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; Katharina Chisholm, PhD; Anita Riecher-Rössler, MD; Laura Egloff, PhD; André Schmidt, MD; 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 Rösler, MD; Oleg Borisov, MD; Peter K. Kwon, 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; Markus Noethen, MD; Paolo Brambilla, MD; Stephen J. Wood, PhD; Rachel Uptegrove, MD, PhD; Frauke Schultze-Lutter, PhD; Anastasia Theodoridou, MD; Eva Meisenzahl, MD; for the PRONIA Consortium

IMPORATANCE 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.
The clinical high-risk (CHR) criteria for psychosis have been established to detect vulnerable individuals as early as possible to intercept disease development. These criteria identify a patient population with increased incidence compared with the general population, yet only 22% of patients with CHR as detected by ultra-high-risk criteria show a psychosis transition during a 3-year period. 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. 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 or machine learning, with studies finding that a first episode can be predicted with clinical data, combinations of clinical and cognitive data, neuroimaging, and, recently, with polygenic risk scores (PRS) for schizophrenia, among other measures. However, reviews have also highlighted methodological shortcomings, such as inadequate sample sizes and model validation strategies that may have inflated accuracy. Moreover, studies suggested that psychosis does not only emerge from CHR states but occurs and can be predicted across a broader spectrum of comorbid conditions commencing in late adolescence and early adulthood. 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.

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. 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.

Addressing these challenges, the European Union–funded PRONIA study (Personalised Prognostic Tools for Early Psychosis Management) 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, 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 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. 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. 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, 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 ILM) or earlier attrition (validation sample [n = 88]; PRONIA minus ILM) (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)* | PRONIA plus 18M cohort (n = 246) | PRONIA minus 18M cohort (n = 88) | Statistical analysis | P value (FDR)* |
|----------|-------------------------------------|--------------------------------|----------------------|----------------|-------------------------------|-------------------------------|----------------------|----------------|
| Samples and study variables | | | | | | | | |
| Sample site, No. of patients | | | | | | | | |
| Munich | 10 | 99 | | | | | | |
| Milan | 2 | 23 | | | | | | |
| Basel | 2 | 32 | | | | | | |
| Cologne | 4 | 59 | $\chi^2 = 6.78$ | .56 | | | | | $\chi^2 = 6.53$ | .83 |
| Birmingham | 2 | 33 | | | | | | |
| Turku | 6 | 32 | | | | | | |
| Udine | 0 | 30 | | | | | | |
| Follow-up interval, mean (SD), d | 628.3 (316.7) | 727.6 (310.7) | $t = -1.45$ | .38 | 842.7 (272.3) | 390.6 (99.6) | $t = 22.10$ | <.001 |
| Sociodemographic data | | | | | | | | |
| Age, mean (SD), y | 23.8 (5.6) | 24.8 (5.8) | $t = -0.81$ | .65 | 24.6 (5.6) | 25.0 (6.1) | $t = -0.59$ | .86 |
| Female, No (%) | 10 (38.5) | 160 (51.9) | $\chi^2 = 1.75$ | .38 | 124 (51.6) | 46 (52.3) | $\chi^2 = 0.09$ | .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) | $\chi^2 = 3.35$ | .65 | 3 (1.2) | 1 (1.1) | | |
| 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) | $t = -0.17$ | .92 | 23.6 (4.6) | 24.0 (4.6) | $t = -0.72$ | .83 |
| Edinburgh Handedness score, mean (SD)$c$ | 56.7 (66.6) | 70.4 (50.6) | $t = -1.28$ | .38 | 68.0 (53.1) | 72.9 (49.4) | $t = -0.70$ | .83 |
| Education, mean (SD), y | 13.3 (2.5) | 14.3 (2.9) | $t = -1.65$ | .28 | 14.3 (3.0) | 14.1 (2.7) | $t = 0.52$ | .86 |
| Educational problems, mean (SD), y repeated | 0.67 (0.88) | 0.26 (0.61) | $t = 3.20$ | .02 | 0.30 (0.64) | 0.27 (0.66) | $t = -0.34$ | .93 |
| Having a partnership most of the time in the year before study inclusion, No. (%) | 8 (30.8) | 88 (28.6) | $\chi^2 = 0.57$ | .90 | 76 (30.9) | 20 (22.7) | $\chi^2 = 2.11$ | .59 |
| CHR criteria met, No. (%) | | | | | | | | |
| Schizotypal personality disorder present | 2 (7.7) | 9 (2.9) | $\chi^2 = 1.71$ | .38 | 10 (4.1) | 1 (1.1) | $\chi^2 = 1.75$ | .67 |
| First-degree relatives with psychosis | 3 (11.5) | 25 (8.1) | $\chi^2 = 0.37$ | .83 | 20 (8.1) | 8 (9.1) | $\chi^2 = 0.08$ | .93 |
| 30% Loss of global functioning compared with highest levels in the year before study inclusion | 9 (34.6) | 88 (28.6) | $\chi^2 = 0.43$ | .81 | 72 (29.3) | 25 (28.4) | $\chi^2 = 0.02$ | .93 |
| Criteria | | | | | | | | |
| GRDP | 2 (7.7) | 23 (7.5) | $\chi^2 = 0.002$ | 1.00 | 18 (7.3) | 7 (8.0) | $\chi^2 = 0.04$ | .93 |
| COGDIS | 11 (42.3) | 83 (26.9) | $\chi^2 = 2.80$ | .29 | 69 (28.0) | 25 (28.4) | $\chi^2 = 0.004$ | .95 |
| APS | 18 (69.2) | 88 (28.6) | $\chi^2 = 18.30$ | <.001 | 85 (34.6) | 21 (23.9) | $\chi^2 = 3.42$ | .58 |
| BLIPS | 2 (7.7) | 5 (1.6) | $\chi^2 = 4.30$ | .28 | 4 (1.6) | 3 (3.4) | $\chi^2 = 1.00$ | .83 |
| CHR | 23 (88.5) | 144 (46.8) | $\chi^2 = 16.68$ | <.001 | 126 (51.2) | 41 (46.6) | $\chi^2 = 0.55$ | .83 |
| Functioning, mean (SD) | | | | | | | | |
| GF:Sd | | | | | | | | |
| 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:Rd | | | | | | | | |
| Highest lifetime | 7.92 (0.74) | 8.11 (0.83) | $z = -1.27$ | .38 | 8.15 (0.78) | 7.97 (0.94) | $z = -1.17$ | .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 | | | | | | | | |
| Positive symptoms | 1.77 (0.81) | 0.83 (0.80) | $t = -5.06$ | <.001 | 0.91 (0.82) | 0.89 (0.90) | $t = 0.25$ | .93 |
| Negative symptoms | 1.90 (1.57) | 1.55 (1.02) | $t = -0.54$ | .75 | 1.58 (1.04) | 1.48 (0.99) | $t = 1.20$ | .70 |
| Disorganized symptoms | 1.07 (1.15) | 0.66 (0.60) | $t = -1.84$ | .26 | 0.69 (0.62) | 0.60 (0.55) | $t = 1.54$ | .59 |
| General psychopathology | 2.29 (1.19) | 1.86 (0.95) | $t = -1.65$ | .28 | 1.92 (0.94) | 1.71 (0.95) | $t = 2.24$ | .47 |

(continued)
Syndromes reached psychotic intensity daily for at least 7 days. Diagnoses of cases with psychosis transition are listed in Table 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 datasets 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

| Variable | Psychosis transition cohort (n = 26) | Nontransition cohort (n = 308) | Statistical analysis | P value (FDR)* | PRONIA plus 18M cohort (n = 246) | PRONIA minus 18M cohort (n = 88) | Statistical analysis | P value (FDR)* |
|----------|-------------------------------------|-----------------------------|----------------------|----------------|-------------------------------|-------------------------------|----------------------|----------------|
| 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) | $X^2 = 8.20$ | .21 | 102 (41.8) | 39 (44.8) | $X^2 = 1.12$ | .93 |
| 1 | 13 (50.0) | 93 (30.4) | | | 78 (32.0) | 27 (31.0) | | |
| 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) | $X^2 = 3.65$ | .26 | 47 (19.2) | 20 (23.3) | $X^2 = 0.58$ | .83 |
| Yes | 17 (65.4) | 248 (81.0) | | | 102 (41.8) | 67 (77.0) | | |
| Affective disorder, No. (%) | | | | | | | | |
| None | 12 (46.2) | 200 (65.4) | | | 154 (62.9) | 59 (67.8) | $X^2 = 1.55$ | .83 |
| 1 | 14 (53.8) | 98 (32.0) | $X^2 = 5.45$ | .26 | 86 (35.1) | 25 (28.7) | | |
| 2 | 0 | 8 (2.6) | | | 5 (2.0) | 3 (3.4) | | |
| Substance use disorder, No. (%) | | | | | | | | |
| No | 25 (96.2) | 302 (98.7) | $X^2 = 1.64$ | <.01 | 241 (98.8) | 86 (98.9) | $X^2 = 0.003$ | .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) | $X^2 = 1.35$ | .86 | 45 (18.4) | 21 (24.1) | | |
| 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) | $X^2 = 0.90$ | .83 | 14 (5.7) | 1 (1.1) | | |
| 2 | 0 | 3 (1.0) | | | 1 (0.4) | 2 (2.3) | | |
| Treatments, No. (%) | | | | | | | | |
| Antipsychotics | 10 (38.5) | 51 (17.2) | $X^2 = 7.08$ | .09 | 40 (20.2) | 14 (15.9) | $X^2 = 0.76$ | .83 |
| Antidepressants | 13 (50.0) | 176 (57.1) | $X^2 = 0.50$ | .74 | 140 (56.7) | 48 (54.5) | $X^2 = 0.25$ | .86 |
| Inpatient | 15 (60.0) | 144 (46.8) | $X^2 = 2.10$ | .38 | 122 (50.2) | 36 (40.9) | $X^2 = 2.24$ | .59 |
| Psychotherapy | 14 (56.0) | 229 (74.4) | $X^2 = 5.09$ | .21 | 179 (73.7) | 62 (70.5) | $X^2 = 0.34$ | .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.

f Excludes major depressive disorder.
hood adversity,45,46 and neurocognitive measures6,47 in the clinical-neurocognitive domain (eTable 5 in the Supplement); PRS for schizophrenia48 in the genetic domain16; and gray matter volume maps in the structural magnetic resonance imaging (sMRI) domain8,9 (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 generalization49 improved prognostic accuracy (eFigures 2 and 3 in the Supplement and Table 2).50 Following the concept of expert-based machine learning,50 we integrated our raters' estimates as additional predictors to produce a cybernetic model51 (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,52 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 nontransitional 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,11 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 I] 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)54 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).57 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 test51 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],

jamapsychiatry.com
Table 2. Prediction Performance of Clinical Raters; Unimodal, Stacked, and Cybernetic Risk Calculators; and Prognostic Workflows

| Model by cohort                        | No. of findings | Sensitivity, % | Specificity, % | BAC, % | PPV   | NPV   | PSI   | Positive LR | AUC  | P value for FDR |
|----------------------------------------|-----------------|----------------|----------------|--------|-------|-------|-------|-------------|------|-----------------|
| **Rater-based estimates**              |                 |                |                |        |       |       |       |             |      |                 |
| PRONIA plus 18M                        | 16              | 61.5           | 84.9           | 73.2   | 32.7  | 94.9  | 27.5  | 4.1         | 0.73 | NA              |
| PRONIA minus 18M                       | NA              | 87.2           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Complete PRONIA                        | 16              | 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              | 84.6           | 66.8           | 75.7   | 23.2  | 97.4  | 20.5  | 2.6         | 0.83 | NA              |
| PRONIA minus 18M                       | NA              | 65.9           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Complete PRONIA                        | 21              | 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                       | NA              | 65.9           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| ZInEP                                 | 14              | 87.5           | 43.1           | 65.3   | 15.9  | 96.6  | 12.5  | 1.5         | 0.67 | NA              |
| BEARS-Kid                              | 10              | 76.9           | 63.9           | 70.4   | 5.8   | 99.0  | 4.8   | 2.1         | 0.68 | NA              |
| **PRS-based risk calculator**          |                 |                |                |        |       |       |       |             |      |                 |
| PRONIA plus 18M                        | 22              | 88.0           | 53.5           | 70.7   | 17.9  | 97.5  | 15.4  | 1.9         | 0.70 | NA              |
| PRONIA minus 18M                       | NA              | 48.6           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Complete PRONIA                        | 22              | 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              | 80.8           | 61.9           | 68.5   | 60.0  | 76.5  | 36.5  | 2.0         | 0.71 | NA              |
| PRONIA minus 18M                       | NA              | 45.5           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Complete PRONIA                        | 22              | 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                       | NA              | 80.8           | 86.0           | 83.4   | 32.8  | 98.1  | 31.0  | 5.8         | 0.89 | NA              |
| Complete PRONIA                        | 21              | 80.8           | 86.9           | 83.8   | 38.9  | 97.4  | 36.3  | 5.4         | 0.88 | NA              |
| **Stacked Clin-NC, PRS, and sMRI risk calculator** | | | | | | | | | | |
| PRONIA plus 18M                        | 21              | 80.8           | 85.0           | 82.9   | 38.9  | 97.4  | 36.3  | 5.4         | 0.88 | NA              |
| PRONIA minus 18M                       | NA              | 80.7           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Complete PRONIA                        | 18              | 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              | 84.6           | 86.4           | 85.5   | 42.3  | 97.9  | 40.2  | 6.2         | 0.90 | NA              |
| PRONIA minus 18M                       | NA              | 80.8           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Complete PRONIA                        | 21              | 80.8           | 86.4           | 83.6   | 33.3  | 98.2  | 31.5  | 5.9         | 0.90 | <.001           |
| **Optimal prognostic workflow**        |                 |                |                |        |       |       |       |             |      |                 |
| PRONIA plus 18M                        | 22              | 84.6           | 87.3           | 85.9   | 44.0  | 98.0  | 42.0  | 6.6         | 0.90 | NA              |
| PRONIA minus 18M                       | NA              | 84.1           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Complete PRONIA                        | 21              | 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                | NA              | 84.1           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Workflow optimized for light examination sparsity | | | | | | | | | | |
| PRONIA plus 18M                        | 21              | 80.8           | 85.9           | 83.3   | 40.4  | 97.4  | 37.8  | 5.7         | 0.87 | NA              |
| PRONIA minus 18M                       | NA              | 84.1           | NA             | NA     | NA    | NA    | NA    | NA          | NA   | NA              |
| Complete PRONIA                        | 19              | 73.1           | 86.4           | 79.7   | 31.1  | 97.4  | 28.6  | 5.4         | 0.85 | <.001           |

(continued)
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 childhood, 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

---

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

| Model by cohort | No. of findings | True positive | True negative | False positive | False negative | Sensitivity, \% | Specificity, \% | BAC, \% | PPV | NPV | PSI | Positive LR | AUC | \( P \) value for FDR |
|-----------------|-----------------|---------------|---------------|----------------|----------------|----------------|----------------|---------|-----|-----|-----|-------------|-----|------------------|
| PRONIA minus 18M cohort\(^{ab}\) | NA | 75 | 13 | NA | NA | 85.2 | NA | NA | NA | NA | NA | NA | NA |
| Workflow optimized for strong examination sparsity\(^{a}\) | 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\(^{a}\) | 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\(^{ab}\) | 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.

\( ^{c} \) \( \Gamma = 0.0 \) (case propagation cutoffs: 25.0% and 100%). Sequence is Clin-NC, Clin-NC plus raters, Clin-NC plus PRS, and Clin-NC plus sMRI.

\( ^{d} \) \( \Gamma = 0.5 \) (case propagation cutoffs: 37.5% and 100%). Sequence is Clin-NC, Clin-NC plus PRS, Clin-NC plus raters, and Clin-NC plus sMRI.

\( ^{e} \) \( \Gamma = 1.0 \) (case propagation cutoffs: 37.5% and 75.0%). Sequence is Clin-NC, Clin-NC plus PRS, Clin-NC plus raters, and Clin-NC plus s MRI.
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 t326 = 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-

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.
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 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 $\Gamma = 0.5$) and 0 (regularization strength $\Gamma = 1.0$) (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 or regularized ($\Gamma = 0.5$) 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...
Multimodal risk calculators outperformed all unimodal counterparts ($t$ range, 3.14–6.20; $P < 3.99 \times 10^{-8}$ 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%]; $t = 3.18; P = .001$ for FDR).
part of the model discovery process, which further supports in patients with CHR, depended on patients with ROD being accuracy, particularly the sensitivity for psychosis transition 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, [72x217]which point to ongoing compensatory mechanisms of resilience to psychosis, as reported previously in a longitudinal sMRI study of adolescents with CHR. 21 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. 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.

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. 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, 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 PRO-NIA discovery sample. This sample size increased the risk of producing overly optimistic prediction results owing to an accidental collection of well-classifiable cases. We implemented a multistep model validation procedure to guard against this possibility, including label permutation testing, strict nested cross-validation of all processing steps, 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. 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, may help to propel a reciprocal and iterative process of clinical validation and refinement of these prognostic tools in real-world early recognition services.
models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset psychosis: 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. Carrón RE, Combblatt 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, Barbeje 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(17):1163-1178. doi:10.1017/S0033291716034949

19. Sanfelici R, Dwyer DB, Antonucci LA, Koutsouris 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, Wallow. Choosing prediction over explanation in psychology: lessons from machine learning. *Perspect Psychol Sci*. 2017; 2(5):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, Hajiiz Z, Stahl D, Steyerberg EW. The science of prognostic psychiatry: a review. *JAMA Psychiatry*. 2018; 75(2):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.20104

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/bst036

26. Koutsouris N, Meisenzl 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, Trender G, am der Heiden W, Schmidt M, Könecke R. Schizophrenia and depression: challenging the paradigm of two separate diseases—a controlled study of schizophrenia, depression and healthy controls. *Schizophr Res*. 2005;77:11-24. doi:10.1016/j.schres.2005.01.004

28. Wigram JT, 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. Seibald C, Khandaler GM, Zannutti S, Jones PB. Association between childhood psychopathological disorders and psychotic experiences in adolescence: a population-based longitudinal study. *Constr Psychopathol*. 2016;6:45-52. doi:10.1007/s10549-016-0116-6

30. Mulsiner KL, Mortensen PB, McGrath J, 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.2018.4866

31. Schmidt A, Cappuccini M, Radaia 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. Rajkumar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-1358. doi:10.1056/NEJMr1814299

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

34. Clark SR, Schubert KA, Baune BT. Towards indicated prevention of psychosis: using probabilistic assessments of transition risk in psychosis premon. *J Neurol 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):142-155. doi:10.1111/j.1600-0447.2006.00854.x

36. Theodoroudou 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, Marz-Imgartinger A, Linder C, Schultz-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, Altmann 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
Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Koutsouleris, N; Dwyer, DB; Degenhardt, F; Maj, C; Urquijo-Castro, MF; Sanfelici, R; Popovic, D; Oeztuerk, O; Haas, SS; Weiske, J; Ruef, A; Kambeitz-Ilankovic, L; Antonucci, LA; Neufang, S; Schmidt-Kraepelin, C; Ruhrmann, S; Penzel, N; Kambeitz, J; Haidl, TK; Rosen, M; Chisholm, K; Riecher-Rossler, A; Egloff, L; Schmidt, A; Andreou, C; Hietala, J; Schirmer, T; Romer, G; Walger, P; Franscini, M; Traber-Walker, N; Schimmelmann, BG; Fluckiger, R; Michel, C; Rossler, W; Borisov, O; Krawitz, PM; Heekeren, K; Buechler, R; Pantelis, C; Falkai, P; Salokangas, RKR; Lencer, R; Bertolino, A; Borgwardt, S; Noethen, M; Brambilla, P; Wood, SJ; Upthegrove, R; Schultze-Lutter, F; Theodoridou, A; Meisenzahl, E

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

Date:
2020-12-02

Citation:
Koutsouleris, N., Dwyer, D. B., Degenhardt, F., Maj, C., Urquijo-Castro, M. F., Sanfelici, R., Popovic, D., Oeztuerk, O., Haas, S. S., Weiske, J., Ruef, A., Kambeitz-Ilankovic, L., Antonucci, L. A., Neufang, S., Schmidt-Kraepelin, C., Ruhrmann, S., Penzel, N., Kambeitz, J., Haidl, T. K., ... Meisenzahl, E. (2020). Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. JAMA PSYCHIATRY, 78 (2), pp.195-209. https://doi.org/10.1001/jamapsychiatry.2020.3604.

Persistent Link:
http://hdl.handle.net/11343/273966

File Description:
Published version