

Original Article

*These authors contributed equally to this work.

Cite this article: Moore TM *et al* (2021). Development of a probability calculator for psychosis risk in children, adolescents, and young adults. *Psychological Medicine* 1–9. <https://doi.org/10.1017/S0033291720005231>

Received: 17 June 2020

Revised: 30 October 2020

Accepted: 9 December 2020

Key words:

Feature-selection; prodrome; psychosis spectrum; recruiting; risk calculator

Author for correspondence:

Tyler M. Moore,

E-mail: tymoore@penmedicine.upenn.edu

Development of a probability calculator for psychosis risk in children, adolescents, and young adults

Tyler M. Moore^{1,2,*} , Monica E. Calkins^{1,2,*}, Adon F. G. Rosen¹, Ellyn R. Butler¹, Kosha Ruparel^{1,2}, Paolo Fusar-Poli^{3,4,5}, Nikolaos Koutsouleris⁶, Philip McGuire³, Tyrone D. Cannon⁷, Ruben C. Gur^{1,2} and Raquel E. Gur^{1,2}

¹Department of Psychiatry, Brain Behavior Laboratory, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; ²Lifespan Brain Institute, Penn Medicine and Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA; ³Department of Psychosis Studies, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK; ⁴OASIS service, South London and Maudsley NHS Foundation Trust, London, UK; ⁵Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy; ⁶Department of Psychiatry and Psychotherapy, Ludwig-Maximilian-University Munich, Germany and ⁷Departments of Psychology and Psychiatry, Yale University, New Haven, CT 06520, USA

Abstract

Background. Assessment of risks of illnesses has been an important part of medicine for decades. We now have hundreds of 'risk calculators' for illnesses, including brain disorders, and these calculators are continually improving as more diverse measures are collected on larger samples.

Methods. We first replicated an existing psychosis risk calculator and then used our own sample to develop a similar calculator for use in recruiting 'psychosis risk' enriched community samples. We assessed 632 participants age 8–21 (52% female; 48% Black) from a community sample with longitudinal data on neurocognitive, clinical, medical, and environmental variables. We used this information to predict psychosis spectrum (PS) status in the future. We selected variables based on lasso, random forest, and statistical inference relief; and predicted future PS using ridge regression, random forest, and support vector machines.

Results. Cross-validated prediction diagnostics were obtained by building and testing models in randomly selected sub-samples of the data, resulting in a distribution of the diagnostics; we report the mean. The strongest predictors of later PS status were the Children's Global Assessment Scale; delusions of predicting the future or having one's thoughts/actions controlled; and the percent married in one's neighborhood. Random forest followed by ridge regression was most accurate, with a cross-validated area under the curve (AUC) of 0.67. Adjustment of the model including only six variables reached an AUC of 0.70.

Conclusions. Results support the potential application of risk calculators for screening and identification of at-risk community youth in prospective investigations of developmental trajectories of the PS.

A person who eventually develops a severe psychotic disorder (e.g. schizophrenia) usually shows signs early in life, years before the disorder is formally diagnosed (Keith & Matthews, 1991; Yung & McGorry, 1996). Symptoms in the early 'pre-disorder' stage—formerly called the 'prodrome'—allow care providers and researchers to assess the risk of future conversion to a disorder like schizophrenia (Nelson & McGorry, 2020; Yung *et al.*, 2003). Indeed, the discovery of the prodrome and even earlier pre-morbid symptoms (Brown, 1963; Mahler, 1952) widened the view of psychosis from a disorder of early adulthood to a disorder of the lifespan (Friedman *et al.*, 2001; Powers *et al.*, 2020). This perspective, in turn, has led to substantial research on signs and symptoms that might be detected before the transition to psychosis (Miller *et al.*, 1999; Woodberry, Shapiro, Bryant, & Seidman, 2016). A promising potential of measuring such symptoms is that psychosis risk and transition can be *predicted*.

Assessing risk – i.e. estimating the probability of an event occurring given some known information – has been an integral part of medicine's role in prognosis (Combe, Donkin, Buchanan, & Mackenzie, 1820). The Framingham study (Dawber, Moore, & Mann, 1957) and subsequent analyses (Mahmood, Levy, Vasani, & Wang, 2014) showed compellingly that statistical models can predict the future better than the average clinician. Some successful contemporary calculators assess risk, for example, of complications from cardiac surgery (Gupta *et al.*, 2011), complications from pancreatectomy (Parikh *et al.*, 2010), general surgical complications (Bilimoria *et al.*, 2013), undiagnosed diabetes (Heikes, Eddy, Arondekar, & Schlessinger, 2008), periodontal disease (Page, Krall, Martin, Mancl, & Garcia, 2002), bone fracture risk (Leslie & Lix, 2014), and hundreds more. Notably, risk calculators have more recently included 'mental' illnesses like psychosis, the focus of the present study. Cannon

et al. (2016) developed a calculator for risk of conversion from clinical high risk (CHR) to frank psychosis within a 2-year window using time-to-event (Cox) regression. They found that psychosis conversion was best predicted by positive psychosis symptoms, declining social function, and poor verbal learning. This calculator was later replicated and extended by Carrión et al. (2016), Zhang et al. (2018) and Osborne and Mittal (2019). Fusar-Poli et al. (2017, 2019) developed a calculator to forecast the transdiagnostic risk of developing psychosis in secondary care, where predictors (demographics and any index diagnosis of non-psychotic mental disorder) were selected based on a priori knowledge [see Riecher-Rossler and Studerus (2017), Radua et al. (2018), Adibi, Sadatsafavi, and Ioannidis, (2020), and Sanfelici, Dwyer, Antonucci, and Koutsouleris (2020) for reviews].

Importantly, currently available psychosis risk calculators were developed in individuals who were seeking clinical care because of psychosis spectrum (PS) symptoms, and thus apply to youth who are already experiencing some distress and/or impairment. A complementary approach to risk identification is through general population or community samples, which aims the ascertainment lens at a broader range of individuals experiencing PS symptoms (Taylor, Calkins, & Gur, 2020; Wigman et al., 2011). This approach may allow earlier identification of at-risk youth and commensurate enhanced opportunities to evaluate varying developmental trajectories and targeted early interventions. Among the few prospective studies in this area, several consistent findings have emerged indicating that persistence and worsening of PS symptoms are associated with particular symptoms, neurocognitive deficits, and neuroimaging parameters and other biomarkers (Calkins et al., 2017; Davies, Sullivan, & Zammit, 2018; Kalman, Bresnahan, Schulze, & Susser, 2019; Taylor et al., 2020). The development and application of a community applied psychosis risk calculator could greatly facilitate the aims of such endeavors, potentially accelerating discoveries and treatment innovations earlier in the pathway to care than is currently feasible.

Given the moderate success of prior CHR calculators, but the different ascertainment strategies of CHR and community-based cohorts, which can have a substantial role in enriching the risk to psychosis (Fusar-Poli et al., 2016), a critical question is whether prior calculators are applicable to community samples. The present study, therefore, had two goals. First, we aimed to evaluate the construct validity (Cronbach & Meehl, 1955, pp. 282–283) of the Cannon et al. (2016) psychosis risk calculator of the North American Prodrome Longitudinal Study (NAPLS) in the Philadelphia Neurodevelopmental Cohort (PNC), using variables as similar as possible to those used in the original study. Second, we aimed to develop and internally validate a new calculator designed to predict the risk of PS status in a community cohort of young people aged 8–21. That is, rather than focus on the transition to threshold psychosis, which may be the optimal focus for clinical applications, we focus on the risk of occurrence of PS symptoms in youth, which has practical scientific purposes such as evaluating neurodevelopmental biobehavioral trajectories in a youth sample enriched with potential for transition to psychosis.

Methods

Participants

Participants ($n = 632$) were recruited for follow-up based on Time 1 PS screening of the PNC (Calkins et al., 2014, 2015; Moore et al., 2016). PNC participants at Time 1 included ~10 000 genotyped

Table 1. Time 1 sample demographic and clinical information for full sample ($N = 632$)

Variable	Value
Age, years (mean, s.d.)	14.8 (3.2)
Female	0.52
African American	0.48
Caucasian	0.41
Hispanic	0.06
Parent Ed., mean yrs. (s.d.)	14.0 (2.2)
Major depressive episode	0.15
Generalized anxiety	0.02
Obsessive-compulsive	0.04
Attention deficit and hyperactivity	0.17

Note. Values are proportions unless otherwise specified; yrs = years; s.d. = standard deviation.

youth aged 8–21 years at enrollment (2009–2011), recruited from pediatric, non-psychiatric services of the Children's Hospital of Philadelphia (CHOP) health care network. The youth were in stable health, proficient in English, and physically and cognitively capable of participating in a clinical assessment interview and computerized neurocognitive testing. Participants provided informed consent/assent and permission to re-contact after receiving a complete description of the study and the Institutional Review Boards at Penn and CHOP approved the protocol. As detailed previously (Calkins et al., 2017), participants who screened either positive ($n = 265$) or negative ($n = 367$) for PS symptoms at Time 1 were identified for follow-up assessment if they were physically healthy at Time 1 (no moderate or severe physical conditions requiring multiple procedures and monitoring), had completed the neuroimaging protocol $> = 18$ months previously, and had good quality neuroimaging data. We emphasized for follow-up individuals from the PNC random subsample ($N = 1601$) who had also received multimodal neuroimaging at T1 (as detailed in Satterthwaite et al., 2014). Follow-up intervals ranged from 2 to 80 months (mean months = 42.9, s.d. = 16.5). Table 1 provides the Time 1 demographic characteristics of the sample, as well as rates of common mental disorders.

Measures

Clinical assessment

Details of Time 1 (Calkins et al., 2014, 2015; Moore et al., 2016) and follow-up (Calkins et al., 2017) assessments have been reported. Briefly, at Time 1, probands (age 11–21) and collaterals (parent or legal guardian for probands aged 8–17) were administered a computerized structured interview (GOASSESS). This instrument assessed psychiatric and psychological treatment history, and lifetime occurrence of major domains of psychopathology – including mood, anxiety, behavioral and eating disorders – and suicidal thinking and behavior (Calkins et al., 2014, 2015). Three screening tools to assess PS symptoms were embedded within the psychopathology screen. Positive sub-psychotic symptoms in the past year were assessed with the 12-item assessor administered PRIME Screen-Revised (PS-R) (Kobayashi et al., 2008; Miller et al., 2004). Items were self-rated on a 7-point scale ranging from 0 ('definitely disagree') to 6 ('definitely agree'). Positive psychotic symptoms (lifetime hallucinations and delusions) were assessed using the Kiddie-Schedule for

Affective Disorders and Schizophrenia (K-SADS) (Kaufman et al., 1997) psychosis screen questions, supplemented with structured questions to reduce false positives. Negative/disorganized symptoms were assessed using six embedded assessor rated items from the Scale of Prodromal Symptoms (SOPS) (McGlashan et al., 2003).

History of exposure to traumatic stressors was tabulated from the post-traumatic stress disorder section of the GOASSESS, in which participants were asked about lifetime history of experiencing eight categories of events (i.e. natural disasters, witnessed violence, attacked physically, sexually assaulted/abused, threatened with a weapon, experienced a serious accident, witnessed serious physical injury/death, observed dead body).

Global function was rated using the Children's Global Assessment Scale (Shaffer et al., 1983).

An abbreviated version of the Family Interview for Genetic Studies (FIGS) (Maxwell, 1996), administered to collaterals (of probands <age 18) and adult probands, screened for presence or absence of a first-degree family history of major domains of psychopathology, with a more detailed assessment of possible psychotic disorders following affirmative responses to psychosis-related screening items. To avoid the influence of proband status on judgments about psychosis family history, the presence/absence was coded based on FIGS data contained in a blinded file, without reference to proband status at either Time 1 or follow-up (Calkins et al., 2017; Taylor et al., 2020).

At follow-up, psychopathology was assessed using a custom protocol (Calkins et al., 2017) consisting of modules of the K-SADS and the Structured Interview for Prodromal Syndromes (SIPS, version 4.0) (McGlashan et al., 2003) administered to probands (age 11 and up) and collaterals (of probands age 8–17). Following each evaluation, assessors integrated information from probands, collaterals, and available medical records to provide combined ratings across symptom domains. Integrated clinical information was then summarized in a narrative case history and presented at a case conference attended by at least two doctoral-level clinicians with expertise in psychosis and/or child psychopathology. Strict blinding was maintained such that recruiters, assessors and clinicians determining consensus ratings and diagnoses were naive to Time 1 PS screening status of all participants. To avoid biasing case assignment or symptom ratings, family history of psychopathology was not disclosed during the case conference. Each SOPS clinical rating ≥ 3 based on the SIPS interview underwent consensus review, and clinical risk status and best estimate final diagnoses for Axis I disorders were determined. Individuals were classified as meeting PS criteria if they had either (a) a DSM-IV psychotic disorder or mood disorder with psychotic features, or (b) at least one SOPS positive symptom currently (past 6 months) rated 3–5 or at least two negative and/or disorganized symptoms rated 3–6. See Calkins et al. (2017) for detailed training and assessment procedures.

Neurocognitive assessment

Time 1 neurocognition was assessed using the Penn Computerized Neurocognitive Battery (Penn CNB) (Gur et al., 2001, 2010; Moore, Reise, Gur, Hakonarson, & Gur, 2015), which comprises 14 tests grouped into five domains of neurobehavioral function. A full description of the Penn CNB, including a description of each individual test, is available in the Supplement.

Environmental exposures

Time 1 environment was assessed using a combination of, (1) self-reported traumatic experiences (as described above), and (2) neighborhood-level characteristics obtained by geocoding participants addresses to census and crime data in the Philadelphia area.

Neighborhood characteristics were measured at the block-group level and included median family income, percent of residents who are married, percent of real estate that is vacant, and several others; see Moore et al. (2016) for further details.

Statistical analyses

Quasi-Replication of Cannon et al. (2016)

The first goal of the present study was to replicate in the PNC the psychosis risk calculator results presented in Cannon et al. (2016), but note that a true replication (using the same variables and coefficients as in the published model) was not possible here. Our approach – testing most of the same variables as in the NAPLS study after re-estimating the coefficients – is best characterized as a ‘quasi-replication’ in the terminology of Coiera, Ammenwerth, Georgiou, and Magrabi (2018).

NAPLS identified the following variables as useful predictors of conversion from a CHR state to frank psychosis within 2 years: age, sum of Structured Interview for Psychosis-risk Syndromes (SIPS) items P1 (Unusual Thought Content) and P2 (Suspiciousness), the Brief Assessment of Cognition in Schizophrenia (BACS) symbol coding raw score, Hopkins Verbal Learning Test (HVLT), stressful life events, family history of psychosis, Global Scale of Functioning-Social (GFS-S) (decline in functioning), and traumatic events (>1). In addition to being useful predictors, the variables identified in the NAPLS study are supported by previous studies and can be obtained in general clinical settings. To replicate the findings of Cannon et al. (2016), we first selected variables in the PNC that most closely match the variables listed above. Online Supplementary Table S1 shows the NAPLS-2 variables used, along with their PNC equivalents. We had perfect or near-perfect matches for Age, Family History, and Traumatic Events, and only partial matches for SIPS P1 & P2, BACS symbol coding, HVLT, and GFS-S. No equivalent was found for stressful life events, though this is partly captured in the traumatic events count. In addition, the PNC sample in this specific replication analysis was limited to those who started the study (Time 1) with sub-threshold PS symptoms (same $N = 265$ PS positive detailed in the Participants sub-section), which is different from the data set (full $N = 632$) used for the construction of the new calculator (see below). This was done because the NAPLS calculator was meant to predict the transition from high-risk to frank psychosis and did not include low-risk people. Thus, the NAPLS-2 calculator was designed to detect a frank psychosis outcome in a sample of people with CHR, whereas the PNC-based calculator was designed to detect CHR/PS in a sample of non-help-seeking community participants. The outcome of interest was (binary) transition to threshold psychosis ($N = 26$ out of the 265) within 2 years of the first visit. As in Cannon et al. (2016), a Cox proportional hazards model (survival analysis) was used. The main metric used for assessment of prediction accuracy was area under the receiver operating characteristic (ROC) curve (Hanley & McNeil, 1982).

PNC-based risk calculator

Next, we wished to build a new psychosis risk calculator within parameters more appropriate to our whole longitudinal sample ($N = 632$, which includes the $N = 265$ CHR persons used above, plus $N = 367$ others, most typically developing). Rather than predicting transition from CHR to threshold psychosis, we aimed to predict the milder PS status. For these purposes someone who does the transition to frank psychosis (not the milder ‘psychosis

spectrum’) would be included as a ‘case’ here – i.e. we wished to predict transition to PS *or* frank psychosis.

To construct the PNC-based risk calculator, we combined three feature-selection methods with three prediction methods, completely crossed (for nine total) in a single cross-validated pipeline. Details are given below, but the core procedure involved splitting the data into testing and training sets, selecting variables and building the model in the training set, and then testing it in the testing set. The cross-validated framework was 10-fold, such that all participants received a predicted value based on variables selected (and model built) in 90% of the sample not including him/her. The 10-fold cross-validation was repeated 1000 times.

Selection of variables for the model

We used three different feature-selection algorithms to ensure multiple variable characteristics were considered in selecting them – e.g. in addition to main effects (Lasso), does the variable have a nonlinear relationship with the outcome (random forest), does the variable interact with other variables (moderation) in determining the outcome (Relieff and STIR)? For each (90% // 10%) split of the sample (each of the 10-folds), the algorithms below were run, giving three different sets of ‘optimal’ features for each split (all saved for subsequent analyses). The integer number of features selected was also saved. Brief descriptions of the algorithms follow, and additional detail is available in the Supplement.

1. Lasso regression. Lasso regression is a type of regularized regression that assesses a ‘penalty’ (forced downward bias of coefficients) for both the number of predictors used in the model and the collinearity among them (Tibshirani, 1996). Usually, the penalty causes most coefficients to become exactly zero, retaining a confined set of non-redundant predictors for prediction (i.e. features with non-zero coefficients are ‘good’).
2. Random forest importance. The random forest algorithm (Liaw & Wiener, 2002) leaves the realm of conventional linear modeling and incorporates decision trees. The first step of these decision trees is to determine which single variable best predicts PS in the training sample. Once that is determined, the algorithm splits the sample into those above ν . below the mean on the ‘important’ variable. In these split sub-samples, the algorithm then looks for the most important variable. Those sub-samples are then further split based on their ‘most important’ variables, etc.
3. Relieff and Statistical Inference Relief (STIR). Relieff is an algorithm designed specifically for feature selection and known for being especially sensitive to interactions among features (Le, Urbanowicz, Moore, & McKinney, 2019). Given n cases and p variables, Relieff first chooses a random case. In p -dimensional Euclidean space, the algorithm finds the nearest neighbor that is the same as the random case on the categorical dependent variable (DV) (a ‘hit’) and the nearest neighbor that is different from the random case on the DV (a ‘miss’). For any given variable, if the value of that variable for the randomly drawn case is closer to the ‘hit’ case than to the ‘miss’ case, the variable importance goes up; otherwise, it goes down. STIR adds to the Relieff procedure by calculating p values for the predictors (not used in the traditional Relieff algorithm). This allowed us to more confidently make decisions about inclusion of variables without accepting an arbitrary cutoff.

Comparing cross-validated prediction models

With the most important variables selected, the next step in the pipeline was to estimate a model using one of the three prediction

methods; therefore, a total of nine models were estimated in each fold, one for each combination of feature-selection and prediction algorithm. Use of multiple algorithms allowed us to answer, generally, which prediction pipeline is likely to perform best in a ‘final’ model. The prediction algorithms were as follows:

1. Ridge regression. Like lasso regression, ridge regression is a form of regularized regression that assesses penalties on the coefficients and is most often used for cross-validation. A major difference is that ridge regression does not shrink coefficients to zero (as does lasso), which was desirable here because the features had already been selected. It is well-established that ridge regression outperforms conventional linear regression in out-of-sample (i.e. cross-validated) prediction (McNeish, 2015).
2. Random forest. The random forest algorithm is described in the above section. Here the algorithm was used for prediction, whereas it had previously been used only for variable selection.
3. Support vector machines (SVMs). SVMs classify cases by finding a hyperplane that separates them (on all variables) with a maximum distance between the hyperplane and the cases (positive or negative). To illustrate an SVM consider the 2-variable case (two continuous predictor variables, X_1 and X_2) predicting a variable with two possible states (say, ‘infected’ or ‘not infected’). Graphing X_1 and X_2 against each other would yield a scatterplot where each point on the scatterplot was either infected or not infected. It would be possible to draw a line through the cloud of points (scatterplot) that maximally separated the infected from the not infected. This line would be the ‘hyperplane’ separating the points; if we added a third variable (X_3), the line would become a plane, and if we added >1 variable (X_4 , X_5 , etc.), the plane would become a hyperplane.

Final proposed risk calculator model

The results from the above analyses revealed which combination of feature-selection and prediction algorithms would likely be best in practice (i.e. predict most accurately if used as a risk calculator). A problem with the optimal result (see below) is that it required far too many variables for a risk calculator meant to be used by the public. We, therefore, estimated 10-fold cross-validated prediction accuracy using the top 2 variables from the final suggested model, top 3 variables, top 4, etc., up to the top 10 variables allowed. As expected, at first the cross-validated area under the curve (AUC) increased as variables were added, but it eventually started to decrease with additional variables. The maximum/optimal number of variables was taken as the final model. Once this number was established (3 variables? 8 variables?), the full sample was used to maximize estimation accuracy of the coefficients. Cross-validated prediction accuracy of the model therefore cannot be obtained until it is used in another, external sample.

R scripts used for all analyses above can be found at <https://www.mooremetrics.com/psy-risk-supplemental-files/>.

Results

Quasi-replication of NAPLS risk calculator

Online Supplementary Table S2 shows the results of the CoxPH model run using the sample of Time 1 PS participants ($N = 265$ who were on the PS at Time 1, not the full $N = 632$ who included typically developing youth). The strongest predictor of conversion to frank psychosis in the NAPLS calculator is Age (32% increased odds per age year), followed by PRIME total score (PS-R Total)

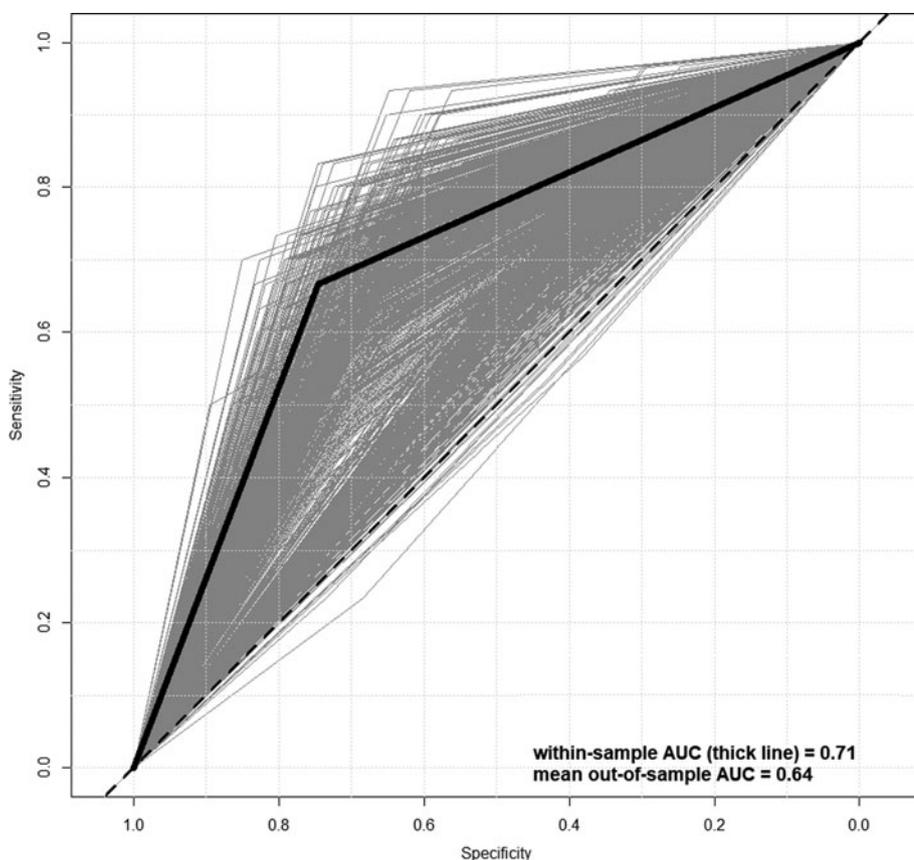


Fig. 1. Receiver operating characteristic curves for within-sample (thick) and out-of-sample (thin, gray) prediction of psychosis conversion.

(4% increased odds per point). **Figure 1** shows the ROC curve (black) corresponding to the model in online Supplementary Table S2. Within-sample prediction achieves an AUC of 0.71. To put these results into context, including how well we would expect them to cross-validate out-of-sample, we implemented two analyses. First, we ran 2-fold cross-validation on the model setup from online Supplementary Table S2—i.e. coefficients were estimated in a random 50% of the sample, and this model was tested (AUC obtained) using the left-out 50%. This was repeated so that each person had an out-of-sample prediction, quality of prediction was assessed using conventional metrics (AUCs, etc.), and this was repeated 10 000 times to get a distribution of cross-validated AUCs. The AUCs of the cross-validated models are shown in green in **Fig. 1**, and as expected, cross-validation reduced the AUC from 0.71 to 0.64, the latter below the conventional cutoff of 0.70.

As a secondary analysis, because we wanted to gauge how ‘impressive’ a within-sample AUC of 0.71 is, we compared the within-sample results (AUC = 0.71; black function in **Fig. 1**) to ‘random’ within-sample results using permuted labels for Psychosis. That is, the binary indicator for frank Psychosis was randomly re-assigned and the model re-estimated, giving a rough indication of what level of within-sample prediction accuracy one could expect purely by chance, given this number of variables distributed in this way with this specific (tiny) proportion of ‘hits’. The above permutation of labels was repeated 10 000 times, and the mean AUC was taken to be the AUC expected by chance. online Supplementary Figure S1 shows the results of the permutation analysis. The pink ‘cloud’ comprises 1000 of the 10 000 ROC curves (limited to 1000 for better visual) estimated for each permutation, and the black function is the

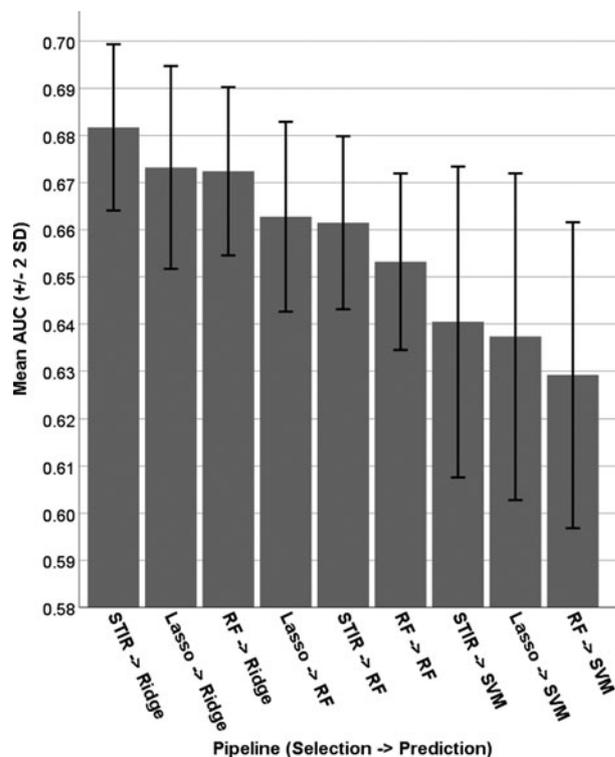


Fig. 2. Area under the ROC curve for nine combinations of feature-selection and prediction algorithms.

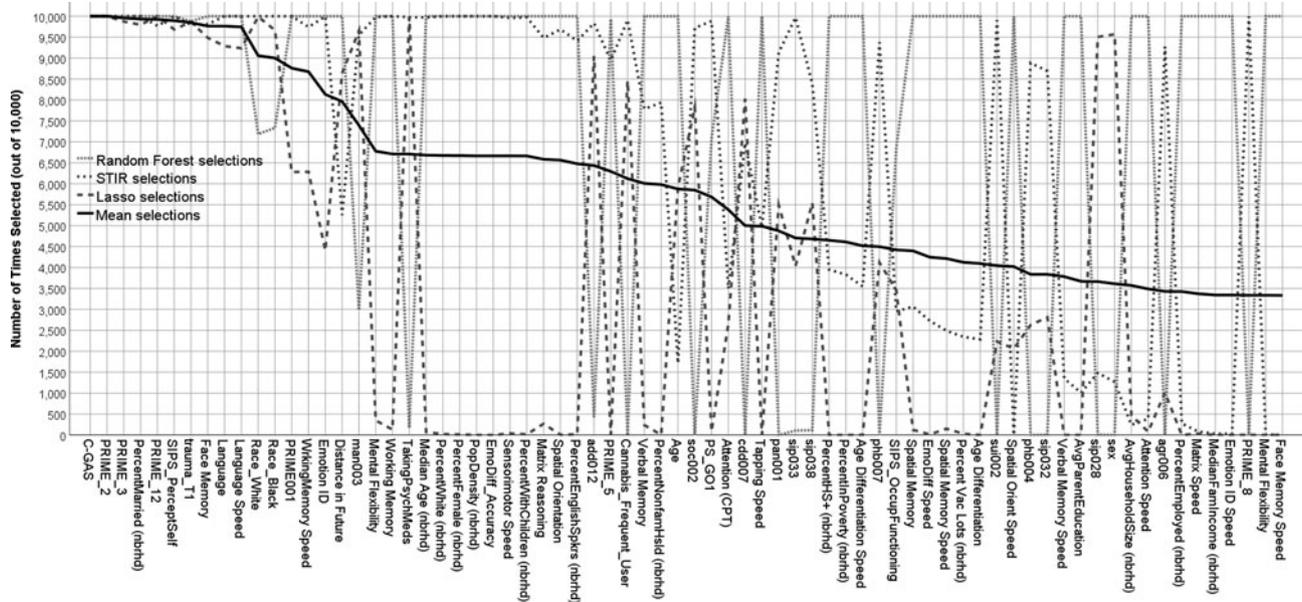


Fig. 3. Frequency of variable selection across random data partitions, by algorithm, in decreasing order of importance.

same within-sample ROC prediction curve resulting from the model in online Supplementary Table S2. Mean AUC for the permuted labels was 0.60, compared to 0.71 using the correct labels. Of central importance in this test is whether the 0.71 falls within the range of AUCs, we would expect by chance, where the range is defined by the 95% confidence interval. The upper bound of the confidence interval is 0.69, meaning the within-sample AUC value of 0.71 indicates prediction significantly better than chance.

PNC-based risk calculator

Figure 2 shows a comparison of the nine pipelines tested here, by AUC. Quality of prediction algorithms was clear, with ridge being the best (three leftmost bars in Fig. 2), followed by random forest, followed by SVMs. Quality of selection algorithms was more variable, with each of the three demonstrating best performance, depending on the algorithm: random forest selection is best for ridge, lasso selection is best for random forest, and STIR selection is best for SVMs. The key result is that the best cross-validated AUC was achieved using random forest selection, followed by ridge regression for prediction. We also examined the balance of sensitivity and specificity achieved by the pipelines in Fig. 2, shown in online Supplementary Figure S2. In all but one pipeline (RF → SVM), sensitivity is prioritized over specificity, and this is especially true for the models using the RF predictor (middle three sets of bars in online Supplementary Figure S2).

Figure 3 shows the frequency of feature-selection across the three algorithms (plus the mean), ordered by decreasing apparent importance overall. The top three most important, on average, were the C-GAS, PS-R item 2 ('I think that I might be able to predict the future'), and PR-R item 3 (something interrupting or controlling thoughts/actions). Regarding agreement among the three algorithms, the top 10 variables (on average) were in the top 1% of all three algorithms, suggesting substantial agreement, at least at the high importance level. Some notable exceptions include, (1) emotion identification performance was considered extremely important by STIR and random forest but only moderately so

Table 2. Final psychosis risk calculator using ridge regression and the top six predictors

Predictor	Coef.
Intercept	2.590
Children's Global Assessment Scale (CGAS)	-0.036
SOPS N4 Experience of emotions and self	0.242
PS-R 2: I think that I might be able to predict the future	0.149
PS-R 3: could possibly be something controlling my thoughts or actions	0.130
PS-R 12: I have been concerned that I might be 'going crazy'	0.169
Percent married in neighborhood	-1.882

Note. SOPS = Scale of Prodromal Symptoms; PS-R (aka 'PRIME') = Prevention through Risk Identification, Management, and Education Screen-Revised; Coef = coefficient; SCR = screen; final result will be in log-odds units, which can be converted to probability by exponentiating (to convert from log-odds to odds) and then using the equation $\text{probability} = \text{odds}/(\text{odds} + 1)$; *p* values and standard errors are not given because they are not meaningful for this type of model (Goeman, Meijer, & Chaturvedi, 2018) due to downwardly biased coefficients (typical rules of general linear model, where the equation is the best linear unbiased estimator, do not apply).

by lasso, (2) working memory performance was considered extremely important by STIR and random forest but not important at all by lasso, and (3) currently, taking psychoactive medications was considered extremely important by STIR and lasso but not important at all by random forest. Breaking the top ten variables down into 'types', five are clinical, three are cognitive, and two (trauma and percent married in neighborhood) are related to the environment or external experiences. Notably, none of the demographic characteristics was in the top 10; race was considered important (almost top 10), and age and sex were considered only moderately important.

One problem with the optimal results from Table 2 (STIR → ridge) is that the average number of variables selected (47.8), on average, was far too many for a risk calculator meant to be used by the public. We, therefore, opted to run a secondary

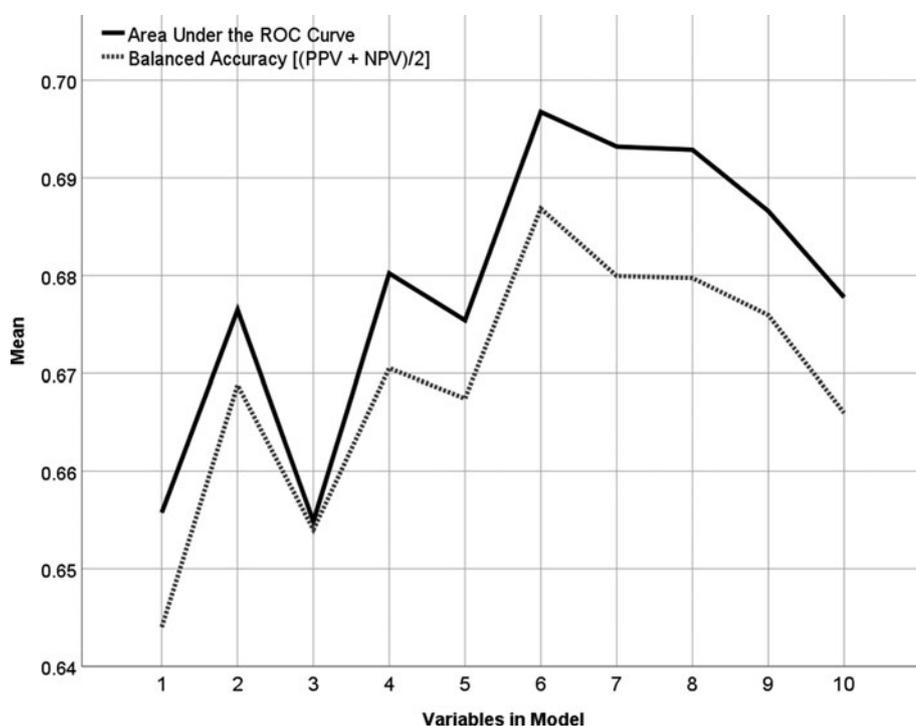


Fig. 4. AUC and balanced accuracy achieved by increasing numbers of variable.

analysis in which we tested the cross-validated prediction performance of increasing numbers of suggested variables. One sensible approach – i.e. to use the importance ranking provided by STIR – was not possible, because too many variables (21) were given the highest possible importance rating by STIR (i.e. selected on all 10 000 runs). To break the 21-way tie, we opted to use the average importance (black line in Fig. 3) in selecting the sequence of variables to add. Note that the final model comprised variables with maximum random forest importance anyway.

Figure 4 shows the cross-validated prediction results using ridge regression and the ‘top x’ variables according to average selection frequency in Fig. 3. With only one variable (C-GAS), the model achieves a CV AUC of almost 0.66. Adding PRIME_2 increases the AUC to almost 0.68, and adding PRIME_3 (for three variables total) brings the CV AUC back down to ~0.66. Adding three more variables (percent married in neighborhood, PRIME_12, and SIPS Perception of Self) brings the CV AUC to its maximum of almost 0.70. The final proposed risk calculation model therefore comprised C-GAS, PRIME_2, PRIME_3, percent married in neighborhood, PRIME_12, and SIPS Perception of Self/Others.

Table 2 shows the coefficients associated with this model. To facilitate use by future researchers, the coefficients in Table 2 are in raw native units (e.g. PS-R responses are on their usual 0–6 scale, C-GAS is out of 100, etc.). Increased risk of psychosis is indicated by low C-GAS; residence in a neighborhood where most people are unmarried; and endorsement of clinical symptoms related to predicting the future, having one’s thoughts controlled, concerns about going crazy, or changes in the experience of self/others.

Discussion

We performed a quasi-replication of a previously developed risk calculator for the transition from PS status to threshold psychosis, and then developed a new calculator for prediction of PS-risk status in a community sample.

Replication of the Cannon et al. (2016) calculator was successful insofar as the within-sample prognostic performance of the calculator was comparable across the two studies. Cross-validation of the results revealed predictive performance (AUC = 0.64) below what is traditionally considered adequate (AUC = 0.70), though all results for this replication should be interpreted with caution. First, there was not an exact match of variables used in the original calculator (Cannon et al., 2016). For example, we could include here only a broad index of global function (C-GAS), which conflates clinical symptoms and multiple domains of function, whereas NAPLS-2 utilized recent decline in social function assessed with the Global Function Social Scale, which differentiates social function from clinical symptoms and other aspects of functioning (Cornblatt et al., 2012). Second, this was a highly unbalanced sample with <10% cases (converters), meaning one could achieve >90% accuracy simply by predicting that no one will convert. This makes the 74% accuracy of the Cannon model seem unacceptably low, but this phenomenon in highly unbalanced samples will confound most available risk calculators with AUCs <0.80. Also, accuracy is not always the primary objective – e.g. the accuracy-maximizing prediction (mentioned above) that no one will convert would be useless in medicine. Finally, the coefficients in the NAPLS-2 model were re-estimated in this new sample, making this study only a quasi-replication focused on construct validity of the calculator. Despite these limitations, our findings appear to support the generalizability of the risk calculator approach in a broader PS community-based cohort.

Development of a new calculator for risk of future PS status (i.e. risk of being at high risk) revealed numerous important predictors of risk and achieved a cross-validated AUC (0.70 rounded up) that was minimally acceptable by contemporary standards. However, there is some information leakage (Boehmke & Greenwell, 2020) caused by the fact that the six variables in the final model were chosen based on importance across multiple algorithms across enough random cross-validations that information used for feature-selection ultimately came from the full sample. Therefore, a more

conservative estimate of the success of the present risk calculator would be to use the max number in Fig. 2, which is ~0.68. Additionally, a critical feature of the risk calculator presented here is that, unlike prior risk calculators, although we did perform our analyses in a community sample enriched for PS symptoms, the risk calculator was not developed on a clinically help seeking sample, characterized by distress and treatment seeking. Thus, the predictive utility of the calculator must be balanced with the potential stigma and anxiety associated with a risk label (Rüsch et al., 2015; Yang et al., 2015). Notably, the model prioritized environment (percent married in neighborhood) over race, suggesting the possibility that, (1) the actual proportion of people married in a neighborhood contains information all the way across the spectrum rather than simply being a proxy for race, and (2) one's environment is at least as important as one's race in determining psychosis risk.

Despite caveats mentioned above, the tool presented here (web link in Supplement) predicts a broader range of the PS continuum than in clinical high-risk samples, which is an advantage since psychosis can originate outside CHR (Lee, Lee, Kim, Choe, & Kwon, 2018). That is, most risk calculators (including Cannon et al., 2016) focus on conversion to frank psychosis, meaning differentiating among people at lower levels of risk (e.g. the difference between someone who responds to PRIME item 1 with a '0' v. someone who responds with a '2') is not a priority. The calculator presented here focuses on assessing risk along the full PS rather than at the moderate-extreme level where one typically sees conversion to frank psychosis. Prediction of PS status in this manner could be useful in recruiting for prospective community cohorts, where predicting that individuals are likely to experience persisting or worsening PS symptoms in the future might be more desirable than predicting likely threshold psychosis in the same time frame. In particular, the risk calculator presented here is applicable for use in a younger cohort of individuals (mean age 15), where conversion to threshold psychosis within only a few years is relatively rarer than in most CHR samples who are, on average, in the late adolescence early adult age range (e.g. Cannon et al., 2016; Osborne & Mittal, 2019; Zhang et al., 2018). This risk strategy may be useful in several ways. First, given that persisting subthreshold psychosis symptoms are associated with increased risk of comorbid psychopathology, including mood, anxiety, substance and suicidal ideation, as well as poor global function (see Taylor et al., 2020 for review), risk prediction can facilitate screening and earlier access to mental health care. In addition to providing referral for relief for current symptoms, screening could lead to improved PS symptom monitoring, resilience-building strategies, and, perhaps, prevention efforts. Second, the approach can facilitate prospective studies aiming to elucidate and characterize biobehavioral and functional features of early developmental trajectories of PS symptoms. Such efforts can potentially facilitate a precision medicine approach by establishing mechanistic links among cellular-molecular aberrations and PS symptoms in the general population.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291720005231>.

Acknowledgements. This work was supported by the Lifespan Brain Institute (LiBI); NIMH grants MH089983, MH107235, MH081902, MH117014, MH09689, MH103654; and the Dowshen Neuroscience Fund.

Conflict of interest. All authors report no conflicts of interest.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and

institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

References

- Adbi, A., Sadatsafavi, M., & Ioannidis, J. P. (2020). Validation and utility testing of clinical prediction models: Time to change the approach. *JAMA*, *324*(3), 235–236.
- Bilimoria, K. Y., Liu, Y., Paruch, J. L., Zhou, L., Kmieciak, T. E., Ko, C. Y., & Cohen, M. E. (2013). Development and evaluation of the universal ACS NSQIP surgical risk calculator: A decision aid and informed consent tool for patients and surgeons. *Journal of the American College of Surgeons*, *217*(5), 833–842.
- Boehmke, B., & Greenwell, B. M. (2020). *Hands-On machine learning with R*. Boca Raton, FL: CRC Press.
- Brown, J. L. (1963). Follow-up of children with atypical development (infantile psychosis). *American Journal of Orthopsychiatry*, *33*(5), 855.
- Calkins, M. E., Merikangas, K. R., Moore, T. M., Burstein, M., Behr, M. A., Satterthwaite, T. D., ... Gur, R. E. (2015). The Philadelphia neurodevelopmental cohort: Constructing a deep phenotyping collaborative. *Journal of Child Psychology and Psychiatry*, *56*(12), 1356–1369. doi:10.1111/jcpp.12416.
- Calkins, M. E., Moore, T. M., Merikangas, K. R., Burstein, M., Satterthwaite, T. D., Bilker, W. B., ... Gur, R. E. (2014). The psychosis spectrum in a young U.S. Community sample: Findings from the Philadelphia neurodevelopmental cohort. *World Psychiatry*, *13*(3), 296–305. doi:10.1002/wps.20152.
- Calkins, M. E., Moore, T. M., Satterthwaite, T. D., Wolf, D. H., Turetsky, B. I., Roalf, D. R., ... Gur, R. C. (2017). Persistence of psychosis spectrum symptoms in the Philadelphia neurodevelopmental cohort: A prospective two-year follow-up. *World Psychiatry*, *16*, 62–76.
- Cannon, T. D., Yu, C., Addington, J., Bearden, C. E., Cadenhead, K. S., Cornblatt, B. A., ... Perkins, D. O. (2016). An individualized risk calculator for research in prodromal psychosis. *American Journal of Psychiatry*, *173*(10), 980–988.
- Carrion, R. E., Cornblatt, B. A., Burton, C. Z., Tso, I. F., Auther, A. M., Adelsheim, S., ... Taylor, S. F. (2016). Personalized prediction of psychosis: External validation of the NAPLS-2 psychosis risk calculator with the EDIPPP project. *American Journal of Psychiatry*, *173*(10), 989–996.
- Coiera, E., Ammenwerth, E., Georgiou, A., & Magrabi, F. (2018). Does health informatics have a replication crisis? *Journal of the American Medical Informatics Association*, *25*(8), 963–968.
- Combe, G., Donkin, B., Buchanan, R., & Mackenzie, G. S. (1820). On inferring natural dispositions and talents from development of brain. *Transactions of the Phrenological Society*, *9*, 306–379.
- Cornblatt, B. A., Carrion, R. E., Addington, J., Seidman, L., Walker, E. F., Cannon, T. D., ... Lencz, T. (2012). Risk factors for psychosis: Impaired social and role functioning. *Schizophrenia Bulletin*, *38*(6), 1247–1257.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*(4), 281.
- Davies, J., Sullivan, S., & Zammit, S. (2018). Adverse life outcomes associated with adolescent psychotic experiences and depressive symptoms. *Society of Psychiatry and Psychiatric Epidemiology*, *53*(5), 497–507.
- Dawber, T. R., Moore, F. E., & Mann, G. V. (1957). Coronary heart disease in the Framingham study. *American Journal of Public Health and the Nation's Health*, *47*, 4–24.
- Friedman, J. I., Harvey, P. D., Coleman, T., Moriarty, P. J., Bowie, C., Parrella, M., ... Davis, K. L. (2001). Six-year follow-up study of cognitive and functional status across the lifespan in schizophrenia: A comparison with Alzheimer's disease and normal aging. *American Journal of Psychiatry*, *158*(9), 1441–1448.
- Fusar-Poli, P., Oliver, D., Spada, G., Patel, R., Stewart, R., Dobson, R., & McGuire, P. (2019). Real-world implementation of a transdiagnostic risk calculator for the automatic detection of individuals at risk of psychosis in clinical routine: Study protocol. *Frontiers in Psychiatry*, *10*, 109.
- Fusar-Poli, P., Rutigliano, G., Stahl, D., Davies, C., Bonoldi, I., Reilly, T., & McGuire, P. (2017). Development and validation of a clinically based risk calculator for the transdiagnostic prediction of psychosis. *JAMA Psychiatry*, *74*(5), 493–500.
- Fusar-Poli, P., Schultze-Lutter, F., Cappucciati, M., Rutigliano, G., Bonoldi, I., Stahl, D., ... Woods, S. W. (2016). The dark side of the moon: Meta-analytical impact of recruitment strategies on risk enrichment in the clinical high risk state for psychosis. *Schizophrenia Bulletin*, *42*(3), 732–743.

- Goeman, J., Meijer, R., & Chaturvedi, N. (2018). L1 and L2 penalized regression models. Vignette of R package *penalized*. Retrieved from <https://cran.r-project.org/web/packages/penalized/penalized.pdf>.
- Gupta, P. K., Gupta, H., Sundaram, A., Kaushik, M., Fang, X., Miller, W. J., ... Lynch, T. G. (2011). Development and validation of a risk calculator for prediction of cardiac risk after surgery. *Circulation*, *124*(4), 381–387.
- Gur, R. C., Ragland, J. D., Moberg, P. J., Turner, T. H., Bilker, W. B., Kohler, C., ... Gur, R. E. (2001). Computerized neurocognitive scanning: I. Methodology and validation in healthy people. *Neuropsychopharmacology*, *25*, 766–776.
- Gur, R. C., Richard, J., Huggett, P., Calkins, M. E., Macy, L., Bilker, W. B., ... Gur, R. E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods*, *187*, 254–262.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36.
- Heikes, K. E., Eddy, D. M., Arondekar, B., & Schlessinger, L. (2008). Diabetes risk calculator: A simple tool for detecting undiagnosed diabetes and pre-diabetes. *Diabetes Care*, *31*(5), 1040–1045.
- Kalman, J. L., Bresnahan, M., Schulze, T. G., & Susser, E. (2019). Predictors of persisting psychotic like experiences in children and adolescents: A scoping review. *Schizophrenia Research*, *209*, 32–39.
- Kaufman, J., Birmaher, B., & Brent, D., Rao, U., Flynn, C., Moreci, P., ... Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children – present and lifetime version (K-SADS-PL): Initial reliability and validity. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*, 980–988.
- Keith, S. J., & Matthews, S. M. (1991). The diagnosis of schizophrenia: A review of onset and duration issues. *Schizophrenia Bulletin*, *17*(1), 51–68.
- Kobayashi, H., Nemoto, T., Koshikawa, H., Osono, Y., Yamazawa, R., Murakami, M., ... Mizuno, M. (2008). A self-reported instrument for prodromal symptoms of psychosis: Testing the clinical validity of the PRIME screen—revised (PS-R) in a Japanese population. *Schizophrenia Research*, *106*(2–3), 356–362.
- Le, T. T., Urbanowicz, R. J., Moore, J. H., & McKinney, B. A. (2019). Statistical inference relief (STIR) feature selection. *Bioinformatics (Oxford, England)*, *35*(8), 1358–1365.
- Lee, T. Y., Lee, J., Kim, M., Choe, E., & Kwon, J. S. (2018). Can we predict psychosis outside the clinical high-risk state? A systematic review of non-psychotic risk syndromes for mental disorders. *Schizophrenia Bulletin*, *44*(2), 276–285.
- Leslie, W. D., & Lix, L. M. (2014). Comparison between various fracture risk assessment tools. *Osteoporosis International*, *25*(1), 1–21.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.
- Mahler, M. S. (1952). On child psychosis and schizophrenia: Autistic and symbiotic infantile psychoses. *The Psychoanalytic Study of the Child*, *7*(1), 286–305.
- Mahmood, S. S., Levy, D., Vasan, R. S., & Wang, T. J. (2014). The Framingham Heart Study and the epidemiology of cardiovascular disease: A historical perspective. *The Lancet*, *383*(9921), 999–1008.
- Maxwell, M. E. (1996). *Manual for the FIGS*. Bethesda: Clinical Neurogenetics Branch, Intramural Research Program, National Institute for Mental Health.
- McGlashan, T. H., Miller, T. J., & Woods, S. W. (2003). *Structured interview for prodromal syndromes, version 4.0*. New Haven: Prime Clinic Yale School of Medicine.
- McNeish, D. M. (2015). Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research*, *50*(5), 471–484.
- Miller, T. J., Cicchetti, D., & Markovich, P. J. (2004). The SIPS screen: A brief self-report screen to detect the schizophrenia prodrome. *Schizophrenia Research*, *70*(Suppl. 1), 78.
- Miller, T. J., McGlashan, T. H., Woods, S. W., Stein, K., Driesen, N., Corcoran, C. M., ... Davidson, L. (1999). Symptom assessment in schizophrenic prodromal states. *Psychiatric Quarterly*, *70*(4), 273–287.
- Moore, T. M., Martin, I. K., Gur, O. M., Jackson, C. T., Scott, J. C., Calkins, M. E., ... Gur, R. E. (2016). Characterizing social environment's association with neurocognition using census and crime data linked to the Philadelphia neurodevelopmental cohort. *Psychological Medicine*, *46*(3), 599–610.
- Moore, T. M., Reise, S. P., Gur, R. E., Hakonarson, H., & Gur, R. C. (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology*, *29*(2), 235.
- Nelson, B., & McGorry, P. (2020). The prodrome of psychotic disorders: Identification, prediction, and preventive treatment. *Child and Adolescent Psychiatric Clinics*, *29*(1), 57–69.
- Osborne, K. J., & Mittal, V. A. (2019). External validation and extension of the NAPLS-2 and SIPS-RC personalized risk calculators in an independent clinical high-risk sample. *Psychiatry Research*, *279*, 9–14.
- Page, R. C., Krall, E. A., Martin, J., Mancl, L., & Garcia, R. I. (2002). Validity and accuracy of a risk calculator in predicting periodontal disease. *The Journal of the American Dental Association*, *133*(5), 569–576.
- Parikh, P., Shiloach, M., Cohen, M. E., Bilimoria, K. Y., Ko, C. Y., Hall, B. L., & Pitt, H. A. (2010). Pancreatectomy risk calculator: An ACS-NSQIP resource. *Hpb*, *12*(7), 488–497.
- Powers, A. R., Addington, J., Perkins, D. O., Bearden, C. E., Cadenhead, K. S., Cannon, T. D., ... Walker, E. F. (2020). Duration of the psychosis prodrome. *Schizophrenia Research*, *216*, 443–449.
- Radua, J., Ramella-Cravaro, V., Ioannidis, J. P., Reichenberg, A., Phiphophthasane, N., Amir, T., ... Fusar-Poli, P. (2018). What causes psychosis? An umbrella review of risk and protective factors. *World Psychiatry*, *17*(1), 49–66.
- Riecher-Rössler, A., & Studerus, E. (2017). Prediction of conversion to psychosis in individuals with an at-risk mental state: A brief update on recent developments. *Current Opinion in Psychiatry*, *30*(3), 209–219.
- Rüsch, N., Heekeren, K., Theodoridou, A., Müller, M., Corrigan, P. W., Mayer, B., ... Rössler, W. (2015). Stigma as a stressor and transition to schizophrenia after one year among young people at risk of psychosis. *Schizophrenia Research*, *166*(1–3), 43–48.
- Sanfelici, R., Dwyer, D. B., Antonucci, L. A., & Koutsouleris, N. (2020). Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: A meta-analytic view on the state-of-the-art. *Biological Psychiatry*, *88*(4), 349–360.
- Satterthwaite, T. D., Elliott, M. A., Ruparel, K., Loughhead, J., Prabhakaran, K., & Calkins, M. E., ... Gur, R. E. (2014). Neuroimaging of the Philadelphia neurodevelopmental cohort. *NeuroImage*, *86*, 544–553.
- Shaffer, D., Gould, M. S., Brasic, J., Ambrosini, P., Fisher, P., Bird, H., & Aluwahlia, S. (1983). A children's global assessment scale (CGAS). *Archives of General Psychiatry*, *40*(11), 1228–1231.
- Taylor, J. H., Asabere, N., Calkins, M. E., Moore, T. M., Tang, S. X., Xavier, R. M., ... Gur, R. E. (2020). Characteristics of youth with reported family history of psychosis spectrum symptoms in the Philadelphia neurodevelopmental cohort. *Schizophrenia Research*, *216*, 104–110.
- Taylor, J. H., Calkins, M. E., & Gur, R. E. (2020). Markers of psychosis risk in the general population. *Biological Psychiatry*, *88*(4), 337–348.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.
- Wigman, J. T., van Winkel, R., Raaijmakers, Q. A., Ormel, J., Verhulst, F. C., Reijneveld, S. A., ... Vollebregt, W. A. (2011). Evidence for a persistent, environment-dependent and deteriorating subtype of subclinical psychotic experiences: A 6-year longitudinal general population study. *Psychological Medicine*, *41*(11), 2317–2329.
- Woodberry, K. A., Shapiro, D. I., Bryant, C., & Seidman, L. J. (2016). Progress and future directions in research on the psychosis prodrome: A review for clinicians. *Harvard Review of Psychiatry*, *24*(2), 87.
- Yang, L. H., Link, B. G., Ben-David, S., Gill, K. E., Girgis, R. R., Brucato, G., ... Corcoran, C. M. (2015). Stigma related to labels and symptoms in individuals at clinical high-risk for psychosis. *Schizophrenia Research*, *168*(1–2), 9–15.
- Yung, A. R., & McGorry, P. D. (1996). The prodromal phase of first-episode psychosis: Past and current conceptualizations. *Schizophrenia Bulletin*, *22*(2), 353–370.
- Yung, A. R., Phillips, L. J., Yuen, H. P., Francey, S. M., McFarlane, C. A., Hallgren, M., & McGorry, P. D. (2003). Psychosis prediction: 12-month follow up of a high-risk (“prodromal”) group. *Schizophrenia Research*, *60*(1), 21–32.
- Zhang, T., Li, H., Tang, Y., Niznikiewicz, M. A., Shenton, M. E., Keshavan, M. S., ... Wang, J. (2018). Validating the predictive accuracy of the NAPLS-2 psychosis risk calculator in a clinical high-risk sample from the SHARP (Shanghai At risk for psychosis) program. *American Journal of Psychiatry*, *175*(9), 906–908.