The relationship between cognitive deficits and impaired short-term functional outcome in clinical high-risk for psychosis participants: A machine learning and modelling approach.

Kate Haining, MSc[1*], Gina Brunner, MSc[1*], Ruchika Gajwani, Ph.D.[2] , Joachim Gross, Ph.D.[1], Andrew I. Gumley, Ph.D.[2], Stephen M. Lawrie M.D.[3], Matthias Schwannauer, Ph.D.[4], Frauke Schultze-Lutter, Ph.D.[5], & Peter J. Uhlhaas, Ph.D.[1,6]

1. Institute for Neuroscience and Psychology, Univ. of Glasgow

2. Institute of Health and Wellbeing, Univ. of Glasgow

3. Department of Psychiatry, Univ. of Edinburgh

4. Department of Clinical Psychology, Univ. Edinburgh

5. Department of Psychiatry and Psychotherapy, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany

6. Department of Child and Adolescent Psychiatry, Charité Universitätsmedizin, Berlin, Germany

* these authors contributed equally

Corresponding Author:

Dr. Peter J. Uhlhaas

Department of Child and Adolescent Psychiatry

Charité Universitätsmedizin

Berlin, Germany

Email: peter.uhlhaas@charit.de

Tel: 0049 30 450 516 193

Abstract: 250 words

Main text: 3936 words

Figures: 1

Tables: 4

Abstract

Poor functional outcomes are common in individuals at clinical high-risk for psychosis (CHR-P), but the contribution of cognitive deficits remains unclear. We examined the potential utility of cognitive variables in predictive models of functioning at baseline and follow-up with machine learning methods. Additional models fitted on baseline functioning variables were used as a benchmark to evaluate model performance.

Data were available for 146 CHR-P individuals of whom 118 completed a 6- and/or 12-month follow-up; as well as 47 participants not fulfilling CHR criteria (CHR-Ns) but displaying affective and substance use disorders; and 55 healthy controls (HCs). Predictors of baseline global assessment of functioning (GAF) scores were selected by L1-regularised least angle regression and then used to train various classifiers, evaluated with 10-fold cross-validation, to predict functional outcome in CHR-P individuals.

In CHR-P participants, cognitive deficits together with clinical and functioning variables explained 41% of the variance in baseline GAF scores while cognitive variables alone explained 12%. These variables allowed classification of functional outcome with an average balanced accuracy (BAC) of 63% in both mixed- and cross-site models. However, higher accuracies (68%-70%) were achieved using classifiers fitted only on baseline functioning variables.

Our findings suggest that cognitive deficits, alongside clinical and functioning variables, displayed robust relationships with impaired functioning in CHR-P participants at baseline and follow-up. Moreover, these variables allow for prediction of functional outcome. However, models based on baseline functioning variables showed a similar performance, highlighting the need to develop more accurate algorithms for predicting functional outcome in CHR-P participants.

1. Introduction

Psychotic disorders, such as schizophrenia, continue to pose a significant challenge for the field given that many patients experience poor outcomes and the absence of significant advances in treatments over the last decades (Millan et al., 2016; Owen et al., 2016). Schizophrenia may be preceded by a clinical high-risk for psychosis (CHR-P) state lasting approximately 5-6 years (Schultze-Lutter et al., 2015) and clinical criteria have been developed to detect individuals prior to the onset of full-blown psychosis (Fusar-Poli et al., 2013). CHR-P criteria include attenuated psychotic symptoms, brief frank psychosis and functional decline with genetic risk (Yung et al., 2005) as well as self-experienced perceptual and cognitive anomalies known as basic symptoms (Schultze-Lutter, 2009; Schultze-Lutter et al., 2012). Approximately 20% of individuals meeting CHR-P criteria will transition to psychosis within a 2-year period (Fusar-Poli et al., 2016). Moreover, around 40-50% of nonconverters continue to experience impairments in social and role functioning (Carrión et al., 2013; Koutsouleris et al., 2018). Therefore, understanding the underlying factors as well as predictors of poor functioning in CHR-P individuals is an important objective for early detection and intervention.

Negative symptoms, disorganised symptoms, impairments in social and role functioning and poor premorbid psychosocial adjustment have been found to predict poor baseline functioning and/or poor functional outcome at follow-up (Carrión et al., 2013; Glenthøj et al., 2016; Koutsouleris et al., 2018; Salokangas et al., 2014). Although positive symptom severity is predictive of transition to psychosis, effects on functioning remain inconsistent (Carrión et al., 2016; Meyer et al., 2014).

While there is emerging evidence for a relationship between cognitive deficits and impaired functioning in CHR-P individuals, the contribution of specific cognitive deficits varies across

studies. Cognitive deficits, predominantly in verbal memory, are an established mediator of functional outcomes in chronic schizophrenia (Green, 1996; Schmidt et al., 2011). Interestingly, in studies of early psychosis, reasoning, problem solving and motor skills more frequently predict short-term (< 2 years) functional outcome while language/verbal skills and global/general cognition more often predict longer-term (> 2 years) functional outcome (Allott et al., 2011). In CHR-P individuals specifically, impairments in verbal memory, emotion recognition and processing speed have been linked with impairments in social and/or role functioning at baseline and follow-up (Carrión et al., 2011, 2013; Glenthøj et al., 2016; Lin et al., 2011; Modinos et al., 2019; Niendam et al., 2006). Moreover, deficits in verbal learning and fluency, motor speed and executive function have also been associated with poor functioning in CHR-P individuals (Carrión et al., 2013; Eslami et al., 2011; Lin et al., 2011; Niendam et al., 2006).

In the current study, we sought to clarify the contribution of cognitive deficits towards impaired functioning in CHR-participants. To identify predictors of functioning, we employed a machine learning approach in which we first identified variables associated with baseline functioning using LASSO-LARS regression and then predicted functional outcome at follow-up with classifiers evaluated using 10-fold cross-validation and permutation testing. While machine learning studies have previously shown potential for identifying predictors of transition to psychosis as well as functional outcomes based on clinical, functional and neuroimaging data (Kambeitz-Ilankovic et al., 2016; Koutsouleris et al., 2009, 2012, 2018), a considerable proportion have also failed to provide convincing evidence (Fusar-Poli et al., 2019; Mechelli et al., 2017; Ramyead et al., 2016). Furthermore, previous studies predicting functional outcome using cognitive measures have applied more traditional logistic regressions without cross-validation or regularisation techniques, potentially carrying a risk of overfitting (Carrión et al., 2013; Eslami et al., 2011; Lin et al., 2011; Meyer et al., 2014;

Modinos et al., 2019). Even in machine learning studies leveraging these techniques, few have attempted to compare their multi-step machine learning pipelines to simpler predictive models in order to justify this added complexity (DeMasi et al., 2017).

To address these gaps, we firstly examined the contribution of clinical, functioning and cognitive variables to impaired functioning at baseline in CHR-P participants. We also included a sample of participants who did not fulfil CHR-P criteria but were characterised by mood, anxiety and substance use (i.e. alcohol and drug) disorders (CHR-Ns), as well as healthy controls (HCs). We then applied a machine learning approach to those variables associated with impaired functioning at baseline in order to predict short-term functional outcome. We additionally created simpler predictive models of functional outcome using only baseline functioning variables to determine whether our more complex machine learning pipeline provided a significant increase in predictive performance. Given the contribution of cognitive impairment to impaired functioning in established schizophrenia (Green, 1996; Schmidt et al., 2011), we hypothesised that the inclusion of cognitive variables in machine learning models would enhance the prediction of functional outcome in CHR-P participants, outperforming simpler models.

2. Methods

2.1. Participants

The data were collected as part of the Youth Mental Health Risk and Resilience (YouR) study (Uhlhaas et al., 2017), an ongoing longitudinal study funded by the Medical Research Council (MRC), which aims to identify neurobiological and psychological mechanisms and predictors of psychosis risk. CHR-P participants were recruited through an online-screening approach (www.your-study.org.uk; McDonald et al., 2019) and via referrals from NHS

patient services and student counselling services. CHR-N participants (N = 47) were also recruited using the former approach while HCs (N = 55) were obtained from an existing volunteer database. CHR-N participants were recruited to allow for a more meaningful comparison with the CHR-P group (Millman et al., 2019). By including participants with affective and substance use disorders (CHR-N group), we aimed to separately assess the impact of psychiatric comorbidity given that such comorbidity is also characteristic of the CHR-P state. Recruitment and assessment visits/ratings were carried out by trained research assistants and MSc/PhD level researchers.

Data were available for 146 CHR-P individuals that were recruited across two sites: Glasgow (n = 109; 74.7%) and Edinburgh (n = 37; 25.3%). One hundred and eighteen participants (80.8%) completed a follow-up session at 6 and/or 12 months. Attrition rates were similar across sites (Glasgow: 20.2%; Edinburgh: 16.2%).

2.2. Baseline Assessments

The positive scale of the Comprehensive Assessment of At-Risk Mental States (CAARMS; Yung et al., 2005) and the Cognitive Disturbances (COGDIS) and Cognitive-Perceptive Basic Symptoms (COPER) items of the Schizophrenia Proneness Instrument, Adult version (SPI-A; Schultze-Lutter et al., 2007) were administered to all participant groups. Participants were recruited into the CHR-P group if they met SPI-A COGDIS/COPER criteria or one of the following CAARMS criteria: Attenuated Psychosis Symptoms (APS), Genetic Risk and Deterioration Syndrome (GRD) or Brief Limited Intermittent Psychotic Symptoms (BLIPS). All participants were also assessed with the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998), Global Assessment of Functioning (GAF) scale from the DSM-IV-TR, Global Functioning: Social (GF: Social) and Role (GF: Role) scales (Cornblatt et al., 2007), Premorbid Adjustment Scale (PAS; Cannon-Spoor et al., 1982) and Adverse

Childhood Experiences Scale (ACES; Felitti et al., 1998). Neuropsychological assessments consisted of the Brief Assessment of Cognition in Schizophrenia (BACS; Keefe et al., 2004) and three tasks from the Penn Computerized Neurocognitive Battery (CNB; Moore et al., 2015): the Continuous Performance Test, the N-Back Test and the Emotion Recognition Task which provide measures of accuracy and response time (RT) for attention, working memory and emotion recognition respectively.

2.3. Follow-Up Assessments

Follow-up interviews were conducted at 6- and 12-months following the baseline assessments for the CHR-P and CHR-N groups and involved the positive scale of the CAARMS as well as the GAF, GF: Social and GF: Role scales. The HC group did not complete follow-up assessments.

2.4. Statistical Analysis

2.4.1. Preprocessing

Data were analysed using Python (3.7) packages Numpy, Pandas and Scikit-learn (SKL). In accordance with Keefe et al. (2008), BACS raw scores were converted into standardized z-scores using the means and standard deviations of gender-specific HCs. This gender correction was applied because gender has been shown to affect BACS performance in a normative sample (Keefe et al., 2008). For consistency, CNB raw scores were calculated in the same way, albeit without correction for gender. CAARMS severity was calculated by multiplying the global score by the frequency score for each of the four domains and summing these products while SPI-A severity was calculated by summing the frequency scores for each basic symptom. Where participants did not experience a symptom, the

associated frequency and distress were set to zero while those with missing data for the outcome variable were removed. Participants and variables with < 70% of the measures of interest were removed and missing values were imputed by Bayesian Ridge regression. Additional columns were generated for GAF scores to define good (GFO) or poor functional outcome (PFO), whereby GAF scores below 65 were coded as PFO. In line with prior studies (Allen et al., 2015; Modinos et al., 2019), this cut-off was selected because the 61-70 range corresponds to the presence of "some difficulty in social, occupational or school functioning but [the person] has some meaningful interpersonal relationships". We additionally calculated how many participants changed GAF category between baseline and follow-up as well as GAF changes over time.

2.4.2. Group comparisons

Group differences were analysed using non-parametric Kruskal-Wallis H tests or Mann–Whitney U tests and chi-square tests, followed by appropriate post hoc tests if required.

2.4.3. Regression Analysis

To determine which variables were associated with baseline GAF scores in CHR-P and CHR-N groups, we fitted combined and cognitive models, whereby the former included clinical, cognitive and functioning variables. We used L1-regularised least angle regression (LASSO-LARS; Efron et al., 2004), with 10-fold cross-validation, as implemented in the SKL function LassoLarsCV. This method is particularly appropriate for addressing the high dimensionality of our candidate predictor set (Fonti, 2017). We excluded attention accuracy as its distribution was highly skewed with a small number of extreme outliers.

2.4.4. Classification Analysis

We trained classifiers to categorise CHR-P individuals into GFO/PFO based on the last available follow-up data (6 months [n = 26] or 12 months [n = 92]). Classifiers included gaussian naive bayes (GNB), linear discriminant analysis (LDA), support vector machines (SVM), random forest classification (RFC) and logistic regression (LR). With the exception of class weights, which were set to 'balanced' due to the unequal numbers of PFO and GFO individuals, default SKL hyperparameters were used. Using only those CHR-P participants with follow-up information, variables not set to zero by the LASSO-LARS model were used for these models (Supplementary Figure 1). All variables were rescaled between zero and one to avoid class separability problems induced by differences in scaling. Due to class imbalance with PFO being more common than GFO, we used area under the curve (AUC) scores to determine whether classifiers performed significantly above chance.

Mixed-site classifiers were evaluated using 10-fold cross-validation, whereby the full dataset was used, and performance metrics are reported as averaged across k-folds. Specifically, the SKL function permutation_test_score (10000 permutations), which implements Test 1 from Ojala and Garriga (2010), was used to conduct permutation tests to evaluate AUC significance. We report performance metrics for all classifiers to evaluate their consistency, as very large discrepancies could suggest that the best performing classifiers were simply overfitting (Vieira et al., 2020).

To determine whether transfer could be established between the two test sites, cross-site classifiers were additionally evaluated with AUC scores obtained by training on the Glasgow data and testing on the Edinburgh data. This split was used as approximately two thirds of the data were collected at the Glasgow site. We report mean AUC, precision, recall, F1 scores (the harmonic mean of precision and recall) and mean balanced accuracy (BAC) for all classifiers. Recall for the two classes (PFO, GFO) corresponds to sensitivity and specificity,

respectively. Precision, recall and F1 scores were generated using the functions cross_val_predict and classification_report, whereby the former reports the prediction given for each data point when in the test set. All other scores reflect the mean across k-folds.
We also created two models which utilised only baseline functioning variables to obtain a stricter benchmark for classifier accuracy. In the first model, we split baseline GAF scores into good and poor functioning at baseline using the same threshold and used these data as predictors. In the second model, we trained the classifiers on social and role functioning as well as GAF scores to determine whether these additional variables could significantly improve classification accuracy compared to those based on baseline GAF scores alone.

3. Results

3.1. Demographic Information

In the CHR-P group, 106 (72.6%) and 70 (59.3%) individuals had poor functioning at baseline and follow-up respectively (Table 1). CHR-P individuals were significantly impaired relative to CHR-N and HC participants, displaying greater symptom severity and distress, increased ACES scores, more comorbid anxiety and mood disorders, lower functioning and poorer attention and processing speed. In addition, CHR-P participants were younger and reported fewer years of education. In the CHR-P group, baseline GAF scores were significantly affected by drug abuse/dependence ($p = .022$), anxiety disorders ($p = .031$) and mood disorders ($p < .001$). Age, gender, education and medication use exerted no such effects. Significant differences across study sites for the CHR-P group are displayed in Supplementary Table 1.

3.2. Baseline Regression Analysis

We fitted combined models where clinical, cognitive and functioning variables were entered as candidate predictors, and a cognitive model which only included cognitive variables.

In the combined model for the CHR-P group, cognitive (verbal memory, working memory RT, emotion recognition accuracy, motor speed), functioning (premorbid adjustment, social and role functioning) and clinical (SPI-A and CAARMS severity and distress, ACES total) variables were associated with baseline GAF scores (Figure 1). The combined model explained 41% of the variance in GAF scores in the CHR-P group, whereas the cognitive model explained 12%. The cognitive model contained verbal memory, working memory accuracy and RT, executive function, emotion recognition accuracy and attention RT (Supplementary Figure 2). Unexpectedly, motor speed and executive function were negatively related to GAF scores in the combined and cognitive models, respectively, while attention RT was positively related to GAF scores in the cognitive model.

Concurring with permutation feature importance scores (Supplementary Table 2), social functioning ($\beta = 2.97$) emerged as the strongest predictor in the combined model (Table 2; Supplementary Table 3) whereas verbal memory ($\beta = 1.88$) was a particularly strong predictor in the cognitive model (Table 2).

The combined model for the CHR-N group explained 17% of the variance in baseline GAF scores. This model included clinical (SPI-A distress) and functioning (social and role functioning) variables (Supplementary Figure 3) with role functioning ($\beta = 2.07$) emerging as the strongest predictor (Table 2, Supplementary Table 2). The cognitive model for the CHR-N group, however, failed to explain any variance in GAF scores.

3.3. Follow-up Classification Analysis

At follow-up, 59.3% of CHR-P individuals presented with PFO. Mixed-site classifiers were

trained to predict PFO versus GFO in CHR-P individuals based on variables associated with baseline GAF (Figure 1). Permutation tests on AUC scores indicated that all mixed-site classifiers performed significantly above chance (Table 3). The classifiers performed consistently, showing a mean BAC of 63% and a mean AUC of 0.72, while LR performed best (mean AUC = 0.74; mean BAC = 0.65). Mean sensitivity and specificity across classifiers was 68% and 56%, respectively, suggesting a bias towards predicting PFO. Performance among the cross-site models was consistently lower than for the mixed-site models (Table 3), with a mean AUC of 0.64 and a mean BAC of 0.63 across classifiers. Again, sensitivity was consistently higher than specificity.

Classifiers using baseline functioning variables only (GAF, social and role functioning) performed better than either mixed- or cross-site models on average (mean AUC = 0.76; mean BAC = 0.68). The baseline GAF model, which used good and poor baseline functioning categories as predictors, yielded an AUC of 0.67 and BAC of 0.70.

3.4. GAF Score Changes

For the CHR-P group, median absolute change in GAF during follow-up was 10.0, whereas median raw change in GAF (including negative change as is) was 0.5 (Supplementary Figure 4A-B). Eighty-three (70.3%) individuals did not change GAF category between baseline and follow-up, whereby 59 (50.0%) and 24 (20.3%) presented with poor and good functioning at both time points respectively. By contrast, 35 (29.7%) individuals did change GAF category between baseline and follow-up, whereby 24 (20.3%) changed from poor to good functioning and 11 (9.3%) changed from good to poor functioning. These results were statistically significant ($p < .001$) with the highest proportion of CHR-P participants presenting with poor functioning at both time points. Notably, raw GAF score changes between baseline and 6-month follow-up were not significantly different from changes between baseline and 12-

month follow-up in CHR-P individuals (n = 84) with both follow-up assessments (p = .590; Supplementary Figure 4C-D).

4. Discussion

We investigated the contribution of cognition towards impaired functioning as well as the potential utility of incorporating cognitive variables into predictive models of functional outcome. Although cognitive deficits explained 41% of the variance in baseline GAF scores when combined with clinical and functioning variables, cognitive variables alone explained only 12%. The combination of cognitive variables with functioning and clinical variables allowed classification of CHR-P individuals into GFO and PFO groups at follow-up with an average BAC of 63% in both mixed- and cross-site models. Furthermore, we were able to predict functional outcomes with acceptable accuracy using simple classifiers incorporating only baseline functioning variables.

4.1. *Predictors of Baseline Functioning*

In addition to clinical and functioning variables, cognitive deficits emerged as predictors of baseline functioning, together explaining 41% of the variance in baseline GAF scores in CHR-P participants. Impaired functioning prior to disorder onset is one of the strongest predictors of functional outcome in CHR-P individuals (Salokangas et al., 2014) and in patients with first-episode psychosis or established schizophrenia (Barajas et al., 2013). Indeed, functioning variables comprised the strongest predictors in the current study, illustrating the importance of interventions targeting functional impairments during early psychosis. Cognitive and clinical variables were weaker predictors and evidenced relatively similar importance scores. In line with previous studies, verbal memory (Meyer et al., 2014;

Niendam et al., 2006), working memory (Goghari et al., 2014), emotion recognition (Glenthøj et al., 2016), motor speed (Carrión et al., 2013), ACES total (Kraan et al., 2015), social and role functioning and premorbid adjustment (Salokangas et al., 2014) emerged as predictors of GAF in the combined CHR-P model. The emergence of CAARMS and SPI-A severity and distress scores as predictors, however, contrasts with previous findings (Carrión et al., 2016; Kim et al., 2019; Lin et al., 2011; Meyer et al., 2014; Rekhi et al., 2019).

In the CHR-P group, the relationship observed between impaired cognition and functioning is consistent with studies in established schizophrenia where cognitive deficits have been linked to decreased ability to live independently, poor social skills and inability to maintain employment (Lepage et al., 2014). Cognitive variables alone only explained 12% of the variance in baseline functioning in the CHR-P group, concurring with previous studies in schizophrenia (Fett et al., 2011) and CHR-P cohorts (Carrión et al., 2011). Notably, one of the strongest cognitive predictors was verbal memory, consistent with previous CHR-P studies predicting social functioning (Meyer et al., 2014; Niendam et al., 2006) and schizophrenia studies predicting a variety of functional outcomes (Green, 1996). Although certain cognitive variables (i.e. motor speed, executive function and attention RT) displayed unexpected relationships with baseline functioning in both combined and cognitive CHR-P models, this may partially reflect a speed-accuracy trade off. Moreover, in our CHR-N sample, cognitive variables were unrelated to GAF, suggesting that this relationship may be specific to the CHR-P state. However, this finding may be explained by the absence of significant cognitive deficits in the CHR-N sample and the smaller sample size.

4.2. Predictors of Functional Outcome

Mixed-site models combining cognitive variables with clinical and functioning variables were able to predict functional outcome in the CHR-P group. All mixed-site models

performed significantly above chance, with a mean AUC of 0.72 and a mean BAC of 63%. Performance was relatively consistent across all algorithms making it unlikely that our best performing classifier (LR; mean AUC = 0.74) was overfitting. These data are in line with previous research utilising clinical, functional and neuroimaging data where functional outcomes have been predicted with AUC scores between 0.70-0.86 and accuracies between 62.5%-82.7% (Kambeitz-Ilankovic et al., 2016; Koutsouleris et al., 2018; Mechelli et al., 2017). Notably, performance in the current study decreased for the cross-site models (mean AUC = 0.64; mean BAC = 63%), which is a common problem noted for machine learning classifiers in the field (Vieira et al., 2020).

We additionally fitted classifiers on baseline functioning variables. Using baseline data to predict later measures of the same variable often predicts outcomes better than chance and baseline models can provide a more stringent method for evaluating classifier accuracy (DeMasi et al., 2017). Indeed, previous studies identified global and social functioning scores as the most useful variables for predicting social functioning at 1-year follow-up in CHR-P participants (Koutsouleris et al., 2018). In the current study, classifiers fitted only on baseline functioning variables performed better, on average, than both mixed- and cross-site models with a mean AUC and BAC of 0.76 and 68%, respectively. This is possibly explained by the fact that GAF scores appear to be relatively stable across time. Overall, nearly two thirds of our sample showed PFO in agreement with previous studies (Carrión et al., 2013; Koutsouleris et al., 2018) and the majority of individuals (70.3%) remained within the same outcome category.

4.3. Limitations

Both the regression and classification analyses could be optimised by increasing the number of participants relative to candidate predictors. Additionally, we only had two test sites,

meaning that cross-site classifiers were only trained on a single site, thus limiting their ability to learn patterns across multiple sites. Given that machine learning models have the potential to outperform human judgement, it is highly probable that models predicting functional outcomes in early psychosis can improve in larger datasets (Fusar-Poli et al., 2019). As accuracy tends to exhibit a strong relationship with sample size for machine learning methods in particular (Floares et al., 2017), standardising data acquisition protocols across research centres and thereby facilitating the collection of much larger collaborative datasets is likely to produce significant performance gains in terms of both accuracy and cross-site transfer. Furthermore, due to the small size of CHR-N participants, strong conclusions regarding the contribution of cognitive deficits towards impaired functioning in this group cannot be drawn and, given that only 55% completed follow-up assessments, GAF outcome/change could not be examined in this group.

The current study also highlights the limitations of current functioning measurements in CHR-P populations. The GAF scale, for example, confounds functioning with symptom severity and shows only limited fluctuations over time. However, the GAF scale was chosen over social and role functioning scales in this study as scores obtained from the latter displayed low variability. Accordingly, more sensitive measures are required that trace changes in functioning across several dimensions. Finally, negative symptoms, which have been shown to mediate the relationship between neurocognition and functioning (Glenthøj et al., 2016; Meyer et al., 2014), as well as treatment use over follow-up were not assessed in the current CHR-P sample.

4.4. Conclusions

Utilising a machine learning approach, we have shown that cognitive variables alongside clinical and functioning variables predict short-term functional outcome with above-chance

performance. With the increasing popularity of complex machine learning models in psychiatry, it is important to consider appropriate benchmark measures to determine whether the potential gains are sufficient to justify their use over simpler alternatives. Our findings suggest, for example, that baseline GAF scores allow a more robust prediction of functional outcomes in CHR-P individuals than complex machine learning approaches. Given the large proportion of CHR-P individuals presenting with PFO, interventions incorporating social skills training, vocational rehabilitation and cognitive remediation are clearly warranted at this stage.

Acknowledgements

This study was supported by project MR/L011689/1 from the Medical Research Council (MRC) and the University of Glasgow's Lord Kelvin/Adam Smith (LKAS) PhD Scholarship. We acknowledge the support of the Scottish Mental Health Research Network (http://www.smhrn.org.uk), now called the NHS Research Scotland Mental Health Network (NRS MHN: http://www.nhsresearchscotland.org.uk/research-areas/mental-health), for providing assistance with participant recruitment, interviews and cognitive assessments. We would like to thank both the participants and patients who took part in the study and the research assistants of the YouR-study for supporting the recruitment and assessment of CHR-P participants. We also would like to thank Professor Roderick Murray-Smith, School of Computing Science, University of Glasgow, for his feedback regarding machine learning methods.

Figure Legends:

Figure 1: Correlation matrix showing the relationship between nonzero predictors and baseline GAF scores for the combined LASSO-LARS regression model for the CHR-P group (N = 146). The latest GAF score is added to this figure for visualisation purposes only and has not been entered in the regression model.

References

Allen, P., Chaddock, C.A., Egerton, A., Howes, O.D., Barker, G., Bonoldi, I., Fusar-Poli, P., Murray, R., McGuire, P., 2015. Functional Outcome in People at High Risk for Psychosis Predicted by Thalamic Glutamate Levels and Prefronto-Striatal Activation. Schizophr. Bull. 41, 429–439. https://doi.org/10.1093/schbul/sbu115

Allott, K., Liu, P., Proffitt, T.M., Killackey, E., 2011. Cognition at illness onset as a predictor of later functional outcome in early psychosis: Systematic review and methodological critique. Schizophr. Res. 125, 221–235. https://doi.org/10.1016/j.schres.2010.11.001

Barajas, A., Usall, J., Baños, I., Dolz, M., Villalta-Gil, V., Vilaplana, M., Autonell, J., Sánchez, B., Cervilla, J.A., Foix, A., Obiols, J.E., Haro, J.M., Ochoa, S., 2013. Three-factor model of premorbid adjustment in a sample with chronic schizophrenia and first-episode psychosis. Schizophr. Res. 151, 252–258. https://doi.org/10.1016/j.schres.2013.10.027

Cannon-Spoor, H.E., Potkin, S.G., Wyatt, R.J., 1982. Measurement of Premorbid Adjustment in Chronic Schizophrenia. Schizophr. Bull. 8, 470–484. https://doi.org/10.1093/schbul/8.3.470

Carrión, R.E., Demmin, D., Auther, A.M., McLaughlin, D., Olsen, R., Lencz, T., Correll, C.U., Cornblatt, B.A., 2016. Duration of attenuated positive and negative symptoms in individuals at clinical high risk: Associations with risk of conversion to psychosis and

functional outcome. J. Psychiatr. Res. 81, 95–101. https://doi.org/10.1016/j.jpsychires.2016.06.021

Carrión, R.E., Goldberg, T.E., McLaughlin, D., Auther, A.M., Correll, C.U., Cornblatt, B.A., 2011. Impact of neurocognition on social and role functioning in individuals at clinical high risk for psychosis. Am. J. Psychiatry 168, 806–813. https://doi.org/10.1176/appi.ajp.2011.10081209

Carrión, R.E., McLaughlin, D., Goldberg, T.E., Auther, A.M., Olsen, R.H., Olvet, D.M., Correll, C.U., Cornblatt, B.A., 2013. Prediction of functional outcome in individuals at clinical high risk for psychosis. JAMA Psychiatry 70, 1133–1142. https://doi.org/10.1001/jamapsychiatry.2013.1909

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. Schizophr. Bull. 33, 688–702. https://doi.org/10.1093/schbul/sbm029

DeMasi, O., Kording, K., Recht, B., 2017. Meaningless comparisons lead to false optimism in medical machine learning. PLoS One 12, 1–17. https://doi.org/10.1371/journal.pone.0184604

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R., Ishwaran, H., Knight, K., Loubes, J.M., Massart, P., Madigan, D., Ridgeway, G., Rosset, S., Zhu, J.I., Stine, R.A., Turlach, B.A., Weisberg, S., Johnstone, I., Tibshirani, R., 2004. Least angle regression. Ann. Stat. 32, 407–499. https://doi.org/10.1214/009053604000000067

Eslami, A., Jahshan, C., Cadenhead, K.S., 2011. Disorganized Symptoms and Executive Functioning Predict Impaired Social Functioning in Subjects at Risk for Psychosis. J. Neuropsychiatr. 23, 457–460. https://doi.org/10.1176/appi.neuropsych.23.4.457

Felitti, V.J., Anda, R.F., Nordenberg, D., Williamson, D.F., Spitz, A.M., Edwards, V., Koss,

M.P., Marks, J.S., 1998. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. Am J Prev Med 14, 245–258. https://doi.org/doi: 10.1016/s0749-3797(98)00017-8

Fett, A.K.J., Viechtbauer, W., Dominguez, M. de G., Penn, D.L., van Os, J., Krabbendam, L., 2011. The relationship between neurocognition and social cognition with functional outcomes in schizophrenia: A meta-analysis. Neurosci. Biobehav. Rev. 35, 573–588. https://doi.org/10.1016/j.neubiorev.2010.07.001

Floares, A.G., Ferisgan, M., Onita, D., Ciuparu, A., Calin, G.A., Manolache, F.B., 2017. The Smallest Sample Size for the Desired Diagnosis Accuracy. Int. J. Oncol. Cancer Ther. 2, 13–19.

Fonti, V., 2017. Feature Selection using LASSO. VU Amsterdam 1–26. https://doi.org/10.1109/ACCESS.2017.2696365

Fusar-Poli, P., Borgwardt, S., Bechdolf, A., Addington, J., Riecher-Rössler, A., Schultze-Lutter, F., Keshavan, M., Wood, S., Ruhrmann, S., Seidman, L.J., Valmaggia, L., Cannon, T., Velthorst, E., De Haan, L., Cornblatt, B., Bonoldi, I., Birchwood, M., McGlashan, T., Carpenter, W., McGorry, P., Klosterkötter, J., McGuire, P., Yung, A., 2013. The psychosis high-risk state: A comprehensive state-of-the-art review. Arch. Gen. Psychiatry 70, 107–120. https://doi.org/10.1001/jamapsychiatry.2013.269

Fusar-Poli, P., Cappucciati, M., Borgwardt, S., Woods, S.W., Addington, J., Nelson, B., Nieman, D.H., Stahl, D.R., Rutigliano, G., Riecher-Rössler, A., Simon, A.E., Mizuno, M., Lee, T.Y., Kwon, J.S., Lam, M.M.L., Perez, J., Keri, S., Amminger, P., Metzler, S., Kawohl, W., Rössler, W., Lee, J., Labad, J., Ziermans, T., An, S.K., Liu, C.C., Woodberry, K.A., Braham, A., Corcoran, C., McGorry, P., Yung, A.R., McGuire, P.K., 2016. Heterogeneity of psychosis risk within individuals at clinical high risk: A meta-

analytical stratification. JAMA Psychiatry 73, 113–120. https://doi.org/10.1001/jamapsychiatry.2015.2324

Fusar-Poli, P., Stringer, D., M. S. Durieux, A., Rutigliano, G., Bonoldi, I., De Micheli, A., Stahl, D., 2019. Clinical-learning versus machine-learning for transdiagnostic prediction of psychosis onset in individuals at-risk. Transl. Psychiatry 9. https://doi.org/10.1038/s41398-019-0600-9

Glenthøj, L.B., Fagerlund, B., Hjorthøj, C., Jepsen, J.R.M., Bak, N., Kristensen, T.D., Wenneberg, C., Krakauer, K., Roberts, D.L., Nordentoft, M., 2016. Social cognition in patients at ultra-high risk for psychosis: What is the relation to social skills and functioning? Schizophr. Res. Cogn. 5, 21–27. https://doi.org/10.1016/j.scog.2016.06.004

Goghari, V.M., Brett, C., Tabraham, P., Johns, L., Valmaggia, L., Broome, M., Woolley, J., Bramon, E., Howes, O., Byrne, M., McGuire, P., 2014. Spatial working memory ability in individuals at ultra high risk for psychosis. J. Psychiatr. Res. 50, 100–105. https://doi.org/10.1016/j.jpsychires.2013.12.010

Green, M.F., 1996. What are the functional consequences of neurocognitive deficits in schizophrenia? Am J Psychiatry 153, 321–330.

Kambeitz-Ilankovic, L., Meisenzahl, E.M., Cabral, C., von Saldern, S., Kambeitz, J., Falkai, P., Möller, H.J., Reiser, M., Koutsouleris, N., 2016. Prediction of outcome in the psychosis prodrome using neuroanatomical pattern classification. Schizophr. Res. 173, 159–165. https://doi.org/10.1016/j.schres.2015.03.005

Keefe, R.S.E., Goldberg, T.E., Harvey, P.D., Gold, J.M., Poe, M.P., Coughenour, L., 2004. The Brief Assessment of Cognition in Schizophrenia: Reliability, sensitivity, and comparison with a standard neurocognitive battery. Schizophr. Res. 68, 283–297. https://doi.org/10.1016/j.schres.2003.09.011

Keefe, R.S.E., 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). Schizophr. Res. 102, 108–115. https://doi.org/10.1016/j.schres.2008.03.024

Kim, H.K., Park, H.Y., Seo, E., Bang, M., Song, Y.Y., Lee, S.Y., Kim, K.R., Park, J.Y., Kang, J.I., Lee, E., An, S.K., 2019. Factors associated with psychosocial functioning and outcome of individuals with recent-onset schizophrenia and at ultra-high risk for psychosis. Front. Psychiatry 10, 1–11. https://doi.org/10.3389/fpsyt.2019.00459

Koutsouleris, N., Borgwardt, S., Meisenzahl, E.M., Bottlender, R., Möller, H.J., Riecher-Rössler, A., 2012. Disease prediction in the at-risk mental state for psychosis using neuroanatomical biomarkers: Results from the fepsy study. Schizophr. Bull. 38, 1234–1246. https://doi.org/10.1093/schbul/sbr145

Koutsouleris, N., Kambeitz-Ilankovic, L., Ruhrmann, S., Rosen, M., Ruef, A., Dwyer, D.B., Paolini, M., Chisholm, K., Kambeitz, J., Haidl, T., Schmidt, A., Gillam, J., Schultze-Lutter, F., Falkai, P., Reiser, M., Riecher-Rössler, A., Upthegrove, R., Hietala, J., Salokangas, R.K.R., Pantelis, C., Meisenzahl, E., Wood, S.J., Beque, D., Brambilla, P., Borgwardt, S., 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, 1156–1172. https://doi.org/10.1001/jamapsychiatry.2018.2165

Koutsouleris, N., Meisenzahl, E.M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., Möller, H.J., Gaser, C., 2009. Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. Arch. Gen. Psychiatry 66, 700–712. https://doi.org/10.1001/archgenpsychiatry.2009.62

Kraan, T., van Dam, D.S., Velthorst, E., de Ruigh, E.L., Nieman, D.H., Durston, S.,

Schothorst, P., van der Gaag, M., de Haan, L., 2015. Childhood trauma and clinical outcome in patients at ultra-high risk of transition to psychosis. Schizophr. Res. 169, 193–198. https://doi.org/10.1016/j.schres.2015.10.030

Lepage, M., Bodnar, M., Bowie, C.R., 2014. Neurocognition: Clinical and functional outcomes in schizophrenia. Can. J. Psychiatry 59, 5–12. https://doi.org/10.1177/070674371405900103

Lin, A., Wood, S.J., Nelson, B., Brewer, W.J., Spiliotacopoulos, D., Bruxner, A., Broussard, C., Pantelis, C., Yung, A.R., 2011. Neurocognitive predictors of functional outcome two to 13years after identification as ultra-high risk for psychosis. Schizophr. Res. 132, 1–7. https://doi.org/10.1016/j.schres.2011.06.014

McDonald, M., Christoforidou, E., Van Rijsbergen, N., Gajwani, R., Gross, J., Gumley, A.I., Lawrie, S.M., Schwannauer, M., Schultze-Lutter, F., Uhlhaas, P.J., 2019. Using online screening in the general population to detect participants at clinical high-risk for psychosis. Schizophr. Bull. 45, 600–609. https://doi.org/10.1093/schbul/sby069

Mechelli, A., Lin, A., Wood, S., McGorry, P., Amminger, P., Tognin, S., McGuire, P., Young, J., Nelson, B., Yung, A., 2017. Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. Schizophr. Res. 184, 32–38. https://doi.org/10.1016/j.schres.2016.11.047

Meyer, E.C., Carrión, R.E., Cornblatt, B.A., Addington, J., Cadenhead, K.S., Cannon, T.D., McGlashan, T.H., Perkins, D.O., Tsuang, M.T., Walker, E.F., Woods, S.W., Heinssen, R., Seidman, L.J., 2014. The relationship of neurocognition and negative symptoms to social and role functioning over time in individuals at clinical high risk in the first phase of the north american prodrome longitudinal study. Schizophr. Bull. 40, 1452–1461. https://doi.org/10.1093/schbul/sbt235

Millan, M.J., Andrieux, A., Bartzokis, G., Cadenhead, K., Dazzan, P., Fusar-Poli, P.,

Gallinat, J., Giedd, J., Grayson, D.R., Heinrichs, M., Kahn, R., Krebs, M.O., Leboyer, M., Lewis, D., Marin, O., Marin, P., Meyer-Lindenberg, A., McGorry, P., McGuire, P., Owen, M.J., Patterson, P., Sawa, A., Spedding, M., Uhlhaas, P., Vaccarino, F., Wahlestedt, C., Weinberger, D., 2016. Altering the course of schizophrenia: Progress and perspectives. Nat. Rev. Drug Discov. 15, 485–515. https://doi.org/10.1038/nrd.2016.28

Millman, Z.B., Gold, J.M., Mittal, V.A., Schiffman, J., 2019. The Critical Need for Help-Seeking Controls in Clinical High-Risk Research. Clin. Psychol. Sci. 7, 1171–1189. https://doi.org/10.1177/2167702619855660

Modinos, G., Kempton, M.J., Tognin, S., Calem, M., Porffy, L., Antoniades, M., Mason, A., Azis, M., Allen, P., Nelson, B., McGorry, P., Pantelis, C., Riecher-Rössler, A., Borgwardt, S., Bressan, R., Barrantes-Vidal, N., Krebs, M.O., Nordentoft, M., Glenthøj, B., Ruhrmann, S., Sachs, G., Rutten, B., Van Os, J., De Haan, L., Velthorst, E., Van Der Gaag, M., Valmaggia, L.R., McGuire, P., 2019. Association of Adverse Outcomes with Emotion Processing and Its Neural Substrate in Individuals at Clinical High Risk for Psychosis. JAMA Psychiatry 77, 190–200. https://doi.org/10.1001/jamapsychiatry.2019.3501

Moore, T.M., Reise, S.P., Gur, R.E., Hakonarson, H., Gur, R.C., 2015. Psychometric properties of the penn computerized neurocognitive battery. Neuropsychology 29, 235–246. https://doi.org/10.1037/neu0000093

Niendam, T.A., Bearden, C.E., Johnson, J.K., McKinley, M., Loewy, R., O'Brien, M., Nuechterlein, K.H., Green, M.F., Cannon, T.D., 2006. Neurocognitive performance and functional disability in the psychosis prodrome. Schizophr. Res. 84, 100–111. https://doi.org/10.1016/j.schres.2006.02.005

Ojala, M., Garriga, G.C., 2010. Permutation tests for studying classifier performance. J.

Mach. Learn. Res. 11, 1833–1863.

Owen, M.J., Sawa, A., Mortensen, P.B., 2016. Schizophrenia. Lancet 388, 86–97. https://doi.org/10.1016/S0140-6736(15)01121-6

Ramyead, A., Studerus, E., Kometer, M., Uttinger, M., Gschwandtner, U., Fuhr, P., Riecher-Rössler, A., 2016. Prediction of psychosis using neural oscillations and machine learning in neuroleptic-naïve at-risk patients. World J. Biol. Psychiatry 17, 285–295. https://doi.org/10.3109/15622975.2015.1083614

Rekhi, G., Rapisarda, A., Lee, J., 2019. Impact of distress related to attenuated psychotic symptoms in individuals at ultra high risk of psychosis: Findings from the Longitudinal Youth at Risk Study. Early Interv. Psychiatry 13, 73–78. https://doi.org/10.1111/eip.12451

Salokangas, R.K.R., Heinimaa, M., From, T., Löyttyniemi, E., Ilonen, T., Luutonen, S., Hietala, J., Svirskis, T., von Reventlow, H.G., Juckel, G., Linszen, D., Dingemans, P., Birchwood, M., Patterson, P., Schultze-Lutter, F., Ruhrmann, S., Klosterkötter, J., McGorry, P.D., McGlashan, T.H., Knapp, M., van der Fliert, R., Klaassen, R., Picker, H., Neumann, M., Brockhaus-Dumke, A., Pukrop, R., Huttunen, J., Ristkari, T., Skeate, A., Gudlowski, Y., Ozgürdal, S., Witthaus, H., Lewis, S., Morrisson, A., 2014. Short-term functional outcome and premorbid adjustment in clinical high-risk patients. Results of the EPOS project. Eur. Psychiatry 29, 371–380. https://doi.org/10.1016/j.eurpsy.2013.10.003

Schmidt, S.J., Mueller, D.R., Roder, V., 2011. Social cognition as a mediator variable between neurocognition and functional outcome in schizophrenia: Empirical review and new results by structural equation modeling. Schizophr. Bull. 37. https://doi.org/10.1093/schbul/sbr079

Schultze-Lutter, F., 2009. Subjective symptoms of schizophrenia in research and the clinic:

The basic symptom concept. Schizophr. Bull. 35, 5–8. https://doi.org/10.1093/schbul/sbn139

Schultze-Lutter, F., Addington, J., Ruhrmann, S., Klosterkötter, J., 2007. Schizophrenia Proneness Instrument-Adult Version (SPI-A). Giovanni Fioriti Editore, Rome, Italy.

Schultze-Lutter, F., Rahman, J., Ruhrmann, S., Michel, C., Schimmelmann, B.G., Maier, W., Klosterkötter, J., 2015. Duration of unspecific prodromal and clinical high risk states, and early help-seeking in first-admission psychosis patients. Soc. Psychiatry Psychiatr. Epidemiol. 50, 1831–1841. https://doi.org/10.1007/s00127-015-1093-3

Schultze-Lutter, F., Ruhrmann, S., Fusar-Poli, P., Bechdolf, A., Schimmelmann, B.G., Klosterkotter, J., 2012. Basic Symptoms and the Prediction of First-Episode Psychosis. Curr. Pharm. Des. 18, 351–357.

Sheehan, D. V, Lecrubier, Y., Sheehan, K., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., Dunbar, G.C., 1998. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 59, 22–33.

Uhlhaas, P.J., Gajwani, R., Gross, J., Gumley, A.I., Lawrie, S.M., Schwannauer, M., 2017. The Youth Mental Health Risk and Resilience Study (YouR-Study). BMC Psychiatry 17, 1–8. https://doi.org/10.1186/s12888-017-1206-5

Vieira, S., Gong, Q.Y., Pinaya, W.H.L., Scarpazza, C., Tognin, S., Crespo-Facorro, B., Tordesillas-Gutierrez, D., Ortiz-García, V., Setien-Suero, E., Scheepers, F.E., van Haren, N.E.M., Marques, T.R., Murray, R.M., David, A., Dazzan, P., McGuire, P., Mechelli, A., 2020. Using machine learning and structural neuroimaging to detect first episode psychosis: Reconsidering the evidence. Schizophr. Bull. 46, 17–26. https://doi.org/10.1093/schbul/sby189

Yung, A.R., Yuen, H.P., McGorry, P.D., Phillips, L.J., Kelly, D., Dell'Olio, M., Francey,

S.M., Cosgrave, E.M., Killackey, E., Stanford, C., Godfrey, K., Buckby, J., 2005. Mapping the onset of psychosis: The Comprehensive Assessment of At-Risk Mental States. Aust. N. Z. J. Psychiatry 39, 964–971. https://doi.org/10.1111/j.1440-1614.2005.01714.x

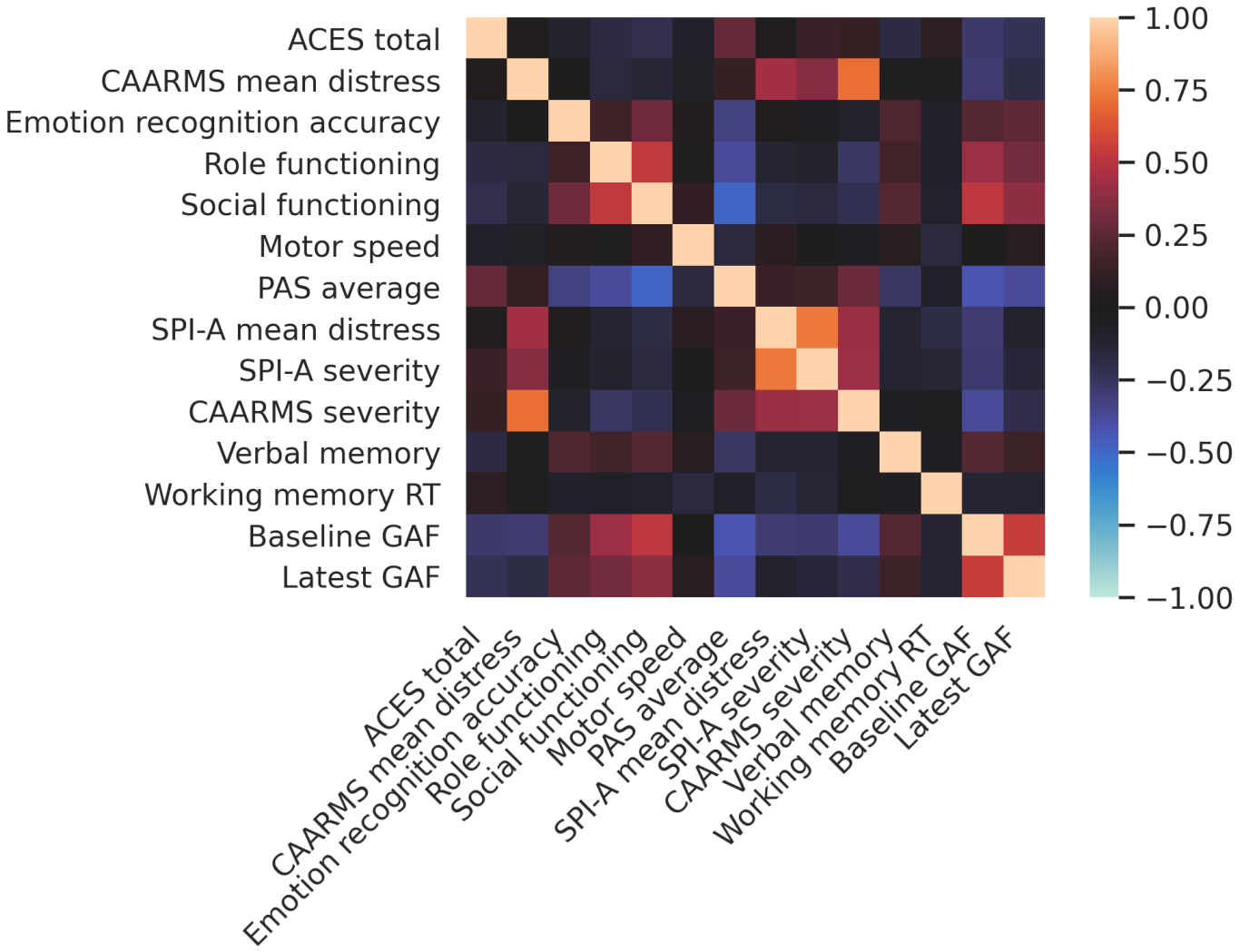
ACES total
CAARMS mean distress
Emotion recognition accuracy
Role functioning
Social functioning
Motor speed
PAS average
SPI-A mean distress
SPI-A severity
CAARMS severity
Verbal memory
Working memory RT
Baseline GAF
Latest GAF
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

Table 1. Demographic, clinical, functioning and cognitive characteristics of the entire sample (N = 248)

Variable	**CHR-P (N = 146)**	**CHR-N (N = 47)**	**HC (N = 55)**	***p*-value**	**Post hoc tests**
Age (years), mean (SD)	21.47 (4.22)	22.94 (4.80)	22.31 (3.39)	.025	CHR-P v HC, CHR-N
Gender, female n (%)	104 (71.2)	30 (63.8)	37 (67.3)	.606	-
Education (years), mean (SD)	15.12 (3.09)	16.45 (3.44)	16.38 (2.84)	.006	CHR-P v HC, CHR-N
CAARMS severity, median (range)	28 (0-74)	6 (0-24)	0 (0-12)	< .001	CHR-P v HC v CHR-N
CAARMS mean distress, median (range)	30 (0-86)	3 (0-55)	0 (0-25)	< .001	CHR-P v HC v CHR-N
SPI-A severity, median (range)	7 (0-74)	0 (0-7)	0 (0-2)	< .001	CHR-P v HC, CHR-N
SPI-A mean distress , median (range)	3 (0-28)	0 (0-6)	0 (0-1)	< .001	CHR-P v HC v CHR-N
CHR-P criteria subgroup, n (%)					
CAARMS	45 (30.8)	-	-	-	-
SPI-A	37 (25.3)	-	-	-	-
CAARMS/SPI-A	64 (43.8)	-	-	-	-
ACES total, median (range)	2 (0-8)	1 (0-4)	0 (0-4)	< .001	CHR-P v HC v CHR-N
Comorbidity, n (%)					
Anxiety disorder	104 (71.2)	22 (46.8)	0 (0)	< .001	CHR-P v HC v CHR-N
Mood disorder	97 (66.4)	14 (29.8)	0 (0)	< .001	CHR-P v HC v CHR-N
Alcohol abuse/dependence	46 (31.5)	11 (23.4)	2 (3.6)	< .001	HC v CHR-P, CHR-N
Drug abuse/dependence	24 (16.4)	3 (6.4)	0 (0)	.002	CHR-P v HC
Eating disorder	11 (7.5)	1 (2.1)	0 (0)	.054	-
Medication, n (%)					
Antipsychotic	4 (2.7)	0 (0)	0 (0)	.242	-
Mood stabiliser	4 (2.7)	0 (0)	0 (0)	.242	-
Antidepressant	53 (36.3)	13 (27.7)	0 (0)	< .001	HC v CHR-P, CHR-N
Anti-anxiety	10 (6.8)	1 (2.1)	0 (0)	.076	-
GAF, median (range)	58 (21-95)	70 (43-94)	88 (67-97)	< .001	CHR-P v HC v CHR-N
Poor baseline functioning, n (%)	106 (72.6)	-	-		-
PFO, n (%)	70 (59.3)	-	-		-
Social functioning, median (range)	8 (3-10)	8 (6-9)	9 (8-10)	< .001	CHR-P v HC v CHR-N
Role functioning, median (range)	8 (3-9)	8 (5-9)	9 (5-9)	< .001	CHR-P v HC v CHR-N
PAS average, median (range)	1.22 (0-3.43)	0.86 (0-3.86)	0.43 (0-1.64)	< .001	CHR-P v HC v CHR-N
BACS, mean (SD)					
Verbal memory	-0.22 (1.20)	0.09 (1.05)	0 (1.01)	.295	-
Motor speed	-0.72 (1.21)	-0.39 (1.01)	0 (1.01)	< .001	CHR-P v HC
Attention & processing speed	-0.48 (1.14)	0.08 (1.19)	0 (1.01)	.001	CHR-P v HC, CHR-N
Verbal fluency	-0.09 (1.24)	-0.23 (1.05)	0 (1.01)	.760	-
Executive function	0 (1.34)	0.05 (1.25)	0 (1.01)	.855	-
Working memory	-0.08 (1.41)	0.24 (1.13)	0 (1.01)	.443	-
Composite score	-0.59 (1.71)	-0.07 (1.36)	0 (1.01)	.022	CHR-P v HC

CNB, mean (SD)					
Emotion recognition accuracy	-0.17 (1.13)	-0.10 (0.91)	0 (1.01)	.565	-
Emotion recognition RT	0.59 (1.58)	0.18 (1.33)	0 (1.01)	.037	CHR-P v HC
Attention accuracy	-0.71 (2.60)	0.10 (1.13)	0 (1.01)	.039	CHR-P v HC
Attention RT	-0.11 (0.86)	-0.26 (0.96)	0 (1.01)	.326	-
Working memory accuracy	-0.41 (1.68)	-0.17 (1.23)	0 (1.01)	.286	-
Working memory RT	-0.05 (0.82)	-0.10 (0.98)	0 (1.01)	.691	-

Note. CHR-P, clinical high-risk for psychosis; CHR-N, clinical high-risk-negative; HC, healthy control; CAARMS, Comprehensive Assessment of At-Risk Mental States; SPI-A, Schizophrenia Proneness Instrument, Adult version; ACES, Adverse Childhood Experience Scale; GAF, Global Assessment of Functioning; PFO, poor functional outcome; PAS, Premorbid Adjustment Scale; BACS, Brief Assessment of Cognition in Schizophrenia; CNB, Penn Computerized Neurocognitive Battery; RT, response time

Table 2: Nonzero LASSO-LARS regression coefficients for CHR-P (N = 146) and CHR-N (N = 47) baseline models

Variable	**β coefficient**		
	CHR-P combined	**CHR-P cognitive**	**CHR-N combined**
Social functioning	2.97		1.12
PAS average	-2.15		
Role functioning	1.24		2.07
Working memory RT	-0.96	-1.88	
SPI-A mean distress	-0.85		-0.63
ACES total	-0.51		
Motor speed	-0.24		
Verbal memory	0.24	1.88	
Emotion recognition accuracy	0.11	1.75	
Total CAARMS severity	-0.10		
SPI-A severity	-0.05		
CAARMS mean distress	-0.02		
Attention RT		1.27	
Executive function		-0.60	
Working memory RT		0.05	

Note. CHR-P, clinical high-risk for psychosis; CHR-N, clinical high-risk-negative; CAARMS, Comprehensive Assessment of At-Risk Mental States; SPI-A, Schizophrenia Proneness Instrument, Adult version; ACES, Adverse Childhood Experience Scale; GAF, Global Assessment of Functioning; PAS, Premorbid Adjustment Scale; RT, response time

Table 3: Mixed-site and cross-site classifiers in the CHR-P sample (N = 146)

Model	Mean AUC	Precision (PFO/GFO)	Recall (PFO/GFO)	F1 (PFO/GFO)	Mean BAC	*p*-value[a]
Mixed-site classifiers						
GNB	0.70	0.68/0.48	0.54/0.62	0.60/0.55	0.59	.002
LDA	0.73	0.71/0.59	0.73/0.56	0.72/0.57	0.65	.005
SVM	0.72	0.73/0.55	0.64/0.65	0.68/0.60	0.65	.014
LR	0.74	0.72/0.56	0.66/0.62	0.69/0.59	0.65	.003
RFC	0.72	0.64/0.55	0.81/0.33	0.72/0.42	0.61	.013
Average	0.72	0.70/0.55	0.68/0.56	0.68/0.55	0.63	-
Cross-site classifiers						
GNB	0.61	0.71/0.50	0.63/0.58	0.67 /0.54	0.61	-
LDA	0.64	0.68/0.50	0.68/0.50	0.68/0.50	0.59	-
SVM	0.62	0.72/0.54	0.68/0.58	0.70/0.56	0.63	-
LR	0.65	0.72/0.54	0.68/0.58	0.70/0.56	0.63	-
RFC	0.66	0.73/0.67	0.84/0.50	0.78/0.57	0.67	-
Average	0.64	0.71/0.55	0.70/0.55	0.71/0.55	0.63	-

Note. GNB, gaussian naive bayes; LDA, linear discriminant analysis; SVM, support vector machines; RFC, random forest classification; LR, logistic regression; AUC, area under the curve; PFO, poor functional outcome; GFO, good functional outcome; BAC, balanced accuracy.

[a] Permutation tests on AUC, corrected for multiple comparisons (Bonferroni).

Table 4: Classifiers using baseline functioning variables in the CHR-P sample (N = 146)

Model	**Mean AUC**	**Precision (PFO/GFO)**	**Recall (PFO/GFO)**	**F1 (PFO/GFO)**	**Mean BAC**	***p*-value[a]**
GNB	0.80	0.76/0.59	0.67/0.69	0.71 /0.63	0.68	.001
LDA	0.78	0.73/0.64	0.77/0.58	0.75/0.61	0.68	.001
SVM	0.79	0.78/0.59	0.66/0.73	0.71/0.65	0.70	.001
LR	0.79	0.77/0.60	0.67/0.71	0.72/0.65	0.69	.001
RFC	0.70	0.69/0.57	0.71/0.54	0.70/0.55	0.65	.019
Simple GAF[b]	0.67	0.71/0.69	0.84/0.50	0.77/0.58	0.70	-
Average	0.76	0.74/0.61	0.73/0.63	0.73/0.61	0.68	-

Note. GNB, gaussian naive bayes; LDA, linear discriminant analysis; SVM, support vector machines; RFC, random forest classification; LR, logistic regression; AUC, area under the curve; PFO, poor functional outcome; GFO, good functional outcome; GAF, Global Assessment of Functioning; BAC, balanced accuracy.

[a] Permutation tests on AUC, corrected for multiple comparisons (Bonferroni).

[b] Associated values reflect single values rather than means due to the nature of the model.