

Original Article

Previous presentation: The 2022 Congress of the Schizophrenia International Research Society, Florence, Italy, 6–10 April 2022.

Cite this article: Worthington MA *et al* (2024). Improving prediction of psychosis in youth at clinical high-risk: pre-baseline symptom duration and cortical thinning as moderators of the NAPLS2 risk calculator. *Psychological Medicine* **54**, 611–619. <https://doi.org/10.1017/S0033291723002301>

Received: 19 January 2023

Revised: 10 July 2023

Accepted: 19 July 2023

First published online: 29 August 2023


Keywords:

Clinical high-risk; cortical thinning; psychosis; risk calculator; schizophrenia

Corresponding author:

Michelle A. Worthington;
Email: michelle.worthington@yale.edu

Improving prediction of psychosis in youth at clinical high-risk: pre-baseline symptom duration and cortical thinning as moderators of the NAPLS2 risk calculator

Michelle A. Worthington¹ , Meghan A. Collins¹, Jean Addington², Carrie E. Bearden³, Kristin S. Cadenhead⁴, Barbara A. Cornblatt⁵, Matcheri Keshavan⁶, Daniel H. Mathalon⁷, Diana O. Perkins⁸, William S. Stone⁶, Elaine F. Walker⁹, Scott W. Woods¹⁰ and Tyrone D. Cannon^{1,10}

¹Department of Psychology, Yale University, New Haven, CT, USA; ²Department of Psychiatry, Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada; ³Department of Psychiatry and Biobehavioral Sciences, UCLA, Los Angeles, CA, USA; ⁴Department of Psychiatry, UCSD, San Diego, CA, USA; ⁵Department of Psychiatry, Zucker Hillside Hospital, Long Island, NY, USA; ⁶Department of Psychiatry, Harvard Medical School at Beth Israel Deaconess Medical Center and Massachusetts General Hospital, Boston, MA, USA; ⁷Department of Psychiatry, UCSF, and SFVA Medical Center, San Francisco, CA, USA; ⁸Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA; ⁹Departments of Psychology and Psychiatry, Emory University, Atlanta, GA, USA and ¹⁰Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA

Abstract

Background. Clinical implementation of risk calculator models in the clinical high-risk for psychosis (CHR-P) population has been hindered by heterogeneous risk distributions across study cohorts which could be attributed to pre-ascertainment illness progression. To examine this, we tested whether the duration of attenuated psychotic symptom (APS) worsening prior to baseline moderated performance of the North American prodrome longitudinal study 2 (NAPLS2) risk calculator. We also examined whether rates of cortical thinning, another marker of illness progression, bolstered clinical prediction models.

Methods. Participants from both the NAPLS2 and NAPLS3 samples were classified as either 'long' or 'short' symptom duration based on time since APS increase prior to baseline. The NAPLS2 risk calculator model was applied to each of these groups. In a subset of NAPLS3 participants who completed follow-up magnetic resonance imaging scans, change in cortical thickness was combined with the individual risk score to predict conversion to psychosis.

Results. The risk calculator models achieved similar performance across the combined NAPLS2/NAPLS3 sample [area under the curve (AUC) = 0.69], the long duration group (AUC = 0.71), and the short duration group (AUC = 0.71). The shorter duration group was younger and had higher baseline APS than the longer duration group. The addition of cortical thinning improved the prediction of conversion significantly for the short duration group (AUC = 0.84), with a moderate improvement in prediction for the longer duration group (AUC = 0.78).

Conclusions. These results suggest that early illness progression differs among CHR-P patients, is detectable with both clinical and neuroimaging measures, and could play an essential role in the prediction of clinical outcomes.

Introduction

One of the primary goals of the clinical high-risk for psychosis (CHR-P) research paradigm is to develop prognostic models estimating an individual's risk for conversion to psychosis. Such models using clinical measures have been developed (Cannon *et al.*, 2008, 2016; Fusar-Poli *et al.*, 2017; Mechelli *et al.*, 2017; Zhang *et al.*, 2018b) and have achieved prognostic accuracies ranging from 65% to 80% in correctly discriminating eventual converters from non-converters (Sanfelici, Dwyer, Antonucci, & Koutsouleris, 2020; Worthington & Cannon, 2021). While these models have shown promise and studies validating these models in external samples have been published (Carrion *et al.*, 2016; Koutsouleris *et al.*, 2021; Osborne & Mittal, 2019; Zhang *et al.*, 2018a), barriers to clinical implementation remain.

A significant barrier in ascertaining a new patient's level of risk is matching characteristics to the appropriate model and risk distribution. Existing risk calculators are developed in samples that differ in symptom severity, pathways of ascertainment, and age at ascertainment (Koutsouleris *et al.*, 2021). Because distributions of the predictor variables as well as the rate of conversion will differ across such samples, predicted risk distributions will differ if

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

the same algorithm is applied, potentially resulting in different individual-level risk predictions for a newly ascertained case. A key step toward implementing prognostic models in clinical settings will be to identify key characteristics that can feasibly and reliably differentiate risk distributions for increased precision in outcome estimation and treatment allocation.

Key differences in symptom severity, conversion rates, and demographic variables could be attributed to the progression of prodromal symptoms prior to the baseline visit. In a recent effort to validate the North American prodrome longitudinal study 2 (NAPLS2) risk calculator in the European-based PRONIA study (www.pronia.eu), validation was possible after statistical adjustments to account for the PRONIA sample's lower symptom severity, lower conversion rate, higher age, and higher neurocognitive performance compared with the North American-based NAPLS2 study (Koutsouleris *et al.*, 2021). These two samples also differ in the criteria used to assess psychosis risk syndromes at a broad level – the PRONIA study uses the comprehensive assessment of at-risk mental state (CAARMS) (Yung *et al.*, 2005) criteria and the NAPLS uses the criteria of psychosis-risk syndromes (COPS) (Miller *et al.*, 2003) – which differ in that the CAARMS permits a broader range and duration of attenuated psychotic symptom (APS). A subset of the PRONIA sample (ultra-high-risk) do match the inclusion criteria of the NAPLS sample, but only make up 20% of the larger sample. Further, these frameworks may be more or less well-suited to employing differing pathways of ascertainment, originating in either the general population as self-referrals or from other healthcare providers as clinician referrals. Thus, illness progression at the point of ascertainment may be captured differently in these two frameworks, resulting in different distributions of risk of conversion.

Evidence of differing rates of illness progression within the CHR-P population has also been detected in measures of cortical thickness. Steeper rates of cortical thinning have been observed among converters compared with non-converters (Cannon *et al.*, 2015), and recent work indicates that accelerated cortical thinning is observable among converters as compared to non-converters in as little as 3 months, on average, and prior to psychosis onset (Collins *et al.*, 2022). These changes could be attributed to disrupted neurodevelopmental processes (e.g. synaptic pruning) which may contribute to the development of psychosis (Cannon *et al.*, 2015; Germann, Brederoo, & Sommer, 2021). The progression of these neural changes may also manifest in APS onset and worsening, and early detection of these changes provides an opportunity (and potential target) for early intervention. For the purpose of predicting outcomes, models using clinical measures alone have provided a relatively high level of accuracy, and the function of neuroimaging measures in clinical prediction has yet to be demonstrated. Studies have examined the additive effects of neuroimaging measures in clinical prediction models with mixed results (Chung *et al.*, 2019a; Koutsouleris *et al.*, 2018); however, no study has examined the additive power of baseline cortical thickness and cortical thickness change in the period preceding psychosis onset to clinical models predicting conversion to psychosis.

In this study, we examined the moderating effect of duration of symptom progression prior to baseline on the resulting risk distribution as calculated by the NAPLS2 risk calculator. We hypothesized that a shorter duration of psychosis-risk symptom worsening would confer a higher risk of conversion and higher APS severity, whereas a longer duration of prodromal symptom worsening would confer a lower risk of conversion and lower

APS severity (Chung *et al.*, 2019b; Koutsouleris *et al.*, 2014). If this approach is successful, it would suggest the use of cut-points specific to each risk distribution for clinical decision-making which should produce comparable performance in sensitivity and specificity across the strata. We also sought to determine whether tracking cortical thinning in the period preceding conversion could provide a potential mechanism that underlies this clinical stratification. As a secondary analysis, we hypothesized that rates of cortical thinning would differentially add to the predictive power of clinical measures such that cortical thinning would have greater predictive power in the shorter duration group as compared to the longer duration group, given the hypothesized differentiation in rates of conversion. Evidence in support of these hypotheses would suggest that the duration of symptom worsening may serve as a meaningful variable on which to stratify CHR-P samples for risk estimation and that processes underlying steeper rates of cortical thinning support this differentiation.

Methods

Participants

Participants were drawn from the (non-overlapping) first and second phases of the NAPLS2 and NAPLS3 (Addington *et al.*, 2012, 2022). NAPLS2 and NAPLS3 are eight- and nine-site observational consortium studies, respectively, examining predictors and mechanisms related to conversion to psychosis in the CHR-P population. Participants were individuals aged 12–35 in NAPLS2 and aged 12–30 in NAPLS3 who met criteria for a psychosis risk syndrome as defined by the COPS (McGlashan, Walsh, & Woods, 2010) and as assessed by the Structured Interview for Psychosis-risk Syndromes (SIPS) (Addington *et al.*, 2007; McGlashan, Walsh, & Woods, 2001). Exclusion criteria included any current or lifetime Diagnostic and Statistical Manual of Mental Disorders-IV diagnosis of a psychotic disorder, IQ < 70, the presence of a neurological disorder, or psychosis-risk symptoms caused by another axis I disorder. In NAPLS3, study visits occurred every 2 months for the first 8 months of the study, and at 12, 18, and 24 months. In NAPLS2, study visits occurred every 6 months for the duration of the 2-year follow-up period. Participants from both NAPLS2 and NAPLS3 provided written informed consent for the study. The protocol and consent forms for each study were approved by the institutional review boards at each site.

Risk calculator assessments

In the original NAPLS2 risk calculator, eight clinical variables that were previously shown to predict conversion to psychosis were included: age, baseline severity of SIPS positive symptom items P1 and P2 (unusual thought content and suspiciousness), score on the brief assessment of cognition in schizophrenia (BACS) symbol coding (SC) test (Keefe *et al.*, 2008); score on the Hopkins Verbal Learning Test-Revised (HVLT-R) (Benedict, Schretlen, Groninger, & Brandt, 1998); decline in social functioning during the prior year as measured by the Global Functioning Social scale (GFS) (Cornblatt *et al.*, 2007); stressful life events as measured by the Research Interview Life Events Scale (Dohrenwend, Askenasy, Krasnoff, & Dohrenwend, 1978); childhood traumas as measured by the Childhood Trauma and Abuse

Scale (Janssen et al., 2004); and family history of psychotic disorder in a first-degree relative (Cannon et al., 2016).

Neuroimaging assessments

We incorporated measures of baseline cortical thickness and change in cortical thinning based on previous findings within the NAPLS3 study showing that rates of cortical thinning were accelerated for converters as compared to non-converters and healthy controls (Collins et al., 2022). Regions of interest included in the percent change measure were originally identified using linear mixed effects models at the vertex level to assess the diagnostic group-by-time interaction on longitudinal cortical thickness while controlling for age, sex, and scanner. These changes were detected specifically in primarily left hemisphere prefrontal cortical regions of interest. Percent change in cortical thickness was calculated based on cortical thickness change from scan 1 (baseline) to scan 2 (first follow-up) and normalized over the average number of months between scan 1 and scan 2 [mean (s.d.) interval = 2.93 (1.81) months]. For full details regarding magnetic resonance imaging (MRI) procedures, quality control, data processing, and scanner information, refer to Collins et al. (2022).

Statistical analysis

All statistical analyses were performed in R version 4.0.4 (R Core Team, 2021). Packages used in this analysis include rms (Harrell, 2021), pec (Mogensen, Ishwaran, & Gerds, 2012), pROC (Robin et al., 2011), and caret (Kuhn et al., 2020).

Sample stratification

Participants from both the NAPLS2 and NAPLS3 samples were included in this study to create a larger, combined sample. This larger combined sample was first divided into training and testing samples with a 50–50 split. In the training dataset, participants were grouped as either ‘long symptom duration’ or ‘short symptom duration’ based on the median number of days since any positive symptom increase prior to baseline as recorded on the SIPS interview. This median was then used to split the testing set into long and short duration testing samples. Models were developed in the training set of each sample iteration (e.g. $\text{full}_{\text{TRAIN}}$, $\text{short duration}_{\text{TRAIN}}$, $\text{long duration}_{\text{TRAIN}}$) and tested in the validation set of each sample iteration (e.g. $\text{full}_{\text{VALIDATION}}$, $\text{short duration}_{\text{VALIDATION}}$, $\text{long duration}_{\text{VALIDATION}}$).

Model training and validation

In line with the original NAPLS2 risk calculator, multivariable Cox proportional hazards regression models were developed to estimate an individual’s likelihood of conversion to psychosis within a 2-year period. In the present study, to maintain degrees of freedom in the smaller samples after stratification and to facilitate future external validation of the present study, we used a pruned version of the risk calculator model (i.e. excluding stressful life events and trauma history, which were not significant predictors in the original risk calculator), of which a similar version has been used in prior studies expanding on the original risk calculator findings (Chung et al., 2019a; Worthington et al., 2021). After pruning, the primary model included age, SIPS items P1 + P2, HVLT-R score, BACS SC score, change in social functioning, and family history of a psychotic disorder.

We first applied the pruned version of the risk calculator in the full training sample to rediscover the model coefficients in this

combined sample and validated how well the model predicted eventual converters in the full validation sample. We then repeated this process within both the short and long duration training/validation samples. Performance metrics for predicting eventual converters in the validation samples were assessed using area under the curve (AUC, 95% confidence interval constructed using 2000 bootstrap resamples) to measure correct discrimination of eventual converters and non-converters, sensitivity to measure the true positive converter cases, specificity to measure the true negative non-converter cases, and balanced accuracy (BAC) as the mean of sensitivity and specificity. Model calibration was assessed using the Brier score and calibration slope (Brier, 1950). In all models and subsamples, the mean level of predicted risk as defined in the training sample was used as a cut-point for determining prediction accuracy.

Incorporating cortical thickness

To test the potential added value of cortical thickness measures, we assessed the increase in predictive power of baseline cortical thickness and rate of cortical thinning (e.g. percent change in gray matter thickness over an average of 3 months) to the clinical model for a subset of NAPLS3 participants who had completed follow-up imaging assessments. This analysis was limited to NAPLS3 participants due to the frequency of follow-up imaging visits available for this sample as compared to NAPLS2 (i.e. imaging assessments were conducted every 2 months in NAPLS3 and every 12 months in NAPLS2) and in line with the goal of assessing short-term cortical changes to predict future outcomes. Risk scores for each individual in the full, short duration, and long duration samples were calculated from the clinical models described above. The resulting risk score was included in three different Cox proportional hazard regression models with and without neuroimaging measures to predict conversion to psychosis. These models were: (1) risk score only; (2) risk score plus baseline cortical thickness; and (3) risk score plus percent change in cortical thickness. Due to the smaller sample sizes and number of converters after stratifying this sample, bootstrap resampling was used to internally validate these models in place of the train/test split used in the prior analysis step.

Results

A total of 1300 participants from the NAPLS2 and NAPLS3 studies who completed any follow-up data were included. Of these, 11.8% experienced eventual conversion to psychosis. The median duration since recent positive symptom increase was 120 days. After sample stratification based on this criterion, the long symptom duration group consisted of 621 participants (9.8% conversion rate) and the short symptom duration group consisted of 618 participants (13.7% conversion rate). Overall, the short duration group was younger, had fewer years of education, and had higher levels of baseline Scale of Prodromal Symptoms (SOPS) items P1 + P2 (unusual thought content and suspiciousness) than the longer duration group. See Table 1 for full results. In both groups, converters had higher ratings on unusual thought content and suspiciousness, worse cognitive functioning on the BACS SC and HVLT-R tests, and a higher proportion had a family history of a psychotic disorder. In the long duration group, but not the short duration group, converters experienced greater decline in GFS functioning and the converter group consisted of a higher proportion of racial minorities than the non-converter group. See online Supplementary eTable 1 for full results.

Table 1. Mean comparisons of demographic and risk calculator variables for long and short symptom duration groups in the combined NAPLS2/NAPLS3 sample

	Long symptom duration			Short symptom duration			Test statistics	
	<i>n</i>	Mean	s.d.	<i>n</i>	Mean	s.d.	<i>t</i> / χ^2 (df)	<i>p</i> value
Demographics								
Age	621	18.56	4.3	618	18.02	4.0	<i>t</i> (1174.3) = 1.97	0.05*
Sex at birth ^a , No. (%) female	621	227	44.6%	618	273	44.2%	χ^2 (1) = 1.66	0.20
Race, No. (%) non-white	619	265	42.7%	618	281	45.5%	χ^2 (1) = 0.89	0.35
Ethnicity, No. (%) Hispanic	619	118	19%	618	130	21%	χ^2 (1) = 0.80	0.37
Years of education	621	11.53	3.0	616	11.16	2.9	<i>t</i> (1234.4) = 2.20	0.03*
Other predictor variables								
SOPS items P1 + P2	589	2.82	1.5	593	3.00	1.5	<i>t</i> (1179.5) = -2.12	0.03*
BACS raw score	589	55.46	13.5	593	55.28	12.9	<i>t</i> (1177.4) = 0.23	0.82
HVLT-R total	589	25.88	5.2	593	25.97	5.1	<i>t</i> (1179.2) = -0.30	0.76
Family history, No. (%) yes	589	66	11.2%	593	74	12.5%	χ^2 (1) = 0.45	0.50
Decline in social functioning	589	0.61	0.9	593	0.70	0.9	<i>t</i> (1179) = -1.65	0.09

SOPS P1 + P2, Scale of Prodromal Symptoms: unusual thought content and suspiciousness; BACS, brief assessment of cognition in schizophrenia; HVLT-R, Hopkins Verbal Learning Test - Revised.

^aAll participants demographic characteristics are self-report.

*Asterisk indicates that mean differences between groups are significant at the $p < 0.05$ level.

When the pruned risk calculator model was applied to the full sample, lower HVLT-R score and higher unusual thought content and suspiciousness scores significantly predicted conversion at the multivariable level. This pattern was seen in the short symptom duration group as well. In the long symptom duration group, greater decline in social functioning and higher levels of unusual thought content and suspiciousness significantly predicted conversion at the multivariable level. See Table 2 for full results.

Model performance was tested in the validation sample using the mean predicted risk scores as a cutoff for conversion prediction in each subsample. In the full sample, the risk calculator achieved an AUC of 0.71 (BAC = 0.64, calibration slope = 1.29). In the long duration group, the model achieved an AUC of 0.74 (BAC = 0.65, calibration slope = 0.92) and in the short duration group, the model achieved an AUC of 0.69 (BAC = 0.69, calibration slope = 1.58). Full results are reported in Table 3. This level of performance is consistent with the performance of existing risk

calculators, notably the original NAPLS2 risk calculator (Cannon et al., 2016). Distributions of predicted risk scores for each stratification group are shown in Fig. 1, which clearly shows the differences in the shapes and central tendencies of risk distributions across shorter and longer duration cases. Receiver operating characteristic curves describing model performance are shown in Fig. 2.

A total of 274 participants (13.4% conversion rate) from NAPLS3 completed follow-up structural MRI scans and were included in the analysis assessing the added value of percent change in cortical thickness as a predictor of conversion to psychosis alongside the clinical risk calculator. When stratified based on symptom duration, the long symptom duration imaging sample consisted of 136 participants (13.2% conversion rate) and the short symptom duration imaging sample consisted of 126 participants (13.5% conversion rate). See online Supplement for group comparisons between short and long symptom duration groups.

Table 2. Statistics for individual predictor variables in the Cox proportional hazard regression pruned risk calculator model predicted conversion to psychosis

	Full sample			Short duration group			Long duration group		
	HR	95% CI	<i>p</i> value	HR	95% CI	<i>p</i> value	HR	95% CI	<i>p</i> value
Age	0.99	0.94–1.05	0.66	0.92	0.85–1.02	0.14	1.02	0.95–1.11	0.59
SIPS items P1 + P2	1.38	1.19–1.64	<0.001**	1.51	1.30–1.94	0.001**	1.30	1.00–1.67	0.04*
BACS raw score	0.99	0.97–1.01	0.68	1.00	0.97–1.03	0.94	0.98	0.95–1.02	0.15
HVLT total	0.95	0.91–0.99	0.03*	0.93	0.87–0.99	0.03*	0.96	0.89–1.03	0.30
Family history, No. (%) yes	1.27	0.71–2.36	0.42	1.92	0.89–3.50	0.11	1.10	0.39–3.35	0.85
GFS decline	1.21	0.88–1.44	0.09	0.80	0.59–1.14	0.29	1.76	1.21–2.09	0.005**

HR, hazard ratio; CI, confidence interval; SOPS items P1 + P2, Scale of Prodromal Symptoms: unusual thought content and suspiciousness; BACS, brief assessment of cognition in schizophrenia; HVLT-R, Hopkins Verbal Learning Test - Revised.

*A single asterisk indicates the predictor variable is significant in the multivariable model at the $p < 0.05$ level.

**A double asterisk indicates the predictor variable is significant in the multivariable model at the $p < 0.01$ level.

Table 3. Performance of models predicting conversion to psychosis using the pruned NAPLS2 risk calculator in the combined NAPLS2/NAPLS3 sample and the predicted risk scores plus cortical thickness measures in the NAPLS3 imaging sample

	AUC	95% CI	Sensitivity	Specificity	BAC	Mean risk score cutoff
Risk calculator model						
Full sample (validation set $n = 650$)	0.69	0.62–0.76	0.63	0.65	0.64	0.12
Short symptom duration (validation set $n = 306$)	0.70	0.61–0.78	0.65	0.65	0.65	0.14
Long symptom duration (validation set $n = 305$)	0.70	0.59–0.81	0.71	0.66	0.69	0.09
<i>With imaging variables (sample $n = 274$)</i>						
Full sample						
Risk score only	0.71	0.62–0.80	0.54	0.72	0.63	0.13
Risk score + baseline cortical thickness	0.71	0.62–0.79	0.57	0.74	0.65	0.13
Risk score + percent difference in cortical thickness	0.78	0.72–0.86	0.69	0.75	0.72	0.13
Short duration						
Risk score only	0.71	0.58–0.85	0.65	0.68	0.66	0.13
Risk score + baseline cortical thickness	0.70	0.55–0.84	0.65	0.68	0.66	0.13
Risk score + percent difference in cortical thickness	0.84	0.75–0.92	0.71	0.75	0.73	0.13
Long duration						
Risk score only	0.71	0.58–0.83	0.44	0.74	0.59	0.12
Risk score + baseline cortical thickness	0.71	0.59–0.83	0.44	0.74	0.56	0.12
Risk score + percent difference in cortical thickness	0.78	0.65–0.90	0.72	0.76	0.74	0.12

AUC, area under the curve; CI, confidence interval; BAC, balanced accuracy.

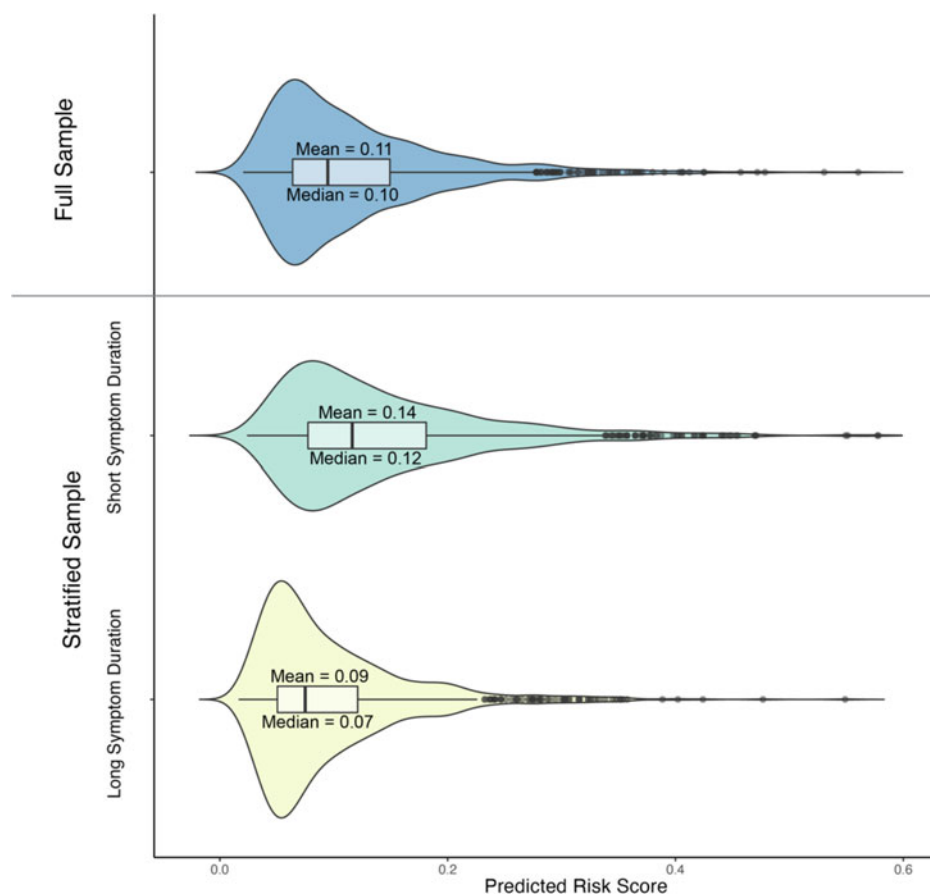


Figure 1. Predicted risk of conversion in the full NAPLS2/NAPLS3 sample, short symptom duration sample, and long symptom duration sample.

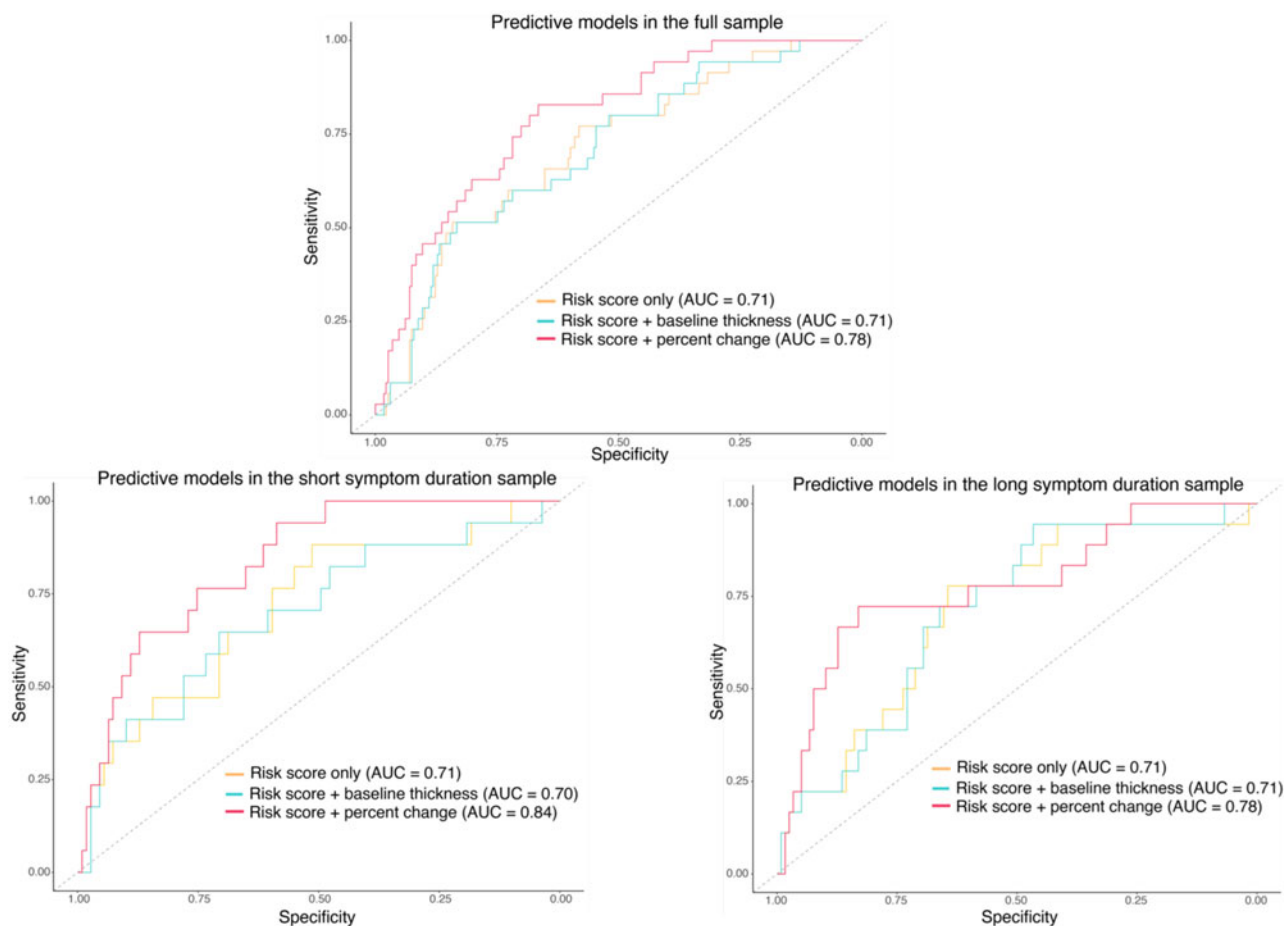


Figure 2. AUC for performance of predictor models without/with imaging variables.

In the full imaging sample, the model including only the clinical risk score achieved an AUC of 0.70 (BAC = 0.63, calibration slope = 0.83), which was comparable to the performance in the full NAPLS2/NAPLS3 sample. The model consisting of the clinical risk score and baseline cortical thickness achieved an AUC of 0.71 (BAC = 0.65, calibration slope = 0.83) and the model consisting of the clinical risk score and the percent change in cortical thickness achieved an AUC of 0.78 (BAC = 0.72, calibration slope = 0.92). In the long symptom duration group, the risk score only model achieved an AUC of 0.71 (BAC = 0.59, calibration slope = 0.99), the risk score and baseline cortical thickness model achieved an AUC of 0.71 (BAC = 0.56, calibration slope = 1.04), and the risk score and percent change in cortical thickness model achieved an AUC of 0.78 (BAC = 0.74, calibration slope = 0.98). In the short duration group, the risk score only model achieved an AUC of 0.71 (BAC = 0.66, calibration slope = 1.02). The risk score and baseline cortical thickness model achieved an AUC of 0.70 (BAC = 0.66, calibration slope = 1.04) while the risk score and percent change in cortical thickness achieved an AUC of 0.84 (BAC = 0.73, calibration slope = 1.14). For full results of model performance, see Table 3 and online Supplementary eTable 5. In the multivariable models that included imaging measures, only the measure of cortical thinning, not the measure of baseline cortical thickness, was a significant predictor of conversion in addition to the risk score for each of the samples. In the short duration group, the addition of cortical thinning to the predicted risk score results in a significant improvement in AUC and

a notable improvement in sensitivity and specificity as compared to the original risk calculator model applied in the short duration group of the combined NAPLS2/NAPLS3 sample. In the long duration group, the addition of cortical thinning represents a moderate improvement in performance as compared to the original risk calculator model applied in the long duration group of the combined NAPLS2/NAPLS3 sample, with improved AUC metrics and notably improved sensitivity and specificity.

Discussion

This is the first study to demonstrate a potential moderator effect of the NAPLS2 risk calculator that could explain differences in risk distributions across cohorts, which was also validated with a neuroimaging measure in an NAPLS3 subgroup. CHR-P participants with a shorter duration between APS increase and ascertainment may best align with risk distributions conferring higher risk of conversion (mean risk score = 0.14) and patients with a longer duration between prodromal symptom increase and ascertainment may best align with risk distributions conferring a lower risk of conversion (mean risk score = 0.09). As observed during the effort to replicate the NAPLS2 risk calculator in the PRONIA study, the mean risk scores of the short symptom duration group and the long symptom duration group in this study map on to the mean risk scores of the NAPLS2 study and the PRONIA study, respectively (Koutsouleris et al., 2021) (i.e. prior to adjusting for sample differences in symptom severity and other risk factors), suggesting

that this symptom duration approach to stratification may be generally applicable to CHR-P populations.

Between the long symptom duration and short symptom duration groups, mean differences in demographic and clinical variables in addition to differences in significant predictor variables at the multivariable level suggest that these groups represent notably different clinical presentations. The shorter symptom duration group may represent an acute onset and worsening of symptoms, with the presence of a decline in cognitive functioning, whereas the longer symptom duration group may represent a steadier decline in both symptoms and social functioning.

Further, in the NAPLS3 subgroup, we found that decrease in cortical thickness adds to the predictive power of the risk calculator for all CHR-P patients (AUC improvement from 0.71 to 0.78), but that it differentially and significantly improves prediction in the shorter symptom duration group (AUC improvement from 0.71 to 0.84) as compared to the longer symptom duration group (AUC improvement from 0.71 to 0.78). Notably, this improvement was not seen when the baseline cortical thickness measure was added to the risk calculator. These findings further support the validity of the distinction between the symptom duration groups and suggest that cortical thinning as a mechanism of psychosis development not only maps on well to clinical indicators of illness progress, but also adds significant power to the prediction of psychosis onset. In the short symptom duration group, which experiences a higher rate of conversion to psychosis, underlying cortical changes may contribute to a more rapid and severe onset of prodromal symptoms, and eventual full-blown psychosis, than in the long symptom duration group.

Beyond neurodevelopmental changes, it will be important to determine underlying social, environmental, and clinical factors that may contribute to these different clinical presentations. There is some evidence to suggest that pathways to ascertainment and referral sources to specialized CHR-P clinics result in risk enrichment heterogeneity (Fusar-Poli et al., 2016a, 2016b). These pathways and referral sources may represent underlying factors such as access to affordable mental healthcare, healthcare resource utilization, and family support in seeking mental health treatment, which could interact with underlying neurodevelopmental changes (Patel, Leathem, Currin, & Karlsgodt, 2021) to result in the clinical presentations observed in this study. Future studies incorporating measures of social support, systematic support, and help-seeking behaviors could further elucidate the nuances contributing to the timing of seeking treatment for prodromal symptoms.

In addition to improving our ability to predict outcomes, the observed patterns of clinical and neurodevelopmental changes could provide insight into treatment selection for CHR-P individuals. Currently, no psychosocial or pharmacological intervention for preventing or delaying the onset of psychosis has emerged as a preferred treatment (Addington, Devoe, & Santesteban-Echarri, 2019), likely in part because of the heterogeneity in CHR-P cohorts not accounted for in intervention studies. Parsing this heterogeneity through improved prediction of outcomes could yield more precise results from treatment studies and tailored interventions based on factors that can be ascertained at or near a baseline visit (Worthington & Cannon, 2021). Further investigation would be warranted to determine whether different interventions could be appropriate for the groups described in this study.

Strengths and limitations

A significant strength of this study was the ability to leverage the very large sample size of the combined NAPLS2/NAPLS3 studies.

This allowed us to not only account for differences across the two samples, but also create a training and validation framework to test the risk calculator models in the full and stratified samples. Nonetheless, it will be important to externally validate these findings in an independent sample. The motivation from this study stemmed from the observed differences in conversion rates and predicted risk scores across the NAPLS2 and PRONIA studies; thus, determining whether the moderator proposed in this study validates across these samples will be an important next step. Because we tested the models with neuroimaging measures only in the NAPLS3 sample, the sample size was significantly reduced, thus predisposing the prediction models to potential overfitting. Further, sample differences between those who completed imaging in the smaller NAPLS3 sample may potentially bias results. The findings incorporating short-term cortical thinning should be externally validated in a larger sample to address sample bias and ensure stability of these results.

Conclusion

This study is the first to provide evidence for a potential moderator of risk prediction in the CHR-P population which could aid clinical implementation of risk calculators to match new patients with appropriate risk distributions. Neuroimaging measures significantly bolstered prediction in the short symptom duration group and improved prediction in the long symptom duration group, and clinical differences were identified between these two groups. These patterns of results suggest that illness progression prior to and shortly after ascertainment not only differs among CHR-P patients, but also could play an essential role in the prediction of outcomes and determination of appropriate risk distribution for prognostication of new cases.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291723002301>

Financial support. This study was supported by the National Institute of Mental Health (grant U01MH081984 to Dr Addington; grant U01MH081928 to Dr Stone; grant U01MH081944 to Dr Cadenhead; grant U01MH081902 to Drs Cannon and Bearden; grant U01MH082004 to Dr Perkins; grant U01MH081988 to Dr Walker; grant U01MH082022 to Dr Woods; grant U01MH076989 to Dr Mathalon; grant U01MH081857 to Dr Cornblatt). This research was supported in part by a grant from the American Psychological Association to Ms. Worthington.

Competing interest. Dr Mathalon is a consultant for Boehringer-Ingelheim, Cadent Therapeutics, Recognify, and Syndisi. Dr Woods reports that during the last 36 months he has received sponsor-initiated research funding support from Teva, Boehringer-Ingelheim, Amarex, and SyneuRx. He has consulted to Boehringer-Ingelheim, New England Research Institute, and Takeda. He has been granted US patent no. 8492418 B2 for a method of treating prodromal schizophrenia with glycine agonists and has received royalties from Oxford University Press. All other authors report no financial relationships with commercial interests.

Ethical standard. 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

Addington, J., Cadenhead, K. S., Cannon, T. D., Cornblatt, B., McGlashan, T. H., Perkins, D. O., ... Woods, S. W. (2007). North American prodrome

- longitudinal study: A collaborative multisite approach to prodromal schizophrenia research. *Schizophrenia Bulletin*, 33(3), 665–672.
- Addington, J., Cadenhead, K. S., Cornblatt, B. A., MATHALON, D. H., McGlashan, T. H., Perkins, D. O., ... Woods, S. W. (2012). North American prodrome longitudinal study (NAPLS 2): Overview and recruitment. *Schizophrenia Research*, 142(1–3), 77–82.
- Addington, J., Devoe, D. J., & Santesteban-Echarri, O. (2019). Multidisciplinary treatment for individuals at clinical high risk of developing psychosis. *Current Treatment Options in Psychiatry*, 6(1), 1–16.
- Addington, J., Liu, L., Brummitt, K., Bearden, C. E., Cadenhead, K. S., Cornblatt, B. A., ... Cannon, T. D. (2022). North American prodrome longitudinal study (NAPLS 3): Methods and baseline description. *Schizophrenia Research*, 243, 262–267.
- Benedict, R. H., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test-Revised: Normative data and analysis of inter-form and test-retest reliability. *Clinical Neuropsychologist*, 12(1), 43–55.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, 78(1), 1–3.
- Cannon, T. D., Cadenhead, K., Cornblatt, B., Woods, S. W., Addington, J., Walker, E., ... McGlashan, T. (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.
- Cannon, T. D., Chung, Y., He, G., Sun, D., Jacobson, A., van Erp, T. G. M., ... Cadenhead, K. (2015). Progressive reduction in cortical thickness as psychosis develops: A multisite longitudinal neuroimaging study of youth at elevated clinical risk. *Biological Psychiatry*, 77(2), 147–157.
- Cannon, T. D., Yu, C., Addington, J., Bearden, C. E., Cadenhead, K. S., Cornblatt, B. A., ... McGlashan, T. H. (2016). An individualized risk calculator for research in prodromal psychosis. *American Journal of Psychiatry*, 173(10), 980–988.
- Carrión, R. E., Cornblatt, B. A., Burton, C. Z., Tso, I. F., Auther, A. M., Adelsheim, S., ... McFarlane, W. R. (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.
- Chung, Y., Addington, J., Bearden, C. E., Cadenhead, K., Cornblatt, B., MATHALON, D. H., ... Cannon, T. D. (2019a). Adding a neuroanatomical biomarker to an individualized risk calculator for psychosis: A proof-of-concept study. *Schizophrenia Research*, 208, 41–43. doi: 10.1016/j.schres.2019.01.026
- Chung, Y., Allswede, D., Addington, J., Bearden, C. E., Cadenhead, K., Cornblatt, B., ... Cannon, T. D. (2019b). Cortical abnormalities in youth at clinical high-risk for psychosis: Findings from the NAPLS2 cohort. *NeuroImage: Clinical*, 23, 1–13. doi: 10.1016/j.nicl.2019.101862
- Collins, M. A., Ji, J. L., Chung, Y., Lympus, C. A., Afriyie-Agyemang, Y., Addington, J. M., ... Cannon, T. D. (2022). Accelerated cortical thinning precedes and predicts conversion to psychosis: The NAPLS3 longitudinal study of youth at clinical high-risk. *Molecular Psychiatry*, 28(3), 1182–1189.
- 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.
- Dohrenwend, B. S., Askenasy, A. R., Krasnoff, L., & Dohrenwend, B. P. (1978). Exemplification of a method for scaling life events: The PERI Life Events Scale. *Journal of Health and Social Behavior*, 19(2), 205–229.
- 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., Rutigliano, G., Stahl, D., Schmidt, A., Ramella-Cravaro, V., Hitesh, S., & McGuire, P. (2016a). Deconstructing pretest risk enrichment to optimize prediction of psychosis in individuals at clinical high risk. *JAMA Psychiatry*, 73(12), 1260–1267.
- Fusar-Poli, P., Schultze-Lutter, F., Cappucciati, M., Rutigliano, G., Bonoldi, I., Stahl, D., ... Perkins, D. O. (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.
- Germann, M., Brederoo, S. G., & Sommer, I. E. (2021). Abnormal synaptic pruning during adolescence underlying the development of psychotic disorders. *Current Opinion in Psychiatry*, 34(3), 222.
- Harrell, F. E. (2021). rms: Regression Modeling Strategies. R package version 6.2-0. <https://CRAN.R-project.org/package=rms>
- Janssen, I., Krabbendam, L., Bak, M., Hanssen, M., Vollebergh, W., de Graaf, R., & van Os, J. (2004). Childhood abuse as a risk factor for psychotic experiences. *Acta Psychiatrica Scandinavica*, 109(1), 38–45.
- Keefe, R. S., Harvey, P. D., Goldberg, T. E., Gold, J. M., Walker, T. M., Kennel, C., & Hawkins, K. (2008). Norms and standardization of the brief assessment of cognition in schizophrenia (BACS). *Schizophrenia Research*, 102(1–3), 108–115.
- Koutsouleris, N., Davatzikos, C., Borgwardt, S., Gaser, C., Bottlender, R., Frodl, T., ... Reiser, M. (2014). Accelerated brain aging in schizophrenia and beyond: A neuroanatomical marker of psychiatric disorders. *Schizophrenia Bulletin*, 40(5), 1140–1153.
- Koutsouleris, N., Kambeitz-Illankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D. B., ... Consortium, P. (2018). Prediction models of functional outcomes for individuals in the clinical high-risk state for psychosis or with recent-onset depression: A multimodal, multisite machine learning analysis. *JAMA Psychiatry*, 75(11), 1156–1172.
- Koutsouleris, N., Worthington, M., Dwyer, D. B., Kambeitz-Illankovic, L., Sanfelici, R., Fusar-Poli, P., ... Addington, J. (2021). Towards generalizable and transdiagnostic tools for psychosis prediction: An independent validation and improvement of the NAPLS-2 risk calculator in the multi-site PRONIA cohort. *Biological Psychiatry*, 90(9), 632–642.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... Kenkel, B. (2020). caret: Classification and Regression Training. R package version 6.0-86. <https://cran.r-project.org/web/packages/caret/caret.pdf>
- McGlashan, T. H., Walsh, B., & Woods, S. (2010). *The psychosis-risk syndrome: Handbook for diagnosis and follow-up*. New York, NY: Oxford University Press.
- McGlashan, T. H., Walsh, B. C., & Woods, S. W. (2001). *Structured interview for psychosis-risk syndromes*. New Haven, CT: Yale School of Medicine.
- 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.
- Miller, T. J., McGlashan, T. H., Rosen, J. L., Cadenhead, K., Ventura, J., McFarlane, W., ... 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.
- Mogensen, U. B., Ishwaran, H., & Gerds, T. A. (2012). Evaluating random forests for survival analysis using prediction error curves. <https://www.jstatsoft.org/v50/i11>
- 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.
- Patel, P. K., Leatham, L. D., Currin, D. L., & Karlsgodt, K. H. (2021). Adolescent neurodevelopment and vulnerability to psychosis. *Biological Psychiatry*, 89(2), 184–193.
- R Core Team. (2021). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Robin, X. R., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 12, 77.
- 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.
- Worthington, M. A., & Cannon, T. D. (2021). Prediction and prevention in the clinical high-risk for psychosis paradigm: A review of the current status and recommendations for future directions of inquiry. *Frontiers in Psychiatry*, 12, 1–10.
- Worthington, M. A., Walker, E. F., Addington, J., Bearden, C. E., Cadenhead, K. S., Cornblatt, B. A., ... Cannon, T. D. (2021). Incorporating cortisol into the NAPLS2 individualized risk calculator for prediction of psychosis. *Schizophrenia Research*, 227, 95–100.

- Yung, A. R., Yung, A. R., Pan Yuen, H., McGorry, P. D., Phillips, L. J., Kelly, D., ... Killackey, E. (2005). Mapping the onset of psychosis: The comprehensive assessment of at-risk mental states. *Australian & New Zealand Journal of Psychiatry*, 39(11–12), 964–971.
- Zhang, T., Li, H., Tang, Y., Niznikiewicz, M. A., Shenton, M. E., Keshavan, M. S., ... Wang, J. (2018a). 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.
- Zhang, T., Xu, L., Tang, Y., Li, H., Tang, X., Cui, H., ... Wang, J. (2018b). Prediction of psychosis in prodrome: Development and validation of a simple, personalized risk calculator. *Psychological Medicine*, 49(12), 1900–1998.