Original Article

CAMBRIDGE
UNIVERSITY PRESS

Predictive validity of psychosis risk models when applied to adolescent psychiatric patients

Maija Lindgren[1] [iD], Heidi Kuvaja[2], Markus Jokela[2] and Sebastian Therman[1] [iD]

[1]Mental Health, Finnish Institute for Health and Welfare, Helsinki, Finland and [2]Department of Psychology and Logopedics, Faculty of Medicine, Helsinki University, Helsinki, Finland

Abstract

Background. Several multivariate algorithms have been developed for predicting psychosis, as attempts to obtain better prognosis prediction than with current clinical high-risk (CHR) criteria. The models have typically been based on samples from specialized clinics. We evaluated the generalizability of 19 prediction models to clinical practice in an unselected adolescent psychiatric sample.

Methods. In total, 153 adolescent psychiatric patients in the Helsinki Prodromal Study underwent an extensive baseline assessment including the SIPS interview and a neurocognitive battery, with 50 participants (33%) fulfilling CHR criteria. The adolescents were followed up for 7 years using comprehensive national registers. Assessed outcomes were (1) any psychotic disorder diagnosis ($n = 18$, 12%) and (2) first psychiatric hospitalization ($n = 25$, 16%) as an index of overall deterioration of functioning.

Results. Most models improved the overall prediction accuracy over standard CHR criteria (area under the curve estimates ranging between 0.51 and 0.82), although the accuracy was worse than that in the samples used to develop the models, also when applied only to the CHR subsample. The best models for transition to psychosis included the severity of positive symptoms, especially delusions, and negative symptoms. Exploratory models revealed baseline negative symptoms, low functioning, delusions, and sleep problems in combination to be the best predictor of psychiatric hospitalization in the upcoming years.

Conclusions. Including the severity levels of both positive and negative symptomatology proved beneficial in predicting psychosis. Despite these advances, the applicability of extended psychosis-risk models to general psychiatric practice appears limited.

Background

Psychosis risk has been extensively studied during the last few decades, with the aim of identifying people who are in the prodromal period for psychosis. Psychosis transition rates vary greatly among studies, with the validity of psychosis risk assessment depending on the population in which it is studied (Fusar-Poli et al., 2016b). Most psychosis risk research has been focused on help-seeking patients in specialized psychosis risk care, fulfilling and enriched with respect to the criteria for clinical high-risk (CHR; Fusar-Poli et al., 2012; Schultze-Lutter et al., 2015). The predictive precision has been lower in more diverse populations, lowering the generalizability of findings from specialized clinics to general psychiatric practice (van Os & Guloksuz, 2017). In addition, non-CHR groups as psychiatric controls are often overlooked (Millman, Gold, Mittal, & Schiffman, 2019).

The most established methods of psychosis risk detection, Comprehensive Assessment of At-Risk Mental States (CAARMS; Yung et al., 2006) and the Structured Interview for Prodromal Syndromes (SIPS; Miller et al., 2003), assess various symptom domains. However, only positive symptoms are actually used for determining psychosis risk status, in addition to a combination of genetic risk and functional level, which are also components of CHR syndrome definitions. Some studies have investigated the predictiveness of symptoms from the other domains – negative and disorganized symptoms (Addington & Heinssen, 2012) – and various psychosis prediction models have been developed combining positive symptoms or CHR status with the other symptom dimensions (Studerus, Ramyead, & Riecher-Rössler, 2017). Since impairments in neurocognitive abilities might be connected to psychosis transition (Bora et al., 2014), these are also included in some models. Other commonly suggested predictors of psychosis include substance abuse, trauma exposure, and various biological markers (Studerus et al., 2017).

As fulfilling criteria for psychosis can be seen as an arbitrary line on the psychosis continuum, transition to a psychotic disorder as the sole outcome of interest has been criticized (van Os & Guloksuz, 2017), especially in the context of positive symptoms. Although positive

symptoms at the CHR level may be precursors of psychosis, especially milder positive symptoms may more often be unspecific risk markers expressing general psychiatric vulnerability (Healy et al., 2019; Jeppesen et al., 2015). Therefore, in addition to the psychosis outcome, the existing CHR assessment tools are also being used to predict unspecific outcomes such as everyday level of functioning or hospitalization – which they may predict better than psychosis (Carrión et al., 2013; Cotter et al., 2014; Lin et al., 2011). In previous prediction algorithms applied to young people at CHR, both psychosis transition and functioning outcomes have repeatedly been best predicted by suspicion/paranoia, delusions, and social functional decline (Worthington, Cao, & Cannon, 2019). Furthermore, we have previously found that, in a general adolescent psychiatric population, the ability of the dichotomous CHR status to predict psychosis was poor; positive symptom intensity predicted psychotic disorders, whereas CHR status predicted psychiatric hospitalization (Lindgren et al., 2014). We now used a longer follow-up and previously published prediction algorithms based on other samples to predict clinical outcomes in a confirmatory manner.

In the current study, the performance of some of the previously suggested psychosis prediction models was tested to determine the generalizability of previous findings to clinical practice (1) in a psychiatric sample consisting of both CHR and non-CHR adolescents, and (2) in the CHR subsample. In addition, the models were used to predict functional outcome, operationalized as first psychiatric hospitalization. Finally, we also fitted exploratory models to identify the best predictors in the current sample.

Methods

Participants and study procedure

The Helsinki Prodromal Study is a prospective psychosis risk study among adolescent in psychiatric care. All 15–18-year-olds entering any public psychiatric clinic or ward in the city of Helsinki during a 3-year period in 2003–2004 and 2007–2008 were invited to fill out the Finnish version of the Prodromal Questionnaire (PQ; Loewy, Bearden, Johnson, Raine, and Cannon, 2005) on their first or second appointment. A total of 819 of questionnaires (~75% of the eligible participants in psychiatric treatment) were completed (Therman et al., 2014). Those who had a current or previous psychotic disorder at baseline or were unable to communicate in Finnish were excluded from the sample.

The review boards of the Finnish Institute for Health and Welfare and the Ethics Committee of the Hospital District of Helsinki and Uusimaa approved the study procedure. The study was carried out in accordance with the sixth version of the Declaration of Helsinki (World Medical Association, 2000). The PQ was part of the standard assessment in the units. Those entering the in-depth assessment gave written informed consent.

In-depth assessment

All 145 adolescents who scored 18 or higher on the Positive Symptoms subscale of the PQ, indicating elevated psychosis risk (Loewy et al., 2005), were invited to the in-depth baseline assessment, together with 87 block-randomized adolescents who scored below this cut-off. Of these invited participants, 174 adolescents completed the in-depth assessment, which included extensive

neuropsychological testing (Lindgren et al., 2010), of which the current study utilizes raw scores of WAIS-III Matrix Reasoning (Wechsler, 1997), measuring non-verbal reasoning, as well as WMS-R Visual Reproduction I (Wechsler, 1987), measuring visual episodic memory. In addition, the adolescents taking part in the in-depth assessment were interviewed with the Finnish translation of the SIPS 3.0 (Miller et al., 2003). The SIPS also includes the modified Global Assessment of Functioning (GAF-M) scale, rated from 1 to 100 (Miller et al., 2003). Medical records preceding the psychiatric treatment were available for most interviewed participants (97.7%). The Global Functioning: Social scale (Cornblatt et al., 2007) was rated from 1 to 10 based on interview data and medical records. Substance abuse and duration of psychiatric symptoms before treatment contact were also rated based on all available information. All interview- and record-based ratings were made by two or more trained reviewers independently, and differences resolved by consensus.

Follow-up

Follow-up information was collected from the national Care Register for Health Care, which includes psychiatric treatment in any public outpatient or inpatient clinic or ward in Finland. As the last participants were enrolled in the year 2008 and the data on the use of psychiatric services were available until the end of the year 2015, a 7-year follow-up was possible for all participants. A psychosis outcome was defined as a psychotic disorder diagnosis of ICD-10 codes F20, F22–F29, F30.2, F31.2, F31.5, F32.3, or F33.3. The psychiatric hospitalization outcome was defined as staying in a psychiatric hospital or in any hospital with a primary psychiatric diagnosis during the follow-up period, excluding those with any such hospitalizations before or at the time of the baseline assessment.

Prediction models

The prediction models were selected from exhaustive reviews by Studerus et al. (2017) and Montemagni, Bellino, Bracale, Bozzatello, and Rocca (2020). They accepted studies that developed or validated a multivariable prediction model on psychosis transition among people estimated to be at high risk for psychosis, some of the studies proposing more than one prediction model. In the current study we did not test models from small samples ($N < 100$) or from studies finding no evidence supporting any tested model. As we were only able to test models with variables available in the Helsinki Prodromal Study, we included 19 prediction models (Table 1).

If the original model used CAARMS or Scale for the Assessment of Negative Symptoms (SANS) scales, we used the SIPS equivalent and thus, for example, CAARMS 'Unusual thought content' was substituted with SIPS scale 'P1 Unusual thought content and delusional ideas', and SANS 'Attention' was considered equivalent to SIPS scale 'D3 Trouble with focus and attention'.

Some models used an equation such as sum of predictor scores; these models were used both as an equation as well as using the predictors individually (Perkins et al., 2015; Ruhrmann et al., 2010; Thompson, Nelson, & Yung, 2011). Furthermore, the prediction model proposed by Ruhrmann et al. (2010) included a criterion of positive symptoms >16; the value of this parameter, however, was questionable in our sample, as only one person scored this high. The model was therefore also

Table 1. Predictors included in the previous psychosis prediction models and their equivalents in the current study

Model		Predictors in original study	Predictors in the current study
Auther		SIPS positive symptoms sum	SIPS positive symptoms sum
Buchy		Low use of alcohol	No alcohol abuse or dependency
Cannon	1. model (best 4 by HR)	GRD, unusual thought content (SIPS P1 >3), suspicion/paranoia (SIPS P2 >2), DSM-IV substance abuse	GRD, SIPS P1 >3, SIPS P2 >2, any substance-related DSM-IV diagnosis
	2. model (best 3 by HR)	Unusual thought content (SIPS P1 >3), global functioning: social scale <7, DSM-IV substance abuse	SIPS P1 >3, global functioning: social scale <7, any substance-related DSM-IV diagnosis
Demjaha		CAARMS negative symptoms factor, disorganization/cognitive symptoms factor	Factor score from confirmatory factor analysis of SIPS equivalents Factor score from confirmatory factor analysis of SIPS equivalents
DeVylder	1. model	Disorganized communication (SIPS P5 ⩾2)	SIPS P5 ⩾2
	2. model	Disorganized communication (SIPS P5 ⩾3)	SIPS P5 ⩾3
	3. model	Disorganized communication (SIPS P5 ⩾4)	SIPS P5 ⩾4
Lin	1. model	GAF, WMS-R Visual Reproduction I	GAF, WMS-R Visual Reproduction I
	2. model	CAARMS thought content, WASI Matrix Reasoning	Unusual thought content (SIPS P1), WAIS-III Matrix Reasoning
Nelson (Nelson et al., 2013)		GAF <44, duration of symptoms >2 years	GAF <44, duration of untreated symptoms >2 years
Perkins		Unusual thought content (SIPS P1), suspicion/paranoia (SIPS P2), decreased ideational richness (SIPS N5), attention (SIPS D3)	SIPS P1, SIPS P2, SIPS N5, SIPS D3
Piskulic		SIPS negative symptoms sum	SIPS negative symptoms sum
Raballo		CAARMS disorganization symptom factor	Factor score from confirmatory factor analysis of SIPS equivalents
Ruhrmann[a]		SIPS positive symptoms sum >16, bizarre thinking (SIPS D2 >2), sleep disturbance (SIPS G1 >2), SIPS schizotypal personality disorder, GAF 12 month maximum, years of education	SIPS positive symptoms sum >16, SIPS D2 >2, SIPS G1 >2, SIPS schizotypal personality features, GAF 12 month maximum
Thompson		At least two of the following: • CAARMS unusual thought content >3 • GAF <50 • CAARMS criteria of Genetic risk with functional decline	At least two of the following: • SIPS P1 >3 • GAF <50 • GRD
Velthorst[b]		GAF <61, BPRS psychotic subscale, SANS negative total score	GAF<61, SIPS positive symptoms sum, SIPS negative symptoms sum
Walder[c]		PAS social functioning, SIPS positive symptoms average	Global functioning: social scale, SIPS positive symptoms sum
Yung (Yung et al., 2004)		Both APS and GRD, duration of symptoms >5 years, GAF <40, attention (SANS attention >2)	Both APS and GRD, duration of untreated symptoms >5 years, GAF <40, attention (SIPS D3 >2)

[a]In the Ruhrmann model, education was left out because of lack of variance in the current adolescent sample. Schizotypal personality disorder criteria were not met anyone in the sample, but a cutoff of at least three out of nine schizotypal personality features was used instead (11 adolescents meeting this criterion). The model included a criterion of sum of positive symptoms >16, but as only one person fulfilled this criterion, the model was run with this predictor replaced by the SIPS positive symptoms sum (used in other models) and these results are presented as 'Ruhrmann modified'.
[b]In the Velthorst model, GAF < 61 predicted psychosis in the opposite direction than was meant to, so also a 'Velthorst modified' model was calculated without this predictor.
[c]The Walder model originated from a male sample, but here it was used for both genders.

estimated with this dichotomous predictor replaced by the SIPS positive symptoms sum, presented separately as a modified model. The models used by Velthorst et al. (2013) used both baseline functioning level and decline in functioning level over the follow-up, controlling for symptoms, however, we only tested the model with baseline functioning and symptom levels, as functioning decline at the outcome time point is confounded with the outcome itself. We also estimated the model without the general functioning variable (Velthorst modified model), due to that variable when estimated having an opposite sign to that of the

original publication. In addition, information on alcohol use in our sample was limited to alcohol use and dependency, and we coded their absence as 'low alcohol use' for testing the model by Buchy, Perkins, Woods, Liu, and Addington (2014), which used a wider spectrum of severity of alcohol use. For two models the required factor score variables were estimated with confirmatory factor analysis (online Supplement S1). As Cannon et al. (2008) had published a wide variety of models without stated preference, we selected the best combinations of both three and four predictors, as measured by the published hazard ratios.

Statistical analysis

The confirmatory factor analyses of two models (Demjaha, Valmaggia, Stahl, Byrne, & McGuire, 2012; Raballo, Nelson, Thompson, & Yung, 2011) were fitted using Mplus 8.3 (Muthén & Muthén, 2017), as described in detail in online Supplement S1.

The performance of each previously proposed model in predicting psychosis transitions and first lifetime hospitalization was tested with penalized (Firth) logistic regression using the *logistf* package (version 1.24; Heinze and Ploner, 2018) in R (version 4.0.4; R Core Team, 2021), with standardized predictors. This method is suitable for small samples. The original coefficients reported by the previous studies were used to calculate the risk predictions when this information was available. In almost all cases, however, the original study had not reported the individual coefficients, and we estimated coefficients based on the current data. To make model comparison fair, we additionally estimated coefficients using the variables of the one model (Ruhrmann et al., 2010) for which we had the original coefficients. For comparison purposes, we also tested how well CHR status alone (as assessed by the SIPS) predicted the outcomes, and how well previous hospitalization alone predicted rehospitalization during follow-up (in this analysis, those with previous hospitalizations were not excluded, in contrast to all other hospitalization analyses). All regression models using the whole sample were weighted to compensate for participation rate and undersampling of questionnaire screen-negatives and males (online Supplement S2; Lindgren et al., 2014). As all the models have been developed within CHR samples, we repeated the analyses in the subsample of CHR participants.

The logarithmic odds ratio is presented for each unstandardized predictor. Akaike information criterion (AIC) values are reported to compare the models, representing the models' quality in the current sample, with a lower AIC value indicating a better fit. The AIC was chosen over the Bayesian information criterion on the recommendation of Vrieze (2012). In addition, the discriminative ability of models was quantified with the weighted area under the curve (AUC) statistics, as calculated with the R package *PRROC* (version 1.3.1; Grau, Grosse, and Keilwagen, 2015), and secondarily with Cohen's κ coefficient, obtained with the *cohen.kappa* function of the R package *psych* (version 2.0.12; Revelle, 2021), as presented in the online Supplementary workbook.

For diagnostic performance assessment, cut-offs for the predicted values of the logistic regressions were set to obtain a fraction of test-positives equal to or higher than the outcome prevalence at follow-up (prevalence matching), with the exception of models where this would have resulted in no screen-negatives, in which case we chose the value corresponding to the closest lower available prevalence with that combination of variables. Setting the cut-off in this way without *post-hoc* data-driven optimization reduces bias (Ewald, 2006) and makes models more comparable. Net benefit or relative utility were considered outside the scope of this paper. Calculations based on the confusion matrix were made both with and without the Haldane correction (0.5 observations added to each cell), and with and without sample weights, with the corrected and weighted results considered primary (Brown & White, 2005). Diagnostic odds ratios (DORs) are reported as the primary single indicator of diagnostic performance, with distribution separation (Cohen's *d*) and the phi coefficient r_φ as secondary indicators.

Finally, we performed exploratory analyses using all the predictors of the proposed models (online Supplement S3), with the exception of the two cognitive variables, as including them would have reduced the number of participants to unacceptably low for the exploratory modeling. As these analyses required complete data, there were 141 cases available for the psychosis outcome and 120 for the hospitalization outcome. As recommended by Studerus et al. (2017), we employed LASSO for these exploratory analyses. Specifically, we used the function *glmnet* of the R package *glmnet* (version 4.1-1; Friedman, Hastie, and Tibshirani, 2010), fitting weighted logistic regression models with default settings, allowing up to four coefficients, which was the maximum used in the previously published models, with the exception of the five-parameter Ruhrmann et al. model (2010). All analysis parameters were set to their defaults.

Results

The sample consisted of 153 adolescents who were starting psychiatric treatment (Fig. 1), most often with a mood disorder diagnosis. Of these participants, 120 (78.4%) were girls (which is representative of the gender distribution of patients in these units). The mean age was 16.5 years (range 15–18), and 10 (6.5%) were inpatients at baseline, the rest being outpatients. Cognitive data were available for 146 adolescents and this smaller sample was used when testing models with cognitive variables.

During the 7-year follow-up, 18 (11.8%) adolescents were diagnosed with a psychotic disorder (11 of whom were CHR at baseline) and 25 (16.3%) had entered their first psychiatric hospital care. We predicted these outcomes with separate models. We then applied the same models to the CHR subsample of 50 adolescents (of whom 40, 80.0% were girls), predicting the 11 psychotic disorders and 11 first psychiatric hospitalizations among them.

Psychosis transition

Most of the models predicted psychosis better than the CHR status alone (Table 2). There were considerable differences in predictive ability between the models, with one of the best performing models using the positive symptom sum score only (Auther et al., 2012). The sample-fitted Velthorst model had good AIC and AUC values but having poor general functioning (GAF < 61) predicted *not developing* psychosis, that is, in the opposite direction of the original model. The modified Velthorst model without the GAF predictor (using merely sums of positive and negative symptoms) was still the best performing model. Other models with good AIC values consisted of the positive symptom sum score and the social functioning scale (Walder et al., 2013) and the negative symptoms sum score (Piskulic et al., 2012), respectively.

The discriminability order of the prediction models differed somewhat from that of the model fit, with a natural advantage for multi-predictor models. Discriminative ability also varied considerably between the models. The highest AUC estimates between 0.71 and 0.75 were observed for models that used the presence of marked positive symptoms, bizarre thinking, sleep disturbance, and schizotypal personality, along with the general functioning score (Ruhrmann et al., 2010); intensity of delusions, suspiciousness, decreased ideational richness, and trouble with attention (Perkins et al., 2015); or the negative symptoms sum score (Piskulic et al., 2012). The original and modified Velthorst models (2013) were also among the models with best AUC values. Figure 2 shows the discrepancy between AIC and
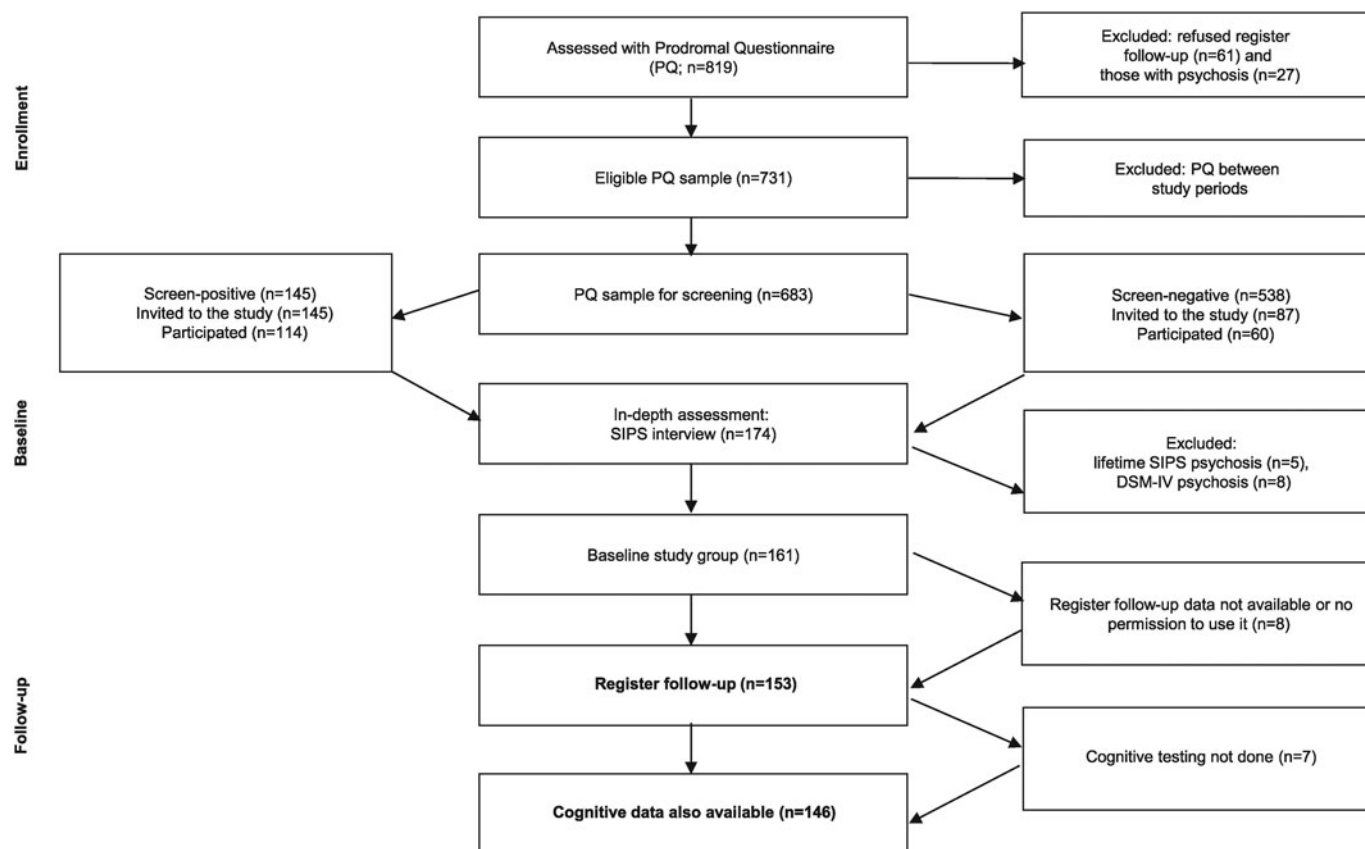
Fig. 1. Forming of the sample.

AUC values of the models in predicting psychosis, using the whole sample.

When the prevalence-matched cut-offs were applied, the most effective classifier, the Thompson et al. (2011) single-criterion model, reached a DOR of 10, with a sensitivity of 26% and a specificity of 97%. For comparison, the geometric mean of the other models' DORs was 4.8, and the SIPS CHR criterion had a DOR of 3.7 (online Supplementary workbook).

The models performed a little better in the CHR subsample than in the unselected psychiatric sample. In the CHR subsample, the two best models had both the best model fit and high discriminability (AUC between 0.81 and 0.83), using the disorganization symptoms latent factor (Raballo et al., 2011) and delusions, social functioning, and substance abuse (Cannon et al., 2008), respectively. In addition, among the best models were those consisting of the sums of positive and negative symptoms (modification of Velthorst et al., 2013), delusion intensity and visual reasoning score (Lin et al., 2013), or positive symptoms, bizarre thinking, sleep disturbance, schizotypal personality, and functioning (Ruhrmann et al., 2010).

With cut-offs specified (matching the 22% transition rate in this subsample) the Perkins et al. (2015) model with sample-estimated parameters obtained a DOR of 13, in contrast to a geometric mean of 5.1 for the other models.

First hospitalization

Hospital care for any psychiatric disorder was best explained by models including presence of marked positive symptoms, bizarre thinking, sleep disturbance, schizotypal personality, and functioning (Ruhrmann et al., 2010), or delusion intensity and nonverbal reasoning score (Lin et al., 2013), both models having the best AIC and AUC values. In addition, the model with the disorganized symptoms latent factor (Raballo et al., 2011) had one of the best AIC values, and a high AUC value was estimated for a model using intensity of delusions, suspiciousness, decreased ideational richness, and trouble with attention (Perkins et al., 2015). However, the existing models had quite poor predictive abilities. As a comparison, previous hospitalization explained rehospitalizations better than most of these models predicted the first hospitalization (online Supplement S4).

When applied only to the CHR subsample, hospitalizations in our sample were best predicted by positive symptoms, bizarre thinking, sleep disturbance, schizotypal personality, and functioning (Ruhrmann et al., 2010), disorganized communication (DeVylder et al., 2014), or delusions and nonverbal reasoning (Lin et al., 2013) (online Supplement S4). As assessed with the AUC, also models consisting of delusions, suspiciousness, decreased ideational richness, and trouble with attention (Perkins et al., 2015), or genetic risk and deterioration (GRD) syndrome, delusions, suspiciousness, and substance abuse (Cannon et al., 2008) were among the most discriminative models.

Exploratory models

The exploratory model results can be found in the online Supplement S3. Online Supplementary Fig. S1 illustrates the

Table 2. Performance of models in predicting psychosis in the whole sample and in the CHR subsample

Model	Parameter	Whole sample, $n = 153$				CHR subsample, $n = 50$			
		log OR (95% CI)	p	AIC[a]	AUC[a]	log OR (95% CI)	p	AIC[a]	AUC[a]
Miller	*CHR status*	1.30 (0.33–2.27)	0.01	−4.78	0.64	0.00 (−0.70 to 0.70)	>0.99	2.00	0.50
Auther	SIPS positive symptoms sum	0.19 (0.09–0.30)	<0.01	**−11.93**	0.70	0.34 (0.09–0.68)	0.01	−5.48	0.78
Buchy	No alcohol abuse or dependency	−1.13 (−2.20 to 0.04)	0.06	−1.63	0.58	−0.13 (−1.66 to 1.67)	0.87	1.97	0.50
Cannon 1. model (best 4 by HR)	SIPS P1 >3	2.01 (0.52–3.59)	0.01	−1.26	0.64	1.54 (0.19–3.06)	0.03	2.23	0.71
	Substance-related diagnosis	−0.05 (−1.59 to 1.19)	0.94			−0.30 (−2.69 to 1.55)	0.76		
	SIPS P2 >2	−0.02 (−1.56 to 1.26)	0.98			0.15 (−1.31 to 1.71)	0.84		
	GRD	1.25 (−3.81 to 4.58)	0.51			0.24 (−4.76 to 3.04)	0.88		
Cannon 2. model (best 3 by HR)	SIPS P1 >3	1.90 (0.58–3.21)	0.01	−6.21	0.66	1.78 (0.29–3.52)	0.02	**−7.95**	**0.83**
	Substance-related diagnosis	0.17 (−1.38 to 1.44)	0.81			0.13 (−2.28 to 2.07)	0.90		
	Social functioning <7	0.93 (−0.08 to 1.93)	0.07			2.02 (0.56–3.73)	0.01		
Demjaha	Negative symptom factor score	1.68 (0.43–3.02)	0.01	−10.02	0.71	1.04 (−0.88 to 3.06)	0.29	**−6.20**	0.77
	Disorg. symptom factor score	−0.06 (−1.30 to 1.13)	0.92			1.79 (−0.20 to 4.19)	0.08		
DeVylder 1. model	SIPS P5 ⩾2	0.82 (−0.53 to 2.00)	0.21	0.46	0.54	0.15 (−1.40 to 1.53)	0.84	1.96	0.51
DeVylder 2. model	SIPS P5 ⩾3	1.89 (0.17–3.57)	0.03	−2.54	0.55	1.37 (−0.61 to 3.36)	0.16	0.06	0.57
DeVylder 3. model	SIPS P5 ⩾4	2.44 (−1.33 to 7.53)	0.18	0.16	0.51	2.42 (−0.58 to 7.44)	0.11	−0.54	0.55
Lin 1. model	Visual Reproduction I	0.14 (−0.06 to 0.38)	0.20	1.08	0.61	−0.01 (−0.26 to 0.27)	0.95	−0.95	0.70
	GAF	−0.02 (−0.05 to 0.01)	0.22			−0.07 (−0.13 to −0.01)	0.03		
Lin 2. model	Matrix Reasoning	0.10 (−0.04 to 0.26)	0.16	−9.47	0.70	0.01 (−0.15 to 0.19)	0.91	−5.43	**0.83**
	SIPS P1	0.59 (0.26–0.94)	<0.01			1.15 (0.38–2.15)	<0.01		
Nelson	Untreated symptoms >2 years	−1.25 (−4.59 to 0.54)	0.20	−0.50	0.62	−0.95 (−5.87 to 1.44)	0.49	3.21	0.59
	GAF <44	0.82 (−0.26 to 1.83)	0.13			0.31 (−1.05 to 1.69)	0.66		
Perkins equation[b]	Sum of SIPS P1, P2, N5, D3	0.20 (0.07–0.34)	<0.01	−6.94	0.65	0.33 (0.07–0.63)	0.01	−4.41	0.72
Perkins, separately	SIPS P1	0.71 (0.22–1.25)	<0.01	−8.80	**0.74**	0.95 (0.19–1.86)	0.01	−2.23	**0.80**
	SIPS P2	−0.10 (−0.61 to 0.41)	0.70			−0.20 (−0.91 to 0.50)	0.57		
	SIPS N5	0.39 (−0.17 to 0.92)	0.17			0.50 (−0.09 to 1.16)	0.10		
	SIPS D3	−0.46 (−1.07 to 0.13)	0.12			−0.06 (−1.15 to 1.11)	0.91		
Piskulic	SIPS negative symptoms sum	0.17 (0.07–0.28)	<0.01	**−10.27**	**0.71**	0.20 (0.06–0.38)	<0.01	**−6.22**	0.74
Raballo	Disorg. symptom factor score	0.83 (0.28–1.43)	<0.01	−7.07	0.68	2.32 (0.94–4.29)	<0.01	**−10.78**	**0.81**
Ruhrmann, separately[b]	SIPS D2 >2	1.72 (0.22–3.19)	0.03	−7.93	0.69	1.86 (−0.18 to 4.34)	0.07	2.24	0.69
	SIPS G1 >2	−0.54 (−1.68 to 0.51)	0.32			−0.42 (−1.89 to 1.10)	0.57		
	SIPS schizotypal personality	0.22 (−1.97 to 2.05)	0.83			−0.90 (−3.84 to 1.55)	0.49		

Model	Predictor	HR (95% CI)	p	AIC	AUC	log OR (95% CI)	p	AIC	AUC
	GAF 12 month max.	−0.04 (−0.08,0.00)	0.04			−0.02 (−0.07 to 0.03)	0.49		
Ruhrmann modified[c]	SIPS positive symptoms sum	0.15 (0.00–0.30)	0.05	−9.49	**0.75**	0.37 (0.04–0.79)	0.02	−0.58	**0.81**
	SIPS D2 >2	0.99 (−0.72 to 2.63)	0.25			1.68 (−0.61 to 4.33)	0.15		
	SIPS G1 >2	−0.80 (−2.03 to 0.30)	0.16			−0.74 (−2.44 to 0.90)	0.37		
	SIPS schizotypal personality	−0.26 (−2.58 to 1.66)	0.80			−2.20 (−6.10 to 0.78)	0.16		
	GAF 12 month max.	−0.02 (−0.07 to 0.02)	0.31			0.00 (−0.06 to 0.06)	0.90		
Ruhrmann equation[d]		0.63 (0.22–1.16)	<0.01	−7.29	0.62	0.48 (−0.06 to 1.17)	0.08	−1.05	0.58
Thompson equation[b]	2 of: SIPS P1 >3, GAF <50, GRD	2.30 (0.93–3.73)	<0.01	−8.13	0.61	1.64 (0.30–3.10)	0.02	−3.74	0.70
Thompson, separately	SIPS P1 >3	1.77 (0.44–3.09)	0.01	−7.25	0.69	1.78 (0.39–3.37)	0.01	−0.70	0.75
	GRD	1.19 (−3.92 to 4.67)	0.54			0.66 (−4.36 to 3.59)	0.71		
	GAF <50	1.02 (0.01–2.13)	0.05			0.82 (−1.01 to 3.22)	0.40		
Velthorst	GAF <61	−2.73 (−4.41 to −1.18)	<0.01	**−20.97**	**0.82**	0.00 (−0.22 to 385.43)	>0.99	6.00	0.78
	SIPS positive symptoms sum	0.17 (0.03–0.32)	0.02			0.00 (−0.15 to 0.00)	>0.99		
	SIPS negative symptoms sum	0.16 (0.02–0.31)	0.02			0.00 (−0.13 to 0.00)	0.99		
Velthorst modified	SIPS positive symptoms sum	0.12 (0.00–0.26)	0.05	**−12.25**	**0.72**	0.23 (−0.04 to 0.58)	0.09	**−6.86**	0.79
	SIPS negative symptoms sum	0.10 (−0.03 to 0.22)	0.12			0.15 (−0.01 to 0.33)	0.06		
Walder	SIPS positive symptoms sum	0.17 (0.06–0.28)	<0.01	**−10.88**	**0.70**	0.28 (0.03–0.62)	0.03	−4.91	0.79
	Social functioning	−0.19 (−0.55 to 0.18)	0.31			−0.35 (−0.89 to 0.19)	0.20		
Yung	SIPS D3 >2	−0.23 (−1.71 to 0.97)	0.73	2.77	0.64	0.16 (−1.24 to 1.61)	0.82	5.17	0.66
	GAF <40	1.23 (0.14–2.28)	0.03			1.12 (−0.26 to 2.59)	0.11		
	Untreated symptoms >5 years	−0.09 (−5.01 to 2.28)	0.95			0.59 (−4.49 to 3.78)	0.75		
	Both APS and GRD	1.26 (−3.79 to 4.60)	0.51			1.86 (−4.96 to 2.74)	0.99		
Four-predictor exploratory model[e]		3.41 (1.79–5.32)	<0.01	−16.20	0.72	4.82 (1.89–8.80)	<0.01	−9.88	0.79

AIC, Akaike information criterion; AUC, area under the curve; APS, Attenuated Psychotic Symptoms risk group; CHR, clinical high-risk; GAF, Global Assessment of Functioning; GRD, genetic risk and deterioration (functional decline) risk group; HR, hazard ratio; log OR, logarithmic odds ratio; SIPS, Structured Interview for Prodromal Syndromes.

[a]For the whole model. The five best values in the AIC and AUC columns are in boldface.

[b]Models presented as an equation such as sum of the predictor scores were used both as an equation as well as using the predictors individually.

[c]The Ruhrmann model included a criterion of sum of positive symptoms >16, but as only one person fulfilled this criterion, the model was run with this predictor replaced by the SIPS positive symptoms sum (used in other models) and these results are presented as 'Ruhrmann modified' model.

[d]Ruhmann equation: 1.571 × (SIPS positive symptoms sum >16) + (0.865 × SIPS D2 >2) + (0.793 × SIPS G1 >2) + (1.037 × SIPS schizotypal personality) + (0.033 × (100 − GAF 12 month max. − 34.64)).

[e]SIPS negative symptoms sum, SIPS positive symptoms sum, SIPS D2 >2, SIPS positive symptoms sum >16.
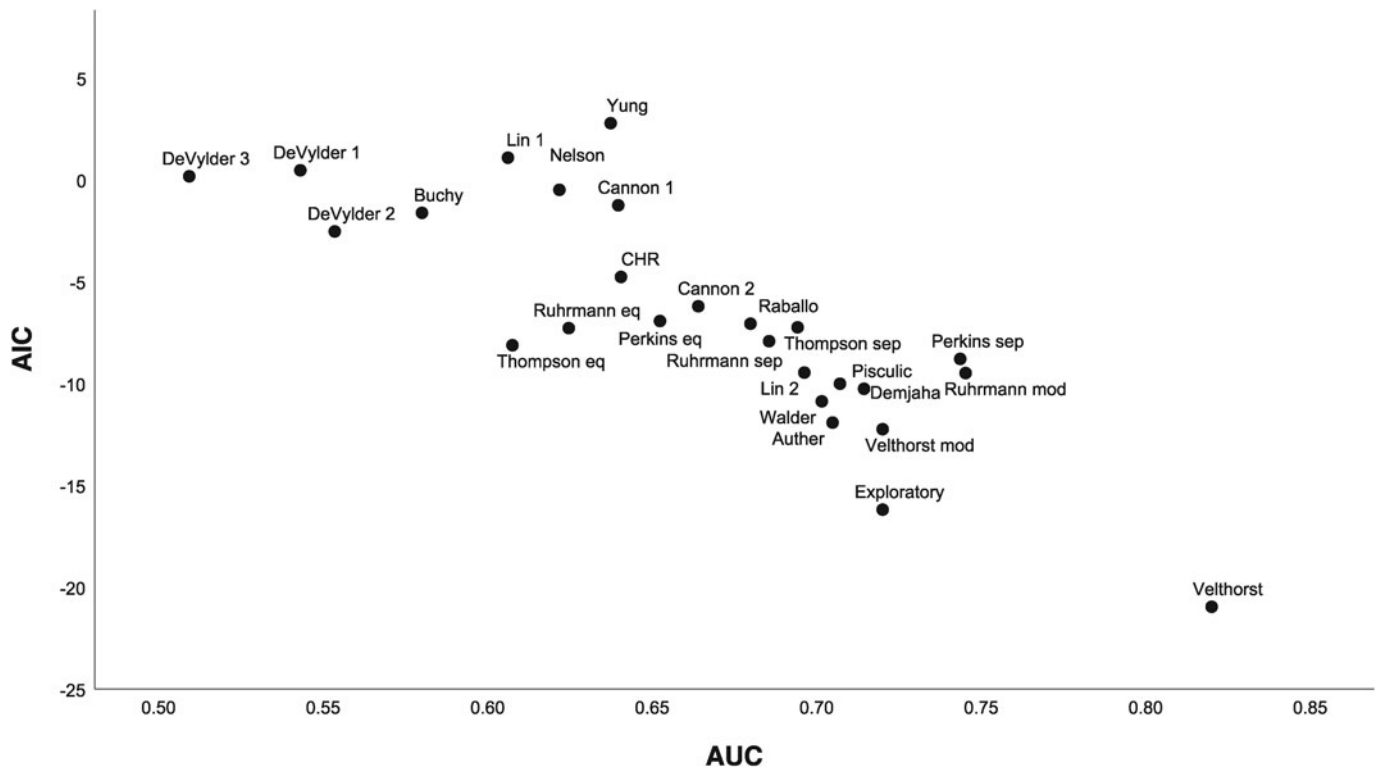
Fig. 2. AIC and AUC values of the models in predicting psychosis, using the whole sample. Note that lower AIC values are better.

standardized coefficients predicting psychosis and hospitalizations with the LASSO at the relevant ranges of λ.

Psychosis transition was best predicted by the SIPS negative symptoms sum score, the SIPS positive symptoms sum score, having a positive symptoms sum score >16, and the presence of bizarre thinking (using a cut-off score of >2 on SIPS D2). Unstandardized coefficients gave similar results. The performance of the one-, two-, and three-predictor exploratory models can be found in online Supplement S3. The four- and three-parameter exploratory models were not substantially better than the two-parameter model including only the negative and positive symptom sum scores (all AUC = 0.72 and DOR = 10) (Table 2; online Supplementary workbook).

Hospitalizations were predicted by low functioning (using a cut-off score of <50 on the GAF), the Demjaha et al. negative symptom factor score, presence of delusions, and sleep problems (using a cut-off of >2 in SIPS G1) (online Supplement S3). The three- and four-predictor exploratory models had AUC values of 0.70 and 0.72, respectively, and the same DOR of 3.1 (online Supplement S4 and Supplementary workbook).

Conclusions

This study tested previously published psychosis prediction models chosen from the exhaustive reviews by Studerus et al. (2017) and Montemagni et al. (2020). Numerous prediction models have been suggested to improve prediction of psychosis incidence beyond CHR status alone, using different combinations of symptoms, functioning, neurocognition, and substance use. However, these models have seldom been tested or externally validated. The models have been developed in CHR clinic settings, and we wanted to test their generalizability in a general adolescent psychiatric context.

The models predicted psychosis better than CHR status alone, but not as well as in the original studies. Comparison is difficult, however, as the overall predictive model accuracy has seldom been reported. The best models used sum of positive and negative symptoms (Velthorst et al., 2013) or just the positive symptom sum score (Auther et al., 2012). Among other best performing models in our sample were the ones with severity of positive symptoms and level of social functioning (Walder et al., 2013) or severity of negative symptoms (Piskulic et al., 2012) as predictors. Additionally, the best discriminating models used the severity of positive symptoms, bizarre thinking, level of functioning, sleep problems, and schizotypal traits (Ruhrmann et al., 2010), or delusions, suspiciousness, decreased ideational richness, and trouble with attention (Perkins et al., 2015). Many of the best models thus avoided dichotomization of baseline variables and explicitly included the severity of positive and negative symptoms. Our results are in line with a recent review, summing prognostic evidence among CHR individuals, and highlighting the significance of baseline severity of positive and negative symptoms as well as functional level as predictors of psychosis (Fusar-Poli et al., 2020). The highest discriminability estimates were in the range of AUC = 0.71–0.82, which exceeds the threshold of a prediction model that can be considered as clinically useful (AUC = 0.7 or even AUC = 0.8; Schummers, Himes, Bodnar, and Hutcheon, 2016).

Because a psychosis risk state may be more indicative of imminent deterioration in functional outcome or severity of general psychopathology than transition to psychosis, prediction of first psychiatric hospitalization was additionally used as a clinically relevant proxy. 'Psychosis risk symptoms', especially in their milder form, are not specific to the psychosis prodrome, but appear as markers of unspecific psychiatric symptomatology and lowered functioning (Healy et al., 2019; Trotta et al., 2020;

Werbeloff et al., 2012; Wigman et al., 2012). Predicting transdiagnostic functional outcomes of these symptoms may therefore be useful in detecting those at most in need of help. Although the models were originally not used to predict hospitalization, they seemed to predict this kind of nonspecific functional decline to a certain extent. The same proposed risk indicator combinations that predicted psychosis best also often were the best in predicting hospitalization for psychiatric disorder: positive symptoms, bizarre thinking, sleep disturbance, schizotypal personality, and general functioning (Ruhrmann et al., 2010), delusional ideas and visual reasoning (Lin et al., 2013), the disorganization symptom factor (Raballo et al., 2011), or delusions, suspiciousness, decreased ideational richness, and attention (Perkins et al., 2015) were among the best models for predicting hospitalization. Although baseline negative symptoms emerged as significant predictors of psychosis, they did not seem to be associated with later hospitalizations in the same way. The best single predictor of rehospitalization was, unsurprisingly, previous hospitalization.

Psychosis risk predictions may work at an acceptable level only in very enriched populations, diminishing their usefulness in clinical use (van Os & Guloksuz, 2017). The original models predicted psychosis specifically in CHR populations suspected to have psychosis risk symptoms already at referral, and many false positives have been anticipated to emerge if used in nonselected clinical populations (Ruhrmann et al., 2010). The current sample consisted of adolescents in their initial psychiatric helpseeking phase, who were at somewhat heightened risk for psychosis based on the questionnaire screening, but also included screen-negatives, and the use of weights allows for generalization to the whole base population of adolescents in psychiatric care. Psychosis outcomes are less likely within this sample than in psychosis risk clinics, and it was thus expected that the accuracy of the models would be reduced. However, we also tested the predictiveness of the models in our CHR subsample, and the models still performed worse than in the samples they were derived from, despite sample-derived parameters. Among the CHR adolescents, a common feature of the best performing models in predicting psychosis was including the intensity of positive symptoms, and especially delusional thought. Psychotic illnesses in the CHR group were best explained by models using the disorganization symptom factor by Raballo et al. (2011) consisting of odd behavior, cognitive difficulties, disorganized communications, and delusions, and a model combining delusions, social functioning, and substance abuse (Cannon et al., 2008). Using delusion intensity and visual reasoning score as predictors (Lin et al., 2013) best discriminated CHR adolescents with and without psychosis during follow-up. First psychiatric hospitalizations in the CHR group were best explained by models combining positive symptoms, bizarre thinking, sleep disturbance, schizotypal personality, and functioning (Ruhrmann et al., 2010).

Using the predictors individually with weights estimated from the current data (sample-optimized version) (Perkins et al., 2015; Ruhrmann et al., 2010; Thompson et al., 2011) led to better prediction values compared to using the original equations. A modified version of the Ruhrmann model was additionally used where the sum of positive symptoms was used to replace the dichotomized >16 criterion which was rare in the sample, leading to best fit. The Ruhrmann model also had to be modified because of the lack of variance in education and schizotypal personality disorder in our adolescent sample (Table 1).

That the cut-off of functioning in the Velthorst model predicted psychosis in the 'wrong direction' stresses how the

predictors depend on the sample used, and in this sample the adolescents with higher baseline functioning were more likely to develop a psychotic disorder. This is also an example of overfitting, highlighting the need for replicable models with published parameters.

In the exploratory analyses using the predictors proposed by the previous studies, psychotic disorders were best predicted by a combination of high overall positive and negative symptoms scores, with a larger weight on the negative symptoms (this was due to the penalty applied in the LASSO, in contrast to the modified Velthorst model, where the weights were approximately equal). There was a modest improvement in model fit when including the third and fourth criteria (at least moderately severe bizarre thinking and exceptionally severe positive psychotic-like symptoms), but adding these predictors did not result in better prediction rates. Our results are in line with a recent meta-analysis among CHR individuals which suggested that despite the large number of putative risk factors evaluated, only positive symptoms and level of functioning predicted transition to psychosis with highly suggestive evidence, whereas negative symptoms showed suggestive evidence, and no other factor showed convincing evidence (Oliver et al., 2020). In another study, employing machine learning among UHR individuals, psychosis was predicted by unusual thought content, severity of positive symptoms, and level of functioning (Mechelli et al., 2017). This previous study also predicted functional outcome with the same symptom measures, finding that it was best predicted by attention deficits, anhedonia-asociality, and unusual thought content (Mechelli et al., 2017), which were somewhat different from the variables best predicting hospitalization in the present study. In our sample, psychiatric hospitalization in the upcoming years was best predicted by serious impairment in functioning, Demjaha's negative symptom factor, delusions, and baseline sleep problems. Demjaha's negative symptom factor was most heavily loaded on avolition, blunted expression of emotion, impaired role functioning, social anhedonia, dysphoria, and odd behavior. Sleep disturbance, on the other hand, is an unspecific general symptom, common in psychiatric illnesses, which may independently add risk to worsening mental health as well as act as a marker of underlying problems, related to both sleep and mental health.

In the current study, comprehensive register data were used for the follow-up. Only information from public services was obtained, but private service use is uncommon among adolescents in Finland and there are no private psychiatric hospitals. The same 7-year follow-up time was used for all participants, they were thus followed until age 22–25. New psychosis cases may still emerge after that age, but for those participants with a longer follow-up time we noticed that only four new transitions emerged in this sample during the next 3 years (data not used).

We used the thorough extensive works by Studerus et al. and Montemagni et al. to select the prediction models. As original coefficients were usually not available, the coefficients were based on the moderate-sized current sample, limiting the comparability with the original studies. SIPS equivalents were substituted for CAARMS variables, but these two methods have been found highly comparable (Fusar-Poli et al., 2016a). We studied adolescents, although most of the original studies also included adults, or only adults. The positive symptoms of adolescents are more transitory than those of adults (Gerstenberg et al., 2016; Welsh & Tiffin, 2014) and caution is recommended when assessing psychosis risk among adolescents (Schultze-Lutter et al., 2015), but psychosis risk has been found a useful concept also in help-

seeking adolescents (Spada et al., 2016; Ulhaq, Thevan, & Adams, 2017). There have also been studies predicting psychosis with baseline diagnostic categories both in routine secondary mental health care (Fusar-Poli et al., 2017) and in the general population (Guloksuz et al., 2020), but these models were not included in the systematic reviews used.

Common methodological problems have been noticed in psychosis prediction studies (Fusar-Poli et al., 2015; Steyerberg & Vergouwe, 2014; Studerus et al., 2017) including overfitting of models, insufficient reporting of analysis procedures and results, and lack of commonly used and easily comparable indicators. All these problems were not completely avoided in the current study; a major limitation is our sample size, which makes overfitting of models a concern. This is especially relevant in relation to models which have many predictors, and was exemplified by the Velthorst et al. (2013) model. Furthermore, categorical scales such as the SIPS subscales (non-psychotic range 0–5) had to be treated as linear in our logistic models; this limitation was offset, however, by several models specifying cut-offs.

To conclude, the applicability and generalizability of psychosis prediction models was found to be only moderate in a general psychiatric sample of adolescents. We were not able to generalize the predictiveness of the majority of previous models which were based on samples from specialized psychosis risk clinics. A clinically significant functional outcome could, however, be partially predicted by models developed for psychosis risk detection, highlighting the importance of assessing psychosis risk factors. Transition to psychosis was best predicted by a parsimonious combination of positive and negative symptom total severity.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0033291721001938

References

Addington, J., & Heinssen, R. (2012). Prediction and prevention of psychosis in youth at clinical high risk. *Annual Review of Clinical Psychology*, 8, 269–289. https://doi.org/10.1146/annurev-clinpsy-032511-143146

Auther, A. M., McLaughlin, D., Carrión, R. E., Nagachandran, P., Correll, C. U., & Cornblatt, B. A. (2012). Prospective study of cannabis use in adolescents at clinical high risk for psychosis: Impact on conversion to psychosis and functional outcome. *Psychological Medicine*, 42(12), 2485–2497. https://doi.org/10.1017/S0033291712000803

Bora, E., Lin, A., Wood, S. J., Yung, A. R., McGorry, P. D., & Pantelis, C. (2014). Cognitive deficits in youth with familial and clinical high risk to psychosis: A systematic review and meta-analysis. *Acta Psychiatrica Scandinavica*, 130(1), 1–15. https://doi.org/10.1111/acps.12261 [doi]

Brown, G. S., & White, K. G. (2005). The optimal correction for estimating extreme discriminability. *Behavior Research Methods*, 37(3), 436–449. https://doi.org/10.3758/BF03192712

Buchy, L., Perkins, D., Woods, S. W., Liu, L., & Addington, J. (2014). Impact of substance use on conversion to psychosis in youth at clinical high risk of psychosis. *Schizophrenia Research*, 156(2–3), 277–280. https://doi.org/10.1016/j.schres.2014.04.021

Cannon, T. D., Cadenhead, K., Cornblatt, B., Woods, S. W., Addington, J., Walker, E., … Heinssen, R. (2008). Prediction of psychosis in youth at high clinical risk: A multisite longitudinal study in North America. *Archives of General Psychiatry*, 65(1), 28–37. https://doi.org/65/1/28[pii]10.1001/archgenpsychiatry.2007.3

Carrión, R. E., McLaughlin, D., Goldberg, T. E., Auther, A. M., Olsen, R. H., Olvet, D. M., … Cornblatt, B. A. (2013). Prediction of functional outcome in individuals at clinical high risk for psychosis. *JAMA Psychiatry*, 70(11), 1133–1142.

Cornblatt, B. A., Auther, A. M., Niendam, T., Smith, C. W., Zinberg, J., Bearden, C. E., & Cannon, T. D. (2007). Preliminary findings for two new measures of social and role functioning in the prodromal phase of schizophrenia. *Schizophrenia Bulletin*, 33(3), 688–702. https://doi.org/sbm029[pii]10.1093/schbul/sbm029

Cotter, J., Drake, R. J., Bucci, S., Firth, J., Edge, D., & Yung, A. R. (2014). What drives poor functioning in the at-risk mental state? A systematic review. *Schizophrenia Research*, 159(2–3), 267–277. https://doi.org/http://dx.doi.org/10.1016/j.schres.2014.09.012

Demjaha, A., Valmaggia, L., Stahl, D., Byrne, M., & McGuire, P. (2012). Disorganization/cognitive and negative symptom dimensions in the at-risk mental state predict subsequent transition to psychosis. *Schizophrenia Bulletin*, 38(2), 351–359. https://doi.org/sbq088[pii]10.1093/schbul/sbq088

DeVylder, J. E., Muchomba, F. M., Gill, K. E., Ben-David, S., Walder, D. J., Malaspina, D., & Corcoran, C. M. (2014). Symptom trajectories and psychosis onset in a clinical high-risk cohort: The relevance of subthreshold thought disorder. *Schizophrenia Research*, 159(2–3), 278–283. https://doi.org/S0920-9964(14)00419-8[pii]

Ewald, B. (2006). Post hoc choice of cut points introduced bias to diagnostic research. *Journal of Clinical Epidemiology*, 59(8), 798–801. https://doi.org/10.1016/j.jclinepi.2005.11.025

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22. https://doi.org/10.18637/jss.v033.i01

Fusar-Poli, P., Bonoldi, I., Yung, A. R., Borgwardt, S., Kempton, M. J., Valmaggia, L., … McGuire, P. (2012). Predicting psychosis: Meta-analysis of transition outcomes in individuals at high clinical risk. *Archives of General Psychiatry*, 69(3), 220–229. https://doi.org/69/3/220[pii]10.1001/archgenpsychiatry.2011.1472

Fusar-Poli, P., Cappucciati, M., Rutigliano, G., Lee, T. Y., Beverly, Q., Bonoldi, I., … McGuire, P. (2016a). Towards a standard psychometric diagnostic interview for subjects at ultra high risk of psychosis: CAARMS versus SIPS. *Psychiatry Journal*, 7146341. doi: 10.1155/2016/7146341. Epub 2016 May 30. https://doi.org/10.1155/2016/7146341

Fusar-Poli, P., Cappucciati, M., Rutigliano, G., Schultze-Lutter, F., Bonoldi, I., Borgwardt, S., … McGuire, P. (2015). At risk or not at risk? A meta-analysis of the prognostic accuracy of psychometric interviews for psychosis prediction. *World Psychiatry*, 14(3), 322–332. https://doi.org/10.1002/wps.20250

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. https://doi.org/10.1001/jamapsychiatry.2017.0284

Fusar-Poli, P., Salazar De Pablo, G., Correll, C. U., Meyer-Lindenberg, A., Millan, M. J., Borgwardt, S., … Arango, C. (2020). Prevention of psychosis: Advances in detection, prognosis, and intervention. *JAMA Psychiatry*, 77(7), 755–765. https://doi.org/10.1001/jamapsychiatry.2019.4779

Fusar-Poli, P., Schultze-Lutter, F., Cappucciati, M., Rutigliano, G., Bonoldi, I., Stahl, D., … Mcguire, P. (2016b). 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. https://doi.org/10.1093/schbul/sbv162

Gerstenberg, M., Theodoridou, A., Traber-Walker, N., Franscini, M., Wotruba, D., Metzler, S., … Heekeren, K. (2016). Adolescents and adults at clinical high-risk for psychosis: Age-related differences in attenuated positive symptoms syndrome prevalence and entanglement with basic symptoms. *Psychological Medicine*, 46(5), 1069–1078. https://doi.org/10.1017/S0033291715002627

Grau, J., Grosse, I., & Keilwagen, J. (2015). PRROC: Computing and visualizing precision-recall and receiver operating characteristic curves in R. *Bioinformatics (Oxford, England)*, 31(15), 2595–2597. https://doi.org/10.1093/bioinformatics/btv153

Guloksuz, S., Pries, L. K., ten Have, M., de Graaf, R., van Dorsselaer, S., Klingenberg, B., … van Os, J. (2020). Association of preceding psychosis risk states and non-psychotic mental disorders with incidence of clinical psychosis in the general population: A prospective study in the NEMESIS-2 cohort. *World Psychiatry*, 19(2), 199–205. https://doi.org/10.1002/wps.20755

Healy, C., Brannigan, R., Dooley, N., Coughlan, H., Clarke, M., Kelleher, I., & Cannon, M. (2019). Childhood and adolescent psychotic experiences and risk of mental disorder: A systematic review and meta-analysis. *Psychological Medicine*, 49(10), 1589–1599. https://doi.org/10.1017/S0033291719000485

Heinze, G., & Ploner, M. (2018). logistf: Firth's bias-reduced logistic regression. R package version 1.23.

Jeppesen, P., Clemmensen, L., Munkholm, A., Rimvall, M. K., Rask, C. U., Jørgensen, T., … Skovgaard, A. M. (2015). Psychotic experiences co-occur with sleep problems, negative affect and mental disorders in preadolescence. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, 56(5), 558–565. https://doi.org/10.1111/jcpp.12319

Lin, A., Wood, S. J., Nelson, B., Brewer, W. J., Spiliotacopoulos, D., Bruxner, A., … Yung, A. R. (2011). Neurocognitive predictors of functional outcome two to 13 years after identification as ultra-high risk for psychosis. *Schizophrenia Research*, 132(1), 1–7. https://doi.org/10.1016/j.schres.2011.06.014 [doi]

Lin, A., Yung, A. R., Nelson, B., Brewer, W. J., Riley, R., Simmons, M., … Wood, S. J. (2013). Neurocognitive predictors of transition to psychosis: Medium- to long-term findings from a sample at ultra-high risk for psychosis. *Psychological Medicine*, 43(11), 2349–2360. https://doi.org/10.1017/S0033291713000123 [doi]

Lindgren, M., Manninen, M., Kalska, H., Mustonen, U., Laajasalo, T., Moilanen, K., … Therman, S. (2014). Predicting psychosis in a general adolescent psychiatric sample. *Schizophrenia Research*, 158(1–3), 1–6. https://doi.org/10.1016/j.schres.2014.06.028

Lindgren, M., Manninen, M., Laajasalo, T., Mustonen, U., Kalska, H., Suvisaari, J., … Therman, S. (2010). The relationship between psychotic-like symptoms and neurocognitive performance in a general adolescent psychiatric sample. *Schizophrenia Research*, 123(1), 77–85. https://doi.org/S0920-9964(10)01441-6[pii]10.1016/j.schres.2010.07.025

Loewy, R. L., Bearden, C. E., Johnson, J. K., Raine, A., & Cannon, T. D. (2005). The prodromal questionnaire (PQ): Preliminary validation of a self-report screening measure for prodromal and psychotic syndromes. *Schizophrenia Research*, 79(1), 117–125.

Mechelli, A., Lin, A., Wood, S., McGorry, P., Amminger, P., Tognin, S., … Yung, A. (2017). Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. *Schizophrenia Research*, 184, 32–38. https://doi.org/10.1016/j.schres.2016.11.047

Miller, T. J., McGlashan, T. H., Rosen, J. L., Cadenhead, K., Cannon, T., Ventura, J., … Woods, S. W. (2003). Prodromal assessment with the structured interview for prodromal syndromes and the scale of prodromal symptoms: Predictive validity, interrater reliability, and training to reliability. *Schizophrenia Bulletin*, 29(4), 703–715.

Millman, Z. B., Gold, J. M., Mittal, V. A., & Schiffman, J. (2019). The critical need for help-seeking controls in clinical high-risk research. *Clinical Psychological Science*, 7(5), 1171–1189. https://doi.org/10.1177/2167702619855660

Montemagni, C., Bellino, S., Bracale, N., Bozzatello, P., & Rocca, P. (2020). Models predicting psychosis in patients with high clinical risk: A systematic review. *Frontiers in Psychiatry*, 11, 223. https://doi.org/10.3389/fpsyt.2020.00223

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén, http://statmodel.com

Nelson, B., Yuen, H., Wood, S. J., **Lin, A., Spiliotacopoulos, D., Bruxner, A., … Yung, A.R.** (2013). Long-term follow-up of a group at ultra high risk ('prodromal') for psychosis: The pace 400 study. *JAMA Psychiatry*, 7(8), 793–802. doi: 10.1001/jamapsychiatry.2013.1270

Oliver, D., Reilly, T. J., Baccaredda Boy, O., Petros, N., Davies, C., Borgwardt, S., … Fusar-Poli, P. (2020). What causes the onset of psychosis in individuals at clinical high risk? A meta-analysis of risk and protective factors. *Schizophrenia Bulletin*, 46(1), 110–120. https://doi.org/10.1093/schbul/sbz039

Perkins, D. O., Jeffries, C. D., Cornblatt, B. A., Woods, S. W., Addington, J., Bearden, C. E., … McGlashan, T. H. (2015). Severity of thought disorder predicts psychosis in persons at clinical high-risk. *Schizophrenia Research*, 169(1–3), 169–177. https://doi.org/10.1016/j.schres.2015.09.008

Piskulic, D., Addington, J., Cadenhead, K. S., Cannon, T. D., Cornblatt, B. A., Heinssen, R., … McGlashan, T. H. (2012). Negative symptoms in individuals at clinical high risk of psychosis. *Psychiatry Research*, 196(2–3), 220–224. https://doi.org/S0165-1781(12)00100-X[pii]10.1016/j.psychres.2012.02.018

Raballo, A., Nelson, B., Thompson, A., & Yung, A. (2011). The comprehensive assessment of at-risk mental states: From mapping the onset to mapping the structure. *Schizophrenia Research*, 127(1–3), 107–114. https://doi.org/10.1016/j.schres.2010.12.021

R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Revelle, W. (2021) *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 2.1.3, https://CRAN.R-project.org/package=psych.

Ruhrmann, S., Schultze-Lutter, F., Salokangas, R. K., Heinimaa, M., Linszen, D., Dingemans, P., … Klosterkötter, J. (2010). Prediction of psychosis in adolescents and young adults at high risk: Results from the prospective European prediction of psychosis study. *Archives of General Psychiatry*, 67(3), 241–251. https://doi.org/67/3/241[pii]10.1001/archgenpsychiatry.2009.206

Schultze-Lutter, F., Michel, C., Schmidt, S. J., Schimmelmann, B. G., Maric, N. P., Salokangas, R. K. R., … Klosterkötter, J. (2015). EPA guidance on the early detection of clinical high risk states of psychoses. *European Psychiatry*, 30(3), 405–416. https://doi.org/http://dx.doi.org/10.1016/j.eurpsy.2015.01.010

Schummers, L., Himes, K. P., Bodnar, L. M., & Hutcheon, J. A. (2016). Predictor characteristics necessary for building a clinically useful risk prediction model: A simulation study. *BMC Medical Research Methodology*, 16(1), 123. https://doi.org/10.1186/s12874-016-0223-2

Spada, G., Molteni, S., Pistone, C., Chiappedi, M., McGuire, P., Fusar-Poli, P., & Balottin, U. (2016). Identifying children and adolescents at ultra high risk of psychosis in Italian neuropsychiatry services: A feasibility study. *European Child and Adolescent Psychiatry*, 25(1), 91–106. https://doi.org/10.1007/s00787-015-0710-8

Steyerberg, E. W., & Vergouwe, Y. (2014). Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *European Heart Journal*, 35(29), 1925–1931. https://doi.org/10.1093/eurheartj/ehu207

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

Therman, S., Lindgren, M., Manninen, M., Loewy, R. L. L., Huttunen, M. O. O., Cannon, T. D. D., & Suvisaari, J. (2014). Predicting psychosis and psychiatric hospital care among adolescent psychiatric patients with the prodromal questionnaire. *Schizophrenia Research*, 158(1–3), 7–10. https://doi.org/S0920-9964(14)00323-5[pii]

Thompson, A., Nelson, B., & Yung, A. (2011). Predictive validity of clinical variables in the 'at risk' for psychosis population: International comparison with results from the North American prodrome longitudinal study. *Schizophrenia Research*, 126(1–3), 51–57. https://doi.org/S0920-9964(10)01553-7[pii]10.1016/j.schres.2010.09.024

Trotta, A., Arseneault, L., Caspi, A., Moffitt, T. E., Danese, A., Pariante, C., & Fisher, H. L. (2020). Mental health and functional outcomes in young adulthood of children with psychotic symptoms: A longitudinal cohort study. *Schizophrenia Bulletin*, *46*(2), 261–271. https://doi.org/10.1093/schbul/sbz069

Ulhaq, S., Thevan, K., & Adams, R. (2017). Innovations in practice: Identifying young people at ultra-high risk of psychosis in a child and adolescent mental health service. *Child and Adolescent Mental Health*, *22*(2), 91–95. https://doi.org/10.1111/camh.12173

van Os, J., & Guloksuz, S. (2017). A critique of the 'ultra-high risk' and 'transition' paradigm. *World Psychiatry*, *16*(2), 200–206. https://doi.org/10.1002/wps.20423

Velthorst, E., Nelson, B., Wiltink, S., De Haan, L., Wood, S. J., Lin, A., & Yung, A. R. (2013). Transition to first episode psychosis in ultra high risk populations: Does baseline functioning hold the key? *Schizophrenia Research*, *143*(1), 132–137. https://doi.org/10.1016/j.schres.2012.10.025

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Walder, D. J., Holtzman, C. W., Addington, J., Cadenhead, K., Tsuang, M., Cornblatt, B., … Walker, E. F. (2013). Sexual dimorphisms and prediction of conversion in the NAPLS psychosis prodrome. *Schizophrenia Research*, *144*(1–3), 43–50. https://doi.org/10.1016/j.schres.2012.11.039

Wechsler, D. (1987). *Wechsler memory scale – revised*. New York, NY: The Psychological Corporation.

Wechsler, D. (1997). *Wechsler adult intelligence scale* (3rd ed.). San Antonio, Texas: The Psychological Corporation.

Welsh, P., & Tiffin, P. A. (2014). The 'at-risk mental state' for psychosis in adolescents: Clinical presentation, transition and remission. *Child Psychiatry and Human Development*, *45*(1), 90–98. https://doi.org/10.1007/s10578-013-0380-z

Werbeloff, N., Drukker, M., Dohrenwend, B. P., Levav, I., Yoffe, R., van Os, J., … Weiser, M. (2012). Self-reported attenuated psychotic symptoms as forerunners of severe mental disorders later in life. *Archives of General Psychiatry*, *69*(5), 467–475. https://doi.org/archgenpsychiatry.2011.1580 [pii]10.1001/archgenpsychiatry.2011.1580

Wigman, J. T., van Nierop, M., Vollebergh, W. A., Lieb, R., Beesdo-Baum, K., Wittchen, H. U., & van Os, J. (2012). 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. *Schizophrenia Bulletin*, *38*(2), 247–257. https://doi.org/sbr196[pii]10.1093/schbul/sbr196

World Medical Association (2000). World medical association declaration of Helsinki: Ethical principles for medical research involving human subjects. *JAMA*, *284*(23), 3043–3045. doi: 10.1001/jama.284.23.3043

Worthington, M. A., Cao, H., & Cannon, T. D. (2019). Discovery and validation of prediction algorithms for psychosis in youth at clinical high risk. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(8), 738–747. https://doi.org/10.1016/j.bpsc.2019.10.006

Yung, A. R., Phillips, L. J., Yuen, H. P., & McGorry, P. D. (2004). Risk factors for psychosis in an ultra high-risk group: Psychopathology and clinical features. *Schizophrenia Research*, *67*(2-3), 131–142.

Yung, A. R., Stanford, C., Cosgrave, E., Killackey, E., Phillips, L., Nelson, B., & McGorry, P. D. (2006). Testing the ultra high risk (prodromal) criteria for the prediction of psychosis in a clinical sample of young people. *Schizophrenia Research*, *84*(1), 57–66. doi: https://doi.org/S0920-9964(06)00108-3[pii]10.1016/j.schres.2006.03.014