supplement. The cybernetic risk calculator analyzed the combined predictions of raters, Clin-NC, polygenic risk score (PRS)–based, and sMRI-based risk calculators; the stacked risk calculator, the combined predictions of Clin-NC, PRS-based, and sMRI-based risk calculators.

[a] Indicates risk calculator encompassing the condensed Clin-NCs and sMRI-based models and specifically trained to externally validate the effect of stacking on prognostic performance in the ZInEP cohort.

(Figure 3). Multimodal risk calculators outperformed all unimodal counterparts (t range, 3.14-6.20; $P < 3.19 \times 10^{-8}$ to $P = .002$ for FDR), whereas the nonregularized prognostic sequence did not differ from the cybernetic model (mean [SD] BAC for nonregularized sequence, 83.7% [9.6%]; mean [SD]

BAC for cybernetic model, 83.4% [9.6%]; $t = 0.21$; $P = .44$ for FDR). In addition, the stacked model (mean [SD] BAC, 79.7% [7.9%]) was outperformed by both the cybernetic model (mean [SD] BAC, 81.5% [9.6%]; $t = 3.18$; $P = .001$ for FDR) and the nonregularized sequential model (mean [SD] BAC, 82.0% [9.6%];

t = 3.82; P < .001 for FDR) in the complete PRONIA sample. Raters were comparable to unimodal risk calculators (mean [SD] BAC for raters, 68.2% [12.6%]; mean [SD] BAC range for unimodal predictors, 67.5% [16.4%] to 74.8% [6.8%]; t range, 0.002-1.56; P = .50 to P = .07 for FDR) but were outperformed by multimodal prediction algorithms in terms of higher BAC and reduced prognostic variability (stacked model vs raters, t = 4.64 [P < .001 for FDR]; cybernetic model vs raters, t = 7.82 [P < .001 for FDR]; sequence model vs raters, t = 8.46 [P < .001 for FDR]). Finally, the nonregularized sequential model reduced raters' false-negative rate from 38.5% to 15.4% (PRONIA plus 18M sample) or 19.2% (complete PRONIA cohort) of cases.

Discussion

Using a thorough model discovery and validation approach,[21] our study demonstrated geographic transportability of expert-based clinical and biological psychosis transition prediction approaches across a transdiagnostic, multinational risk population. We found that combined risk calculators outperformed all unimodal counterparts and clinical raters in terms of prognostic accuracy and cross-site stability. Importantly, our study revealed that the increased diagnostic burden arising from data fusion could be mitigated through optimized sequential testing that arranges clinicians and risk calculators into clinically scalable prognostic workflows. Based on this form of deferral learning,[55] we showed that the complete assessment battery is only needed in 23.2% of the initial population (eFigure 15B in the Supplement). This subgroup was enriched for patients who received a prediction of psychosis transition in the initial clinical-neurocognitive examination, suggesting that biological markers of psychosis transition are useful for delineating true-positive from false-positive findings at the later steps of a multistep prognostic assessment.

Examining the baseline heterogeneity of our transdiagnostic population, we found functional-neurocognitive impairments in the ROD group akin to the CHR group and low between-group neuroanatomical and genetic separability (eTable 12 in the Supplement), supporting the neurobiological proximity between early-onset affective and psychotic disorders.[26,30,56] Although CHR syndromes expectedly separated CHR and ROD groups at the cross-sectional level, we observed that these syndromes emerged in 19.2% of patients with ROD during the follow-up period, which led to psychosis transition in 1.8% of cases.[56] Strikingly, our analyses also showed that the models' prognostic accuracy, particularly the sensitivity for psychosis transition in patients with CHR, depended on patients with ROD being part of the model discovery process, which further supports the pooling of both groups into a broader risk population (eTable 14 and eFigure 12 in the Supplement). Our finding of a transdiagnostic predictability of psychosis was corroborated by the generalizability of the clinical-neurocognitive and neuroanatomical models to external samples, which showed markedly different risk levels, age distributions, and diagnostic compositions.

The in-depth analysis of the clinical-neurocognitive domain revealed that the presence of APSs facilitated a good baseline separability of CHR syndromes vs ROD and substantially informed the prediction of psychosis transition (eTable 13 in the Supplement). However, measures of childhood adversity,[46] motor disturbances,[57] and facial affect recognition[58] did not overlap between diagnostic and prognostic models (eFigure 16 in the Supplement) and thus could be regarded as transdiagnostic markers[59-62] of poor psychosis-related outcomes, including transition to psychosis. This interpretation was supported by the prognostic generalization of the clinical-neurocognitive model to the clinically relevant separation of patients with (1) nonremitting/de novo and nonsymptomatic CHR syndrome courses (eTable 15 in the Supplement) or (2) unfavorable perceptual, affective, functional, and basic symptom trajectories (eFigure 13 in the Supplement).[63,64] Importantly, the model's prognostic generalization capacity did not encompass nonpsychotic diagnoses (eTable 16 in the Supplement), and, thus, its prognostic pluripotency was confined to diverse CHR-specific symptom courses.[65]

Furthermore, we confirmed the prognostic value of PRS for schizophenia,[48] as reported recently,[16] and extended those findings by showing that genetic information augments the performance of clinical-neurocognitive models and prognostic workflows in a broader risk population (Table 2 and eTable 20 in the Supplement). Within this transdiagnostic setting, we replicated group-level differences among patients with psychosis transition, patients with nontransition, and healthy controls[16] but also found that PRS-based prognostic assignments specifically differentiated APS-related trajectories (eFigure 13 in the Supplement).[66] They also delineated patient groups with abnormally high and low genetic risk compared with healthy controls (eFigure 8 in the Supplement)—a finding that may point to distinct environmental and/or neurobiological pathways conferring risk and resilience to psychosis.[67]

The analysis of the structural neuroimaging data revealed a psychosis-predictive brain signature that generalized well across 3 independent cohorts. This signature overlapped with brain alterations previously reported to correlate with perceptual abnormalities, disorganization of speech and thought, and poor insight in early, subsyndromal, or prodromal stages of psychosis.[9,68-70] Interestingly, nontransition-assigned patients showed reversed temporo-occipital volume reductions, which differentiated them from healthy controls (eFigure 9 in the Supplement). These findings may point to ongoing compensatory mechanisms of resilience to psychosis, as reported previously in a longitudinal sMRI study of adolescents with CHR.[71] In this regard, our sMRI-based risk calculator may serve as a useful tool for enriching future observational studies and clinical trials with at-risk patients who express potential brain mechanisms of resilience to psychosis transition.

We observed that our raters matched unimodal risk calculators in predicting psychosis as measured by their BAC. However, raters also showed a pronounced optimism bias (low sensitivity and high specificity) toward the true risk of poor clinical outcome.[13] It is noteworthy that their prognoses were based on all information collected in an extended study-

related assessment and likely would be less accurate in routine, time-restricted diagnostic settings. Because the algorithmic counterparts showed exactly the inverted bias (high sensitivity and low specificity), the integration of clinicians and risk calculators into the cybernetic model produced a superior predictive system.[50] Furthermore, our prognostic workflows demonstrated that similar levels of prognostic accuracy can be achieved by reducing the false-positive rate through sequential model application in patients with an estimated higher risk for psychosis transition (eFigure 14 in the Supplement). In this subgroup, the removal of the final sMRI-based assessment step increased false-positive findings (eFigure 15 in the Supplement), suggesting that the cost-benefit ratio of expensive neuromarkers needs to be individually adjusted according to the patient's predicted risk.[72]

The finding that prognostic workflows always started with the clinical-neurocognitive model places the recognition of the clinical gestalt of emerging psychosis at the gateway of more precise early detection techniques.[73] Our scalability experiments suggest that the laborious recognition of this pattern currently practiced in early recognition services could be effectively condensed to a few clinical-neurocognitive variables,[6] thus enhancing the clinical utility of the proposed workflow. Nonetheless, future studies should revisit the validity of the selected 7 variables because they have been taken out of their original assessment context. Further studies also need to quantitatively explore the information patterns guiding clinicians' gut-feeling estimates of psychosis transition and in turn foster more effective clinical early recognition strategies that integrate with cybernetic systems.

Limitations

Psychosis transitions were limited to 26 individuals in the PRONIA discovery sample. This sample size increased the risk of producing overly optimistic prediction results owing to an accidental collection of well-classifiable cases. We imple-

mented a multistep model validation procedure to guard against this possibility, including label permutation testing, strict nested cross-validation of all processing steps,[74] in-depth model analysis to assess possible prognostic confounds and moderators, specificity testing of all models in a completely held-back portion of the PRONIA sample, and model benchmarking in 3 independent data sets, which provided a further 45 individuals with psychosis transition and 600 with nontransition for external validation. Owing to limited data availability in these samples, only the condensed clinical-neurocognitive, sMRI-based, and a specific stacked risk calculator trained on the outputs of the former 2 models could be externally validated. However, our internal-external validation approach followed established guidelines for model construction and validation.[21] In keeping with this literature, the similar performance levels observed in our LOSOCV and independent validation experiments support the validity of the models not tested in external samples.

Conclusions

In this prognostic study, we identified generalizable risk assessment tools that can be arranged into a multimodal prognostic workflow for a clinically viable, individualized prediction of psychosis in patients with CHR states and ROD. Our study showed for the first time, to our knowledge, that the augmentation of human prognostic abilities with algorithmic pattern recognition improves prognostic accuracy to margins that likely justify the clinical implementation of cybernetic decision-support tools. New international collaborations, such as the HARMONY (Harmonization of At Risk Multisite Observational Networks for Youth) initiative,[75] may help to propel a reciprocal and iterative process of clinical validation and refinement of these prognostic tools in real-world early recognition services.

Author Affiliations: Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University, Munich, Germany (Koutsouleris, Dwyer, Urquijo-Castro, Sanfelici, Popovic, Oeztuerk, Weiske, Ruef, Falkai); Max-Planck Institute of Psychiatry, Munich, Germany (Koutsouleris, Falkai); Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom (Koutsouleris); Institute of Human Genetics, Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany (Degenhardt, Noethen); Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany (Degenhardt); Institute of Genomic Statistics and Bioinformatics, University of Bonn, Bonn, Germany

(Maj, Borisov, Krawitz); Max-Planck School of Cognition, Leipzig, Germany (Sanfelici); International Max-Planck Research School for Translational Psychiatry, Munich, Germany (Popovic, Oezturk); Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York (Haas); Department of Psychiatry and Psychotherapy, Faculty of Medicine and University Hospital, University of Cologne, Cologne, Germany (Kambeitz-Ilankovic, Ruhrmann, Penzel, Kambeitz, Haidl, Rosen); Department of Education, Psychology, and Communication, University of Bari Aldo Moro, Bari, Italy (Antonucci); Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany (Neufang, Schmidt-Kraepelin, Schultze-Lutter, Meisenzahl); Institute for Mental Health, University of Birmingham, Birmingham, United Kingdom (Chisholm, Upthegrove); Department of Psychiatry, Psychiatric University Hospital, University of Basel, Switzerland (Riecher-Rössler, Egloff, Schmidt, Andreou, Borgwardt); Department of Psychiatry, University of Turku, Turku, Finland (Hietala, Salokangas); GE Healthcare GmbH (previously GE Global Research GmbH), Munich, Germany (Schirmer); Department of Child and Adolescent

Psychiatry, University of Münster, Münster, Germany (Romer); Department of Child and Adolescent Psychiatry, Psychotherapy and Psychosomatics, LVR Clinic Düsseldorf, Düsseldorf, Germany (Walger); Department of Child and Adolescent Psychiatry and Psychotherapy, University of Zürich, Zürich, Switzerland (Franscini, Traber-Walker); University Hospital of Child and Adolescent Psychiatry, University Hospital Hamburg-Eppendorf, Hamburg, Germany (Schimmelmann); University Hospital of Child and Adolescent Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland (Schimmelmann, Flückiger, Michel); Department of Psychiatry, Psychotherapy and Psychosomatics, University Hospital of Psychiatry Zurich, Zurich, Switzerland (Rössler, Heekeren, Buechler, Theodoridou); Department of Psychiatry and Psychotherapy I, LVR Hospital Cologne, Cologne, Germany (Heekeren); Department of Neuroradiology, University Hospital of Zurich, Zurich, Switzerland (Buechler); Melbourne Neuropsychiatry Centre, University of Melbourne and Melbourne Health, Melbourne, Australia (Pantelis); Department of Psychiatry and Psychotherapy, University of Münster, Münster,

Germany (Lencer); Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany (Lencer, Borgwardt); Department of Basic Medical Science, Neuroscience and Sense Organs, University of Bari Aldo Moro, Bari, Italy (Bertolino); Department of Neurosciences and Mental Health, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy (Brambilla); Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy (Brambilla); Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia (Wood); Orygen, the National Centre of Excellence for Youth Mental Health, Melbourne, Australia (Wood); Department of Psychology and Mental Health, Faculty of Psychology, Airlangga University, Surabaya, Indonesia (Schultze-Lutter).

Balestrieri, Carolina Bonivento, Giuseppe Cabras, and Franco Fabbro (Department of Medical Area, Workgroup of Paolo Brambilla, University of Udine); and Marco Garzitto and Sara Piccin (IRCCS Scientific Institute E. Medea, Polo FVG, Workgroup of Paolo Brambilla, University of Udine).

REFERENCES

1. McGorry PD, Mei C. Ultra-high-risk paradigm: lessons learnt and new directions. *Evid Based Ment Health*. 2018;21(4):131-133. doi:10.1136/ebmental-2018-300061

2. Fusar-Poli P, Salazar de Pablo G, Correll CU, et al. Prevention of psychosis: advances in detection, prognosis, and intervention. *JAMA Psychiatry*. 2020;77(7):755-765. doi:10.1001/jamapsychiatry.2019.4779

3. Fusar-Poli P, Rutigliano G, Stahl D, et al. Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry*. 2017;74(5):493-500. doi:10.1001/jamapsychiatry.2017.0284

4. Schultze-Lutter F, Michel C, Schmidt SJ, et al. EPA guidance on the early detection of clinical high risk states of psychoses. *Eur Psychiatry*. 2015;30(3):405-416. doi:10.1016/j.eurpsy.2015.01.010

5. 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. doi:10.1001/jamapsychiatry.2013.1909

6. 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. doi:10.1176/appi.ajp.2016.15070890

7. Zhang T, Xu L, Tang Y, et al; SHARP (ShangHai At Risk for Psychosis) Study Group. Prediction of psychosis in prodrome: development and validation of a simple, personalized risk calculator. *Psychol Med*. 2019;49(12):1990-1998. doi:10.1017/S0033291718002738

8. Koutsouleris N, Meisenzahl EM, Davatzikos C, et al. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Arch Gen Psychiatry*. 2009;66(7):700-712. doi:10.1001/archgenpsychiatry.2009.62

9. Koutsouleris N, Riecher-Rössler A, Meisenzahl EM, et al. Detecting the psychosis prodrome across high-risk populations using neuroanatomical biomarkers. *Schizophr Bull*. 2015;41(2):471-482. doi:10.1093/schbul/sbu078

10. Das T, Borgwardt S, Hauke DJ, et al. Disorganized gyrification network properties during the transition to psychosis. *JAMA Psychiatry*. 2018;75(6):613-622. doi:10.1001/jamapsychiatry.2018.0391

11. 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. doi:10.1016/j.schres.2016.11.047

12. Clark SR, Baune BT, Schubert KO, et al. Prediction of transition from ultra-high risk to first-episode psychosis using a probabilistic model combining history, clinical assessment and fatty-acid biomarkers. *Transl Psychiatry*. 2016;6(9):e897. doi:10.1038/tp.2016.170

13. 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. doi:10.1001/jamapsychiatry.2018.2165

14. Fusar-Poli P, Werbeloff N, Rutigliano G, et al. Transdiagnostic risk calculator for the automatic detection of individuals at risk and the prediction of psychosis: second replication in an independent national health service trust. *Schizophr Bull*. 2019;45(3):562-570. doi:10.1093/schbul/sby070

15. Carrión RE, Cornblatt BA, Burton CZ, et al. Personalized prediction of psychosis: external validation of the NAPLS-2 psychosis risk calculator with the EDIPPP project. *Am J Psychiatry*. 2016;173(10):989-996. doi:10.1176/appi.ajp.2016.15121565

16. Perkins DO, Olde Loohuis L, Barbee J, et al. Polygenic risk score contribution to psychosis prediction in a target population of persons at clinical high risk. *Am J Psychiatry*. 2020;177(2):155-163. doi:10.1176/appi.ajp.2019.18060721

17. Worthington MA, Cao H, Cannon TD. Discovery and validation of prediction algorithms for psychosis in youth at clinical high risk. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2020;5(8):738-747. doi:10.1016/j.bpsc.2019.10.006

18. Studerus E, Ramyead A, Riecher-Rössler A. Prediction of transition to psychosis in patients with a clinical high risk for psychosis: a systematic review of methodology and reporting. *Psychol Med*. 2017;47(7):1163-1178. doi:10.1017/S0033291716003494

19. Sanfelici R, Dwyer DB, Antonucci LA, Koutsouleris N. Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: a meta-analytic view on the state-of-the-art. *Biol Psychiatry*. 2020;88(4):349-360. doi:10.1016/j.biopsych.2020.02.009

20. Yarkoni T, Westfall J. Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect Psychol Sci*. 2017;12(6):1100-1122. doi:10.1177/1745691617693393

21. Steyerberg EW, Harrell FE Jr. Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*. 2016;69:245-247. doi:10.1016/j.jclinepi.2015.04.005

22. Fusar-Poli P, Hijazi Z, Stahl D, Steyerberg EW. The science of prognosis in psychiatry: a review. *JAMA Psychiatry*. 2018;75(12):1289-1297. doi:10.1001/jamapsychiatry.2018.2530

23. Lee TY, Lee J, Kim M, Choe E, Kwon JS. Can we predict psychosis outside the clinical high-risk state? a systematic review of non-psychotic risk syndromes for mental disorders. *Schizophr Bull*. 2018;44(2):276-285. doi:10.1093/schbul/sbx173

24. McGorry PD, Hartmann JA, Spooner R, Nelson B. Beyond the "at risk mental state" concept: transitioning to transdiagnostic psychiatry. *World Psychiatry*. 2018;17(2):133-142. doi:10.1002/wps.20514

25. Fusar-Poli P, Nelson B, Valmaggia L, Yung AR, McGuire PK. Comorbid depressive and anxiety disorders in 509 individuals with an at-risk mental state: impact on psychopathology and transition to psychosis. *Schizophr Bull*. 2014;40(1):120-131. doi:10.1093/schbul/sbs136

26. Koutsouleris N, Meisenzahl EM, Borgwardt S, et al. Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain*. 2015;138(pt 7):2059-2073. doi:10.1093/brain/awv111

27. Häfner H, Maurer K, Trendler G, an der Heiden W, Schmidt M, Könnecke R. Schizophrenia and depression: challenging the paradigm of two separate diseases—a controlled study of schizophrenia, depression and healthy controls. *Schizophr Res*. 2005;77(1):11-24. doi:10.1016/j.schres.2005.01.004

28. Wigman JTW, van Nierop M, Vollebergh WA, et al. Evidence that psychotic symptoms are prevalent in disorders of anxiety and depression, impacting on illness onset, risk, and severity—implications for diagnosis and ultra-high risk research. *Schizophr Bull*. 2012;38(2):247-257. doi:10.1093/schbul/sbr196

29. Siebald C, Khandaker GM, Zammit S, Lewis G, Jones PB. Association between childhood psychiatric disorders and psychotic experiences in adolescence: a population-based longitudinal study. *Compr Psychiatry*. 2016;69:45-52. doi:10.1016/j.comppsych.2016.05.004

30. Musliner KL, Mortensen PB, McGrath JJ, et al; Bipolar Disorder Working Group of the Psychiatric Genomics Consortium. Association of polygenic liabilities for major depression, bipolar disorder, and schizophrenia with risk for depression in the Danish population. *JAMA Psychiatry*. 2019;76(5):516-525. doi:10.1001/jamapsychiatry.2018.4166

31. Schmidt A, Cappucciati M, Radua J, et al. Improving prognostic accuracy in subjects at clinical high risk for psychosis: systematic review of predictive models and meta-analytical sequential testing simulation. *Schizophr Bull*. 2017;43(2):375-388. doi:10.1093/schbul/sbw098

32. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-1358. doi:10.1056/NEJMra1814259

33. Wiener N. *The Human Use of Human Beings*. Houghton Mifflin; 1950.

34. Clark SR, Schubert KO, Baune BT. Towards indicated prevention of psychosis: using probabilistic assessments of transition risk in psychosis prodrome. *J Neural Transm (Vienna)*. 2015;122(1):155-169. doi:10.1007/s00702-014-1325-9

35. Riecher-Rössler A, Gschwandtner U, Aston J, et al. The Basel early-detection-of-psychosis (FEPSY)-study—design and preliminary results. *Acta Psychiatr Scand*. 2007;115(2):114-125. doi:10.1111/j.1600-0447.2006.00854.x

36. Theodoridou A, Heekeren K, Dvorsky D, et al. Early recognition of high risk of bipolar disorder and psychosis: an overview of the ZInEP "early recognition" study. *Front Public Health*. 2014;2:166. doi:10.3389/fpubh.2014.00166

37. Schimmelmann BG, Michel C, Martz-Irngartinger A, Linder C, Schultze-Lutter F. Age matters in the prevalence and clinical significance of ultra-high-risk for psychosis symptoms and criteria in the general population: findings from the BEAR and BEARS-Kid studies. *World Psychiatry*. 2015;14(2):189-197. doi:10.1002/wps.20216

38. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg*. 2015;102(3):148-158. doi:10.1002/bjs.9736

39. McGlashan T, Walsh B, Woods S. *The Psychosis-Risk Syndrome: Handbook for Diagnosis and Follow-up.* Oxford University Press; 2010.

40. Yung AR, Nelson B, Thompson A, Wood SJ. The psychosis threshold in ultra high risk (prodromal) research: is it valid? *Schizophr Res*. 2010;120(1-3): 1-6. doi:10.1016/j.schres.2010.03.014

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

42. Schultze-Lutter F, Addington J, Ruhrmann S, Klosterkötter J. *Schizophrenia Proneness Instrument, Adult Version*. Giovanni Fioriti Editore; 2007.

43. 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. doi:10.1093/schbul/sbm029

44. Pedersen G, Hagtvet KA, Karterud S. Generalizability studies of the Global Assessment of Functioning–Split version. *Compr Psychiatry*. 2007; 48(1):88-94. doi:10.1016/j.comppsych.2006.03.008

45. Bernstein D, Fink L. *CTQ: Childhood Trauma Questionnaire: A Retrospective Self-report*. Psychological Corp; 1998.

46. Redman SL, Corcoran CM, Kimhy D, Malaspina D. Effects of early trauma on psychosis development in clinical high-risk individuals and stability of trauma assessment across studies: a review. *Arch Psychol (Chic)*. 2017;1(3):28.

47. Koutsouleris N, Davatzikos C, Bottlender R, et al. Early recognition and disease prediction in the at-risk mental states for psychosis using neurocognitive pattern classification. *Schizophr Bull*. 2012;38(6):1200-1215. doi:10.1093/schbul/sbr037

48. Ripke S, Neale BM, Corin A, et al; Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature*. 2014;511(7510):421-427. doi:10.1038/nature13595

49. Karrer TM, Bassett DS, Derntl B, et al. Brain-based ranking of cognitive domains to predict schizophrenia. *Hum Brain Mapp*. 2019;40(15): 4487-4507. doi:10.1002/hbm.24716

50. Gennatas ED, Friedman JH, Ungar LH, et al. Expert-augmented machine learning. *Proc Natl Acad Sci U S A*. 2020;117(9):4571-4577. doi:10.1073/pnas.1906831117

51. Quade D. Using weighted rankings in the analysis of complete blocks with additive block effects. *J Am Stat Assoc*. 1979;74(367):680-683. doi:10.1080/01621459.1979.10481670

52. Heckert NA, Filliben JJ. *Dataplot Reference Manual, Volume 2: LET Subcommands and Library Functions*. Published 2003. Accessed November 2, 2020. https://www.itl.nist.gov/div898/handbook/eda/section3/eda3672.htm

53. Gómez-Verdejo V, Parrado-Hernández E, Tohka J; Alzheimer's Disease Neuroimaging Initiative. Sign-consistency based variable importance for machine learning in brain imaging. *Neuroinformatics*. 2019;17(4):593-609. doi:10.1007/s12021-019-9415-3

54. Koutsouleris N, Borgwardt S, Meisenzahl EM, Bottlender R, Möller HJ, Riecher-Rössler A. Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: results from the FePsy study. *Schizophr Bull*. 2012;38(6):1234-1246. doi:10.1093/schbul/sbr145

55. Madras D, Pitassi T, Zemel R. Predict responsibly: improving fairness and accuracy by learning to defer. In: Bengio S, Wallach H, Larochelle H, Grauman K, CesaBianchi N, Garnett R, eds. *Advances in Neural Information Processing Systems*. Volume 31. Curran Associates Inc; 2018:6147–6157.

56. Zavos HMS, Eley TC, McGuire P, et al. Shared etiology of psychotic experiences and depressive symptoms in adolescence: a longitudinal twin study. *Schizophr Bull*. 2016;42(5):1197-1206. doi:10.1093/schbul/sbw021

57. Callaway DA, Perkins DO, Woods SW, Liu L, Addington J. Movement abnormalities predict transitioning to psychosis in individuals at clinical high risk for psychosis. *Schizophr Res*. 2014;159(2-3):263-266. doi:10.1016/j.schres.2014.09.031

58. Green MF, Horan WP, Lee J. Nonsocial and social cognition in schizophrenia: current evidence and future directions. *World Psychiatry*. 2019;18(2): 146-161. doi:10.1002/wps.20624

59. Thompson A, Papas A, Bartholomeusz C, et al. Social cognition in clinical "at risk" for psychosis and first episode psychosis populations. *Schizophr Res*. 2012;141(2-3):204-209. doi:10.1016/j.schres.2012.08.007

60. Barbato M, Liu L, Cadenhead KS, et al. Theory of mind, emotion recognition and social perception in individuals at clinical high risk for psychosis: findings from the NAPLS-2 cohort. *Schizophr Res Cogn*. 2015;2(3):133-139. doi:10.1016/j.scog.2015.04.004

61. Schiffman J. Motor issues in the clinical high risk phase of psychosis. *Schizophr Bull*. 2017;43(5):937-938. doi:10.1093/schbul/sbx086

62. Bora E, Berk M. Theory of mind in major depressive disorder: a meta-analysis. *J Affect Disord*. 2016;191:49-55. doi:10.1016/j.jad.2015.11.023

63. Lindgren M, Jonninen M, Jokela M, Therman S. Adolescent psychosis risk symptoms predicting persistent psychiatric service use: a 7-year follow-up study. *Eur Psychiatry*. 2019;55:102-108. doi:10.1016/j.eurpsy.2018.10.004

64. Beck K, Studerus E, Andreou C, et al. Clinical and functional ultra-long-term outcome of patients with a clinical high risk (CHR) for psychosis. *Eur Psychiatry*. 2019;62:30-37. doi:10.1016/j.eurpsy.2019.08.005

65. Woods SW, Powers AR III, Taylor JH, et al. Lack of diagnostic pluripotentiality in patients at clinical high risk for psychosis: specificity of comorbidity persistence and search for pluripotential subgroups. *Schizophr Bull*. 2018;44(2):254-263. doi:10.1093/schbul/sbx138

66. Jonas KG, Lencz T, Li K, et al. Schizophrenia polygenic risk score and 20-year course of illness in psychotic disorders. *Transl Psychiatry*. 2019;9(1):300. doi:10.1038/s41398-019-0612-5

67. Hess JL, Tylee DS, Mattheisen M, et al; Schizophrenia Working Group of the Psychiatric Genomics Consortium; Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH). A polygenic resilience score moderates the genetic risk for schizophrenia. Published online September 6, 2019. *Mol Psychiatry*. doi:10.1038/s41380-019-0463-8

68. Takahashi T, Suzuki M. Brain morphologic changes in early stages of psychosis: implications for clinical application and early intervention. *Psychiatry Clin Neurosci*. 2018;72(8):556-571. doi:10.1111/pcn.12670

69. Sasabayashi D, Takayanagi Y, Takahashi T, et al. Increased occipital gyrification and development of psychotic disorders in individuals with an at-risk mental state: a multicenter study. *Biol Psychiatry*. 2017;82(10):737-745. doi:10.1016/j.biopsych.2017.05.018

70. Okada N, Yahata N, Koshiyama D, et al. Abnormal asymmetries in subcortical brain volume in early adolescents with subclinical psychotic experiences. *Transl Psychiatry*. 2018;8(1):254. doi:10.1038/s41398-018-0312-6

71. de Wit S, Wierenga LM, Oranje B, et al. Brain development in adolescents at ultra-high risk for psychosis: longitudinal changes related to resilience. *Neuroimage Clin*. 2016;12:542-549. doi:10.1016/j.nicl.2016.08.013

72. Escudero J, Ifeachor E, Zajicek JP, Green C, Shearer J, Pearson S; Alzheimer's Disease Neuroimaging Initiative. Machine learning-based method for personalized and cost-effective detection of Alzheimer's disease. *IEEE Trans Biomed Eng*. 2013;60(1):164-168. doi:10.1109/TBME.2012.2212278

73. Schultze-Lutter F, Schmidt SJ, Theodoridou A. Psychopathology: a precision tool in need of re-sharpening. *Front Psychiatry*. 2018;9:446. doi:10.3389/fpsyt.2018.00446

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

75. Brady LS, Potter WZ, Gordon JA. Redirecting the revolution: new developments in drug development for psychiatry. *Expert Opin Drug Discov*. 2019;14(12):1213-1219. doi:10.1080/17460441.2019.1666102