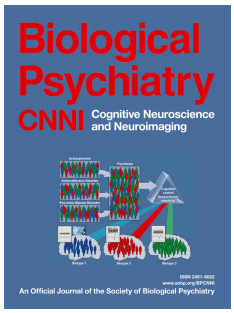

Combining clinical with cognitive or MRI data for predicting transition to psychosis in ultra-high risk patients: Data from the PACE 400 cohort

Simon Hartmann, Micah Cearns, Christos Pantelis, Dominic Dwyer, Blake Cavve, Enda Byrne, Isabelle Scott, Hok Pan Yuen, Caroline Gao, Kelly Allott, Ashleigh Lin, Stephen J. Wood, Johanna T.W. Wigman, Paul Amminger, Patrick D. McGorry, Alison R. Yung, Barnaby Nelson, Scott R. Clark

PII: S2451-9022(23)00320-8

DOI: https://doi.org/10.1016/j.bpsc.2023.11.009

Reference: BPSC 1165

To appear in: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*

Received Date: 27 July 2023

Revised Date: 19 October 2023

Accepted Date: 26 November 2023

Please cite this article as: Hartmann S., Cearns M., Pantelis C., Dwyer D., Cavve B., Byrne E., Scott I., Yuen H.P., Gao C., Allott K., Lin A., Wood S.J, Wigman J.T.W, Amminger P., McGorry P.D, Yung A.R, Nelson B. & Clark S.R, Combining clinical with cognitive or MRI data for predicting transition to psychosis in ultra-high risk patients: Data from the PACE 400 cohort, *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* (2024), doi: https://doi.org/10.1016/j.bpsc.2023.11.009.

Combining clinical with cognitive or MRI data for predicting transition to psychosis in ultra-high risk patients: Data from the PACE 400 cohort

Short title: Multimodal psychosis prediction modelling in PACE 400

Simon Hartmann[1,2,3], Micah Cearns[1], Christos Pantelis[4,5], Dominic Dwyer[2,3], Blake Cavve[6], Enda Byrne[7], Isabelle Scott[2,3], Hok Pan Yuen[2,3], Caroline Gao[2,3], Kelly Allott[2,3], Ashleigh Lin[6], Stephen J Wood[2,3,8], Johanna T W Wigman[9], Paul Amminger[2,3], Patrick D McGorry[2,3], Alison R Yung[10], Barnaby Nelson[2,3], Scott R Clark[1]

1. The University of Adelaide, Discipline of Psychiatry, Adelaide Medical School, Adelaide, Australia
2. Orygen, Melbourne, Australia
3. Centre for Youth Mental Health, The University of Melbourne, Melbourne, Australia
4. Melbourne Neuropsychiatry Centre, Department of Psychiatry, The University of Melbourne, Carlton South, Victoria, Australia
5. Western Centre for Health Research & Education, Western Hospital Sunshine, The University of Melbourne, St Albans, Victoria, Australia
6. Telethon Kids Institute, The University of Western Australia, Perth, Australia
7. Child Health Research Centre, The University of Queensland, Queensland, Australia
8. School of Psychology, The University of Birmingham
9. Department of Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation, University of Groningen, University Medical Center Groningen
10. Institute for Mental and Physical Health and Clinical Translation, Deakin University, Melbourne, Australia

Address for Correspondence:

Simon Hartmann,
Adelaide Medical School, Discipline of Psychiatry
Adelaide, SA 5005, Australia.
Email: simon.hartmann@adelaide.edu.au.

Abstract

Background: Multimodal modelling that combines biological and clinical data shows promise in predicting transition to psychosis in individuals at ultra high risk (UHR). Individuals who transition to psychosis are known to have deficits at baseline in cognitive function and reductions in grey matter volume in multiple brain regions identified by magnetic resonance imaging (MRI).

Methods: In this study, we used Cox proportional hazards regression models to assess the additive predictive value of each modality – cognition, cortical structure information, and the neuroanatomical measure of "brain age gap" – to a previously developed clinical model using functioning and duration of symptoms prior to service entry as predictors in the Personal Assessment and Crisis Evaluation (PACE) 400 cohort. The PACE 400 study is a well characterised cohort of Australian youth identified as UHR using the Comprehensive Assessment of At Risk Mental States followed for up to 18 years, containing clinical data (from N=416 participants), cognitive data (N=213), and MRI cortical parameters extracted using Freesurfer (N=231).

Results: The results showed that neuroimaging, brain age gap, and cognition added marginal predictive information to the previously developed clinical model (fraction of new information: neuroimaging 0-12%, brain age gap 7%, cognition 0-16%).

Conclusions: In summary, adding a second modality to a clinical risk model predicting the onset of a psychotic disorder in the PACE 400 cohort showed little improvement in fit of the model for long term prediction of transition to psychosis.

Introduction

The development of criteria for ultra high risk (UHR) of psychosis has facilitated early intervention strategies to promote better clinical outcome[1]. Although there is meta-analytic evidence that 25% of UHR individuals transition to first episode psychosis over a 3-year period[2], we are currently unable to identify the level of risk at the individual level. Being able to do this would enable individualised treatment strategies to be developed using currently available treatments and also enable efficient stratification of UHR individuals in clinical trials of new treatments.

The majority of approaches to date that have attempted to generate individualised prediction models use either traditional multivariate techniques such as Cox proportional hazard[3]–[6] and logistic regression[7], [8], or machine learning models such as support vector machines[9]–[11] and greedy algorithms[12]. Recently, prediction models that combine multiple domains such as clinical, structural magnetic resonance imaging (MRI), cognition, genetic markers, and blood markers have been shown to improve psychosis prediction accuracy in UHR cohorts e.g., as demonstrated by the PRONIA consortium in recent studies using multimodal, multisite machine learning models[11], [13]. Such multimodal models can provide important information regarding the value of more expensive and complex assessment workflows including genomic testing and MRI as compared to structured clinical and cognitive assessments[14], [15]. To drive the implementation of prediction models in practice, there is a need to understand the benefit of including complex assessments as a low number of predictors or modalities, in particular non-invasive, lowers the difficulty of translation into clinical practice and should be included as objective during the development of prediction models besides a high predictive accuracy. Here, we validate the new information introduced by new predictors in a nested Cox regression model by determining the fraction of new information added to the total predictive information over an extended follow up period investigating the relevance of adding complex modalities.

Clinical variables known to predict transition to psychosis in UHR cohorts include: long duration of symptoms prior to presentation to clinical services[20], [21], severity of positive[22], [23] and negative psychotic symptoms[24], [25], poor functioning and quality of life[26], [27]. Cognition is impaired across domains in UHR and is a key prognostic biomarker of transition to first episode psychosis (FEP)[28]. Neuroimaging studies have found the surface area in rostral anterior cingulate, lateral and medial prefrontal regions, and parahippocampal gyrus[29], the mean anterior genu thickness[30], and the cortical thinning rate[31] to be predictive of transition to psychosis. One relatively new imaging concept, *brain age gap,* shows potential for prediction for transition to FEP[32]. Magnetic Resonance Imaging (MRI) scans can be used to estimate an individual's brain age by using prediction models that were trained on normative population data[60]. Brain age gap refers to the difference

between the estimate of an individual's brain age and the individual's chronological age[61]. A positive brain age gap indicates an 'older' brain as compared to the chronological age whereas a negative brain age gap suggests a 'younger' brain. Brain age gap has been part of an increasing number of studies over the past decade showing that higher brain age gap scores are associated with cognitive impairment and with schizophrenia or bipolar disorder [11], [33], [62]–[64].

In the current analysis, we investigated the potential benefit of using a multimodal model – compared to a clinical risk model alone – to estimate the transition hazard in UHR individuals using the PACE 400 data set. The aim was to assess the individual additive predictive value of cognition, cortical structure information, and brain age gap to a clinical Cox proportional hazards model developed by Nelson *et al*.[16]. The clinical model consisted of poor functioning (Global Assessment of Functioning, GAF), duration of symptoms prior to service entry, and UHR subgroup. The aim of this study was to quantify the benefits of including additional modalities in predicting transition to FEP in the PACE 400 cohort rather than finding the most generalisable prediction model.

Method

The PACE 400 study

The PACE 400 study is the first long-term follow-up of a UHR cohort (up to 15 years after entry to the Personal Assessment and Crisis Evaluation [PACE] clinic). The PACE 400 cohort[16] (N = 416) comprised all UHR patients participating across seven studies (3 intervention[34]–[36], 4 cohort[37]–[40]) at the PACE clinic in Melbourne, Australia, between 1993 and 2006.

The enrollment criteria and assessment of UHR status at baseline are outlined in the supplemental material. The main outcome of interest in the PACE 400 study was transition to psychotic disorder. Details on how the psychosis status was determined in the PACE 400 study are described in the supplemental material. Time to follow-up ranged from 2.4 to 18.6 years after baseline with a mean follow-up time of 7.5 years (SD = 3.2 years)[16]. The study combined individual information from multiple substudies across multiple domains including clinical assessments, cognition, neuroimaging, and in some cases fluid biospecimens. Previous studies have investigated cortical structure in the PACE 400 cohort but either in a smaller cohort[17] or only in individuals who did not transition[18]. Further, cognitive predictors in the PACE 400 cohort were previously

assessed in a study [19] but not in terms of the model fit.

Measures

Clinical measures

At baseline, negative symptoms were assessed using the Scale of Assessment for Negative Symptoms (SANS)[41], positive symptoms with the Brief Psychiatric Rating Scale, psychotic subscale (BPRS)[42] and the Comprehensive Assessment of At Risk Mental States (CAARMS)[1], and depressive symptoms using the Hamilton Rating Scale for Depression[43].

Functioning

Functioning was determined using the Quality of Life Scale (QLS)[44] and the Global Assessment of Function (GAF)[45].

Structural imaging

Details on MRI scanners used for MRI acquisition, cortical reconstruction, and volumetric segmentation using FreeSurfer[46] are outlined in the supplemental material.

The neuroimaging measures demonstrated a large variance between scanner sites due to different types of scanners that were deployed (see Figure S1). We applied the ComBat method[49] prior to our analysis to harmonize neuroimaging measures across sites. The ComBat method assumes an additive and multiplicative scanner or site effect which can be estimated from the data using conditional posterior means and subsequently be removed[50]. Hence, ComBat requires a sufficient sample size from each site or scanner to successfully estimate the multiplicative effect. The outcome measure of transitioning to FEP was included as covariate to align the distributions of individuals transitioning to FEP and individuals who did not transition across sites[50]. To reduce the dimension of the feature space for each neuroimaging domain, we applied bilateral principal component analysis (PCA) [51] to maximize the variance in the data. We also included cortical thickness values for fusiform, superior temporal, and paracentral regions as candidate predictors as they have been associated with psychosis conversion in the ENIGMA clinical high risk (CHR) for psychosis initiative[52].

Cognition

IQ at baseline was measured using a range of age-appropriate scales across studies[53] including the Wechsler Adult Intelligence Scale—Revised (WAIS-R)[54], the Wechsler Abbreviated Scale of Intelligence[55], or the

Wechsler Intelligence Scale for Children[56]. Verbal list learning and memory was assessed with the Rey Auditory Verbal Learning Test (RAVLT)[57]. Here, we used the age-adjusted scores for WAIS-R subtests Arithmetic and Digit Symbol Coding as well as the total score from a three-trial version of the RAVLT as cognitive predictors. Verbal learning and memory (RAVLT), processing speed (Digit Symbol Coding), and auditory verbal working memory (Arithmetic) have shown strong associations with transition to psychosis and changes in functioning in previous studies [28], [58], [59].

Brain age gap

We used the publicly available pre-trained ENIGMA brain age model (https://photon-ai.com/enigma_brainage) to estimate the brain age in the PACE 400 cohort. The model was trained using ridge regression to estimate normative models of the association between chronological age and 14 subcortical gray matter regions (nucleus accumbens, amygdala, caudate, hippocampus, pallidum, putamen, and thalamus), 2 lateral ventricles, 68 cortical thickness, and 68 surface area measures, and total intracranial volume in a healthy sample of 952 males (16 scanning sites) and 1236 females (22 scanning sites) aged 18–75 years[65]. Standardised protocols were used for image processing and feature extraction across sites (http://enigma.ini.usc.edu/protocols/imaging-protocols/). To control for regression dilution, a common phenomenon in brain age prediction models resulting in a systematic overestimation of the brain age for younger individuals and a systematic underestimation of the brain age for older individuals[66], we included chronological age as covariate in our analysis as suggested by the ENIGMA brain age model[65]. An overview of the estimated brain age gap for individuals with neuroimaging in the PACE 400 cohort using the ENIGMA Photon Brain Age Model without correction and with correction by removing the linear trend caused by chronological age is shown in Figure S2 in the supplemental material.

Model

Survival analysis was applied to analyse transition to FEP. Cox proportional hazards regression[67] was used to investigate the predictive value of clinical predictors combined with cognition, neuroimaging, or brain age gap. We fitted a base model including three clinical variables as well as “enhanced” models that added one of the following modalities: CAARMS subscales, cognition, MRI, or brain age gap. For each additional modality, we added intially one additional predictor and in a further analysis a maximum of two predictors to remain within a maximum of five predictors (three clinical predictors plus two predictors for each additional modality) resulting in 10-15 events per predictor[68]–[72].The analysis plan is summarized in Figure 1.

Three predictors GAF, duration of symptoms prior to service entry, and UHR subgroup were included in the base model based on univariate analysis in Nelson *et al*[16]. More information on the clinical predictors is provided in the supplemental material. With regard to the additional modalities, the CAARMS subscales disorders of thought content and conceptual disorganization, most significant additional variables identified in Nelson *et al.*[16], were included to provide additional information on the severity of positive psychotic symptoms and to control for the effect of adding additional variables to the base model. The first principal component of each MRI domain as well as cortical thickness values for left fusiform, right superior temporal, and left and right paracentral regions were included as neuroimaging predictors. As neuroanatomical predictor, brain age gap plus chronological age were included. For the subset of participants with cognition measures, the base model was repeated and compared in a separate sample with two enhanced models: first adding CAARMS subscales and second adding cognition predictors measured by age-adjusted scores for Wechsler subtests Arithmetic and Digit Symbol Coding and the RAVLT total score. Due to a difference in the number of participants with neuroimaging and cognition data, we performed our analysis in separate neuroimaging and cognition data sets from the PACE 400 sample. Each model was internally validated using bootstrapping (1,000 samples)[74].

The enhanced models were then compared to the base model (GAF, duration of symptoms prior to service entry, and UHR subgroup) to assess the additional predictive value of each modality. For each enhanced model, the likelihood-ratio (LR) test (LRT) for added value was obtained by comparing log likelihoods of the base and full models. The significance level for P values from the LRT was <0.05. We also determined the fraction of new information as the proportion of total predictive information that was added by cognition, MRI predictors, or brain age gap. More information on the calculation of the fraction of new information is provided in the supplemental material. Prior to the analysis, we checked if all included variables in the analysis and the variable describing the treatment groups (treatment-as-usual participants and participants that received trial treatments) satisfy the proportional hazards assumption. The analysis was performed in R 4.1.1[75] using the rms[76] and glmnet[77] package. Code for this analysis will be made available at https://github.com/preempt-centre-for-research-excellence/MultiPredModelPACE400.

Results

Table 1 details the descriptive statistics of the neuroimaging and cognition sample at baseline and follow-up. A total of 212 UHR individuals (49% female) were included in the neuroimaging data set (age at baseline (mean ±

SD) = 19 ± 5 years). There were 65 transitioned cases (31%) in the neuroimaging sample with an average time to transition of 168 days (SD = 461 days). The cognition data set contained a total of 94 UHR individuals (51% female) with an average age of 21 years (SD = 3.5 years). In the cognition sample, there were 39 transitioned cases (41%) with an average time to transition of 217 days (SD = 528 days). The demographic and clinical characteristics of the n=416 total sample have been reported and discussed in detail in a previous publication[16].

Clinical measures plus neuroimaging

Table 2 lists the regression coefficients and test scores after internal validation using bootstrapping for clinical and neuroimaging variables in a multivariate Cox regression model to predict transition to FEP in PACE 400. The base model with GAF, duration of symptoms prior to service entry, and UHR subgroup as predictors in the neuroimaging sample achieved a concordance index (CI) of 0.68. There was strong evidence that all three individual predictors had an effect on the risk for transition to FEP (GAF: HR=0.51 [95% CI: 0.33,0.71], P=0.001; Duration of symptoms prior to service entry – log transformed: HR=1.68 [95% CI: 1.16,2.64], P=0.015; UHR subgroup – BLIPS vs. Vulnerability: HR=3.13 [95% CI: 1.26,10.50], P=0.017; UHR subgroup – Attenuated Psychosis vs. Vulnerability: HR=1.28 [95% CI: 0.62,3.55], P = 0.017).

The addition of the CAARMS subscales 'disorders of thought content' or 'conceptual disorganization' to the base model increased the model fit by 3-4% adding marginal new information (LRT disorders of thought content: P=0.240, LRT conceptual disorganization; P=0.302). The addition of the first bilateral principal component of cortical surface area, curvature, volume, or thickness did not add new information to the clinical model (LRT surface area: P=0.946, LRT curve: P=0.789, LRT volume: P=0.687, LRT thickness: P=0.463). Subsequently, a combination of the first principal component of thickness and volume or the first and second principal components for cortical thickness, as cortical thickness and volume appeared to add the most information to the clinical model out of the four cortical domains, resulted in a marginal increase of new information (2%) with no effect (LRT thickness & volume: P=0.752, LRT thicknes 1. & 2. PCA: P=0.701) (see Table S1).

Out of the four individual regions identified in the ENIGMA clinical high risk for psychosis initiative to be associated with psychosis conversion, cortical thickness for the right paracentral region added the most new

information to the clinical model (7%) with an increase in CI to 0.69. However, adding the regional cortical thickness values individually as predictors to the base model did not have a significant effect on the model fit (LRT right paracentral: P=0.119, LRT left paracentral: P=0.131, LRT right superior temporal: P=0.412, LRT left fusiform: P=0.137). The largest addition of new information of 12% to the base model was achieved by adding cortical thickness values of left paracentral and left fusiform together, although with small effect (LRT: P=0.101) (see Table S1).

Clinical measures plus brain age gap

Table 3 lists the regression coefficients and test scores after internal validation using bootstrapping for clinical and brain age gap variables. Adding brain age gap and chronological age to the clinical model resulted in 7% of new information and an increase in CI to 0.69, although not significant (LRT: P=0.291). The fraction of new information was predominantly due to the addition of age as shown by the individual analysis in Table 3.

Clinical measures plus cognition

Table 4 lists the regression coefficients and test scores after internal validation using bootstrapping for clinical and cognition variables. The base model with GAF, duration of symptoms prior to service entry, and UHR subgroup as predictors in the cognition sample achieved a concordance index (CI) of 0.69. As opposed to the base model in the neuroimaging data set, there was strong evidence that in the base model only GAF had an effect on the risk of transition to FEP (GAF: HR=0.33 [0.15,0.54], P=0.001) but not duration of symptoms prior to service entry or UHR subgroup categories. Similar to the results in the neuroimaging data set, the addition of the CAARMS subscales disorders of thought content or conceptual disorganization to the base model only marginally increased the model fit by 1-4% (LRT disorders of thought content: P=0.305, LRT conceptual disorganization: P=0.605) with no improvement in CI. Adding RAVLT total score and the age-adjusted scores for Arithmetic and Digit Symbol Coding individually (+0-9%, LRT Digit Symbol Coding: P=0.492, LRT Arithmetic: P=0.113, LRT RAVLT total: P=0.975) or combined (+2-16%, see Table S2) as cognitive predictors to the base model did not result in any large improvement of the model fit.

Discussion

In this study, we assessed the predictive value of additional modalities including cognition, structural neuroimaging, or the neuroanatomical measure brain age gap to a base clinical model of transition to FEP (GAF, duration of symptoms prior to service entry, and UHR subgroup) in the PACE 400 sample, derived using Cox proportional hazards regression models. The cognitive variables, verbal learning and memory (RAVLT), processing speed (Digit Symbol Coding), and auditory verbal working memory (Arithmetic), added marginal additional predictive information to the clinical model. Additionally, the addition of neuroimaging measures such as cortical surface area, curvature, volume, or thickness resulted in no significant improvement of the model fit and accuracy. The neuroimaging composite measure brain age gap plus chronological age added 7% of new information and increased CI from 0.68 to 0.69, but this effect was predominantly a result of the addition of chronological age as a predictor rather than specific differences in brain structure.

Previous studies have shown that multimodal approaches, particularly machine learning models, may help to more accurately estimate the individual transition risk in UHR samples compared to unimodal approaches[11], [15], [78], [79]. Most commonly, the complementary predictive value of cognition, neuroimaging, and genetic features have been investigated. Our results suggest that the combination of MRI and clinical assessment only marginally improved the fit of a psychosis transition prediction model in the PACE 400 cohort. The combination of neuroimaging with the base clinical model resulted in a similar model fit and CI when controlling for adding the next significant clinical variables identified in Nelson *et al*.[16], the CAARMS subscales for disorders of thought content and conceptual disorganization.

The discrepancy in outcomes with previous multimodal UHR studies could possibly be related to the heterogeneity of the PACE 400 cohort in that it is a collection of cohort studies and clinical trials conducted over an extended period of time (14 years). Moreover, studies investigating the benefit of multimodal prediction models either suggest only a marginal improvement compared to unimodal approaches[78] or use a small sample size[79] resulting in a strong risk of misestimation[80], [81]. More promising results have been achieved when stacking different modalities[15] e.g., using generalised stacked models[11], as stacking determines how to optimally combine the predictions from each modality. However, stacked Cox proportional hazards regression models are particularly complex due to the inclusion of time-to-event information. Furthermore, previous studies suggest that the change in cortical structure, especially cortical thickness, may potentially be a more

suitable predictor for transition to psychosis compared to cortical measures assessed at baseline[31], [82], [83].

The addition of cognitive measures to the clinical model did not result in an improvement of fit. Our results are consistent with previous studies analysing the predictive value of cognition in the PACE 400 cohort[19], [58] showing that cognition is not a strong predictor of transition to psychosis. Our results on the additive predictive value of cognition are restricted by the differences in cognitive batteries across studies and the resultant small size of the cognition sample in this study. The neuroanatomical measure brain age gap did not improve model accuracy when considered alongside age. Our findings align with the results from the North American Prodromal Longitudinal Study that found that the predictive variance of brain age gap overlapped entirely with age[4]. Further, our results in terms of brain age gap are limited by the usage of an publicly available external model that was trained on healthy individuals aged 18-75 years. Although the ENIGMA Photon Brain Age model has proven to be accurate in a previous study, the lack of a validation sample in our cohort and the slightly different age range with an average age of 20 years (SD=±3.5 years) may have potentially influenced our results. Finally, the predictive value of brain age gap for transition to psychosis could be reduced by the large age range in our sample resulting in a discrepancy of predictive information in similar brain age values for younger and older participants. This could be accounted for by dividing the sample into age groups besides adding age as covariate, but was not performed in this sample due to the low sample size.

Our study is limited by the low number of events (transitioned cases) in the subset of the PACE 400 cohort that had cognitive or neuroimaging data available, highlighting a key drawback to adding modalities to the structured clinical assessment routine as they multiply the costs and workload of the assessment. Potentially, the low number of events could partially explain the lack of predictive benefit of multimodal models observed in this study as the characteristics of individuals with complete data may differ from individuals who were excluded from this study. A low number of events restricts the number of potential predictors, the optimization of the model fit, and the validation of the fitted model. Additionally, we did not account for nonstandard treatment due to randomization to intervention trials that are part of the PACE 400 sample as testing the proportional hazards assumption did not indicate a need for stratification based on treatment received. Moreover, a sensitivity analysis in the original PACE 400 study indicated the same results for the treatment-as-usual participants (i.e., excluding 244 who had received trial treatments) and the entire cohort[16]. Another limitation

is the possibility that some transitioned cases were not detected, i.e., if they were unavailable for interview and had not attended a public mental health service[16].

Another major limitation of our study was the heterogeneity in neuroimaging measures due to different MRI scanners used at assessment sites. Figure S1 illustrates the inherited bias across sites for cortical thickness in right rostral middle frontal region. Harmonizing the neuroimaging measures across scanner sites using the ComBat method successfully removed the site bias although the ComBat method has shown to have the potential of causing distortion in the absence of a scanner or site effect [84] and is outperformed by traveling-subject based harmonization methods [85]. The need for harmonization raises a number of questions with regard to the clinical application of multimodal models for the prediction of transition to psychosis in UHR individuals. Harmonization performs well during the implementation and evaluation phase of a model as the distributions of each scanner or site can be determined in the training and test set. However, harmonization in a clinical application relies on a-priori knowledge of the deployed scanner to remove the inherited bias. Moreover, there is no agreed upon way to standardize MRI measures within a cross-validation framework used to train machine learning models[63]. Thus, the heterogeneity in MRI measures across sites and scanners severely limits the broad clinical applicability of multimodal prediction models that include neuroimaging and highlights the need for local recalibrations of models.

In sum, our results show that the inclusion of neuroimaging or cognitive information in a risk model that estimates the proportional hazard of transition to psychosis in UHR subjects in the PACE 400 study appears to add little information to improve the fit of the clinical based model. Hence, these findings raise the question whether adding baseline cognitive and structural MRI assessments provides sufficient additional predictive information to warrant the associated computational and economical costs, and the increased workflow complexity of actioning these assessments in a clinical setting besides their value in a clinical setting. However, it is important to acknowledge that our findings are limited by the constraints on methodological choices given by the nature of the cohort which could potentially lowered the importance of our findings.

Acknowledgement

This work was funded through the Prediction of Early Mental Disorder and Preventive Treatment (PRE-EMPT,

www.pre-empt.org.au) - Centre of Research Excellence (NHMRC grant: #1198304).

Conflict of Interest/Financial Disclosure

PDM reported grants from the National Institute of Mental Health during the conduct of the study; in addition, Dr McGorry had a patent for AU 2015203289 issued, a patent for US 9884034 issued, a patent for US 15/844444 issued, and a patent for CA 2773031 issued; and he has received past unrestricted grant funding from Janssen-Cilag, AstraZeneca, Eli Lilly, Novartis, and Pfizer, and honoraria for consultancy and teaching from Janssen-Cilag, Eli Lilly, Pfizer, AstraZeneca, Roche, Bristol Meyers Squibb, and Lundbeck. He has received grant funding from the Colonial Foundation, the National Health and Medical Research Council of Australia, Australian Research Council, NARSAD, the Stanley Foundation, NIH, the Wellcome Trust, and the Australian and Victorian governments. SC received speaker/consultation fees from: Janssen-Cilag, Lundbeck, Otsuka and Servier and research funding from Janssen-Cilag, Lundbeck, Otsuka and Gilead. BN was supported by a NHMRC Senior Research Fellowship (#1137687) and a University of Melbourne Dame Kate Campbell Fellowship—all unrelated to this work. CP was supported by a National Health and Medical Research Council (NHMRC) L3 Investigator Grant (#1196508) outside the submitted work. KA was supported by a NHMRC Career Development Fellowship (#1141207) and a University of Melbourne Dame Kate Campbell Fellowship. KA has received funding from the NHMRC, MRFF and Wellcome Trust—all unrelated to this work. AL was supported by an NHMRC Emerging Leadership Fellowship (#2010063). ARY was supported by an NHMRC Principal Research Fellowship (#1136829). GPA was supported by an NHMRC Senior Research Fellowship (#1080963). JTWW. was funded by a Netherlands Organization for Scientific Research (NWO) Veni grant no. 016.156.019. All other authors report no biomedical financial interests or potential conflicts of interest.

References

[1] A. R. Yung *et al.*, “Mapping the Onset of Psychosis: The Comprehensive Assessment of At-Risk Mental States,” *Aust. New Zeal. J. Psychiatry*, vol. 39, no. 11–12, pp. 964–971, Nov. 2005, doi: 10.1080/j.1440-1614.2005.01714.x.

[2] G. Salazar de Pablo *et al.*, “Probability of Transition to Psychosis in Individuals at Clinical High Risk: An Updated Meta-analysis,” *JAMA Psychiatry*, vol. 78, no. 9, pp. 970–978, Sep. 2021, doi: 10.1001/jamapsychiatry.2021.0830.

[3] J. Addington, L. Liu, D. O. Perkins, R. E. Carrion, R. S. E. Keefe, and S. W. Woods, “The Role of Cognition and Social Functioning as Predictors in the Transition to Psychosis for Youth With Attenuated Psychotic

Symptoms," *Schizophr. Bull.*, vol. 43, no. 1, pp. 57–63, Jan. 2017, doi: 10.1093/schbul/sbw152.

[4] Y. Chung *et al.*, "Adding a neuroanatomical biomarker to an individualized risk calculator for psychosis: A proof-of-concept study," *Schizophr. Res.*, vol. 208, pp. 41–43, 2019, doi: https://doi.org/10.1016/j.schres.2019.01.026.

[5] T. D. Cannon *et al.*, "An Individualized Risk Calculator for Research in Prodromal Psychosis," *Am. J. Psychiatry*, vol. 173, no. 10, pp. 980–988, Oct. 2016, doi: 10.1176/appi.ajp.2016.15070890.

[6] P. Fusar-Poli *et al.*, "Development and Validation of a Clinically Based Risk Calculator for the Transdiagnostic Prediction of Psychosis," *JAMA Psychiatry*, vol. 74, no. 5, pp. 493–500, May 2017, doi: 10.1001/jamapsychiatry.2017.0284.

[7] T. Zhang *et al.*, "Prediction of psychosis in prodrome: development and validation of a simple, personalized risk calculator," *Psychol. Med.*, vol. 49, no. 12, pp. 1990–1998, 2019, doi: DOI: 10.1017/S0033291718002738.

[8] C. M. Corcoran *et al.*, "Prediction of psychosis across protocols and risk cohorts using automated language analysis," *World Psychiatry*, vol. 17, no. 1, pp. 67–75, Feb. 2018, doi: https://doi.org/10.1002/wps.20491.

[9] N. Koutsouleris *et al.*, "Use of Neuroanatomical Pattern Classification to Identify Subjects in At-Risk Mental States of Psychosis and Predict Disease Transition," *Arch. Gen. Psychiatry*, vol. 66, no. 7, pp. 700–712, Jul. 2009, doi: 10.1001/archgenpsychiatry.2009.62.

[10] N. Koutsouleris *et al.*, "Toward Generalizable and Transdiagnostic Tools for Psychosis Prediction: An Independent Validation and Improvement of the NAPLS-2 Risk Calculator in the Multisite PRONIA Cohort," *Biol. Psychiatry*, vol. 90, no. 9, pp. 632–642, 2021, doi: https://doi.org/10.1016/j.biopsych.2021.06.023.

[11] N. Koutsouleris *et al.*, "Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression," *JAMA Psychiatry*, vol. 78, no. 2, pp. 195–209, Feb. 2021, doi: 10.1001/jamapsychiatry.2020.3604.

[12] D. O. Perkins *et al.*, "Towards a Psychosis Risk Blood Diagnostic for Persons Experiencing High-Risk Symptoms: Preliminary Results From the NAPLS Project," *Schizophr. Bull.*, vol. 41, no. 2, pp. 419–428, Mar. 2015, doi: 10.1093/schbul/sbu099.

[13] N. Koutsouleris *et al.*, "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*, vol. 75, no. 11, pp. 1156–1172, Nov. 2018, doi: 10.1001/jamapsychiatry.2018.2165.

[14] S. R. Clark, K. O. Schubert, and B. T. Baune, "Towards indicated prevention of psychosis: using probabilistic assessments of transition risk in psychosis prodrome," *J. Neural Transm.*, vol. 122, no. 1, pp. 155–169, 2015, doi: 10.1007/s00702-014-1325-9.

[15] A. Schmidt *et al.*, "Improving Prognostic Accuracy in Subjects at Clinical High Risk for Psychosis: Systematic Review of Predictive Models and Meta-analytical Sequential Testing Simulation," *Schizophr. Bull.*, vol. 43, no. 2, pp. 375–388, Mar. 2017, doi: 10.1093/schbul/sbw098.

[16] B. Nelson *et al.*, "Long-term Follow-up of a Group at Ultra High Risk ('Prodromal') for Psychosis: The PACE 400 Study," *JAMA Psychiatry*, vol. 70, no. 8, pp. 793–802, Aug. 2013, doi: 10.1001/jamapsychiatry.2013.1270.

[17] M. Rapado-Castro *et al.*, "Does cortical brain morphology act as a mediator between childhood trauma and transition to psychosis in young individuals at ultra-high risk?," *Schizophr. Res.*, vol. 224, pp. 116–125, 2020, doi: https://doi.org/10.1016/j.schres.2020.09.017.

[18] V. L. Cropley *et al.*, "Baseline grey matter volume of non-transitioned 'ultra high risk' for psychosis individuals with and without attenuated psychotic symptoms at long-term follow-up," *Schizophr. Res.*, vol. 173, no. 3, pp. 152–158, 2016, doi: https://doi.org/10.1016/j.schres.2015.05.014.

[19] A. Lin *et al.*, "Neurocognitive predictors of transition to psychosis: medium- to long-term findings from a sample at ultra-high risk for psychosis," *Psychol. Med.*, vol. 43, no. 11, pp. 2349–2360, 2013, doi: DOI: 10.1017/S0033291713000123.

[20] M. Penttilä, E. Jääskeläinen, N. Hirvonen, M. Isohanni, and J. Miettunen, "Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis," *Br. J. Psychiatry*, vol. 205, no. 2, pp. 88–94, 2014, doi: DOI: 10.1192/bjp.bp.113.127753.

[21] B. Nelson *et al.*, "Further examination of the reducing transition rate in ultra high risk for psychosis samples: The possible role of earlier intervention," *Schizophr. Res.*, vol. 174, no. 1, pp. 43–49, 2016, doi: https://doi.org/10.1016/j.schres.2016.04.040.

[22] M. P. Hengartner, K. Heekeren, D. Dvorsky, S. Walitza, W. Rössler, and A. Theodoridou, "Checking the predictive accuracy of basic symptoms against ultra high-risk criteria and testing of a multivariable prediction model: Evidence from a prospective three-year observational study of persons at clinical high-risk for psychosis," *Eur. Psychiatry*, vol. 45, pp. 27–35, 2017, doi: DOI: 10.1016/j.eurpsy.2017.05.026.

[23] T. Ziermans *et al.*, "Neurocognitive and Clinical Predictors of Long-Term Outcome in Adolescents at Ultra-High Risk for Psychosis: A 6-Year Follow-Up," *PLoS One*, vol. 9, no. 4, p. e93994, Apr. 2014, [Online]. Available: https://doi.org/10.1371/journal.pone.0093994.

[24] D. Piskulic *et al.*, "Negative symptoms in individuals at clinical high risk of psychosis," *Psychiatry Res.*, vol. 196, no. 2, pp. 220–224, 2012, doi: https://doi.org/10.1016/j.psychres.2012.02.018.

[25] L. R. Valmaggia *et al.*, "Negative psychotic symptoms and impaired role functioning predict transition outcomes in the at-risk mental state: a latent class cluster analysis study," *Psychol. Med.*, vol. 43, no. 11, pp. 2311–2325, 2013, doi: DOI: 10.1017/S0033291713000251.

[26] A. Malla and J. Payne, "First-Episode Psychosis: Psychopathology, Quality of Life, and Functional Outcome," *Schizophr. Bull.*, vol. 31, no. 3, pp. 650–671, Jan. 2005, doi: 10.1093/schbul/sbi031.

[27] A. R. Yung *et al.*, "Testing the Ultra High Risk (prodromal) criteria for the prediction of psychosis in a clinical sample of young people," *Schizophr. Res.*, vol. 84, no. 1, pp. 57–66, 2006, doi: https://doi.org/10.1016/j.schres.2006.03.014.

[28] A. Catalan *et al.*, "Neurocognitive Functioning in Individuals at Clinical High Risk for Psychosis: A Systematic Review and Meta-analysis," *JAMA Psychiatry*, vol. 78, no. 8, pp. 859–867, Aug. 2021, doi: 10.1001/jamapsychiatry.2021.1290.

[29] Y. Chung *et al.*, "Cortical abnormalities in youth at clinical high-risk for psychosis: Findings from the NAPLS2 cohort," *NeuroImage Clin.*, vol. 23, p. 101862, 2019, doi: https://doi.org/10.1016/j.nicl.2019.101862.

[30] M. Walterfang *et al.*, "Corpus callosum shape alterations in individuals prior to the onset of psychosis," *Schizophr. Res.*, vol. 103, no. 1, pp. 1–10, 2008, doi: https://doi.org/10.1016/j.schres.2008.04.042.

[31] M. A. Collins *et al.*, "Accelerated cortical thinning precedes and predicts conversion to psychosis: The NAPLS3 longitudinal study of youth at clinical high-risk," *Mol. Psychiatry*, 2022, doi: 10.1038/s41380-022-01870-7.

[32] Y. Chung *et al.*, "Use of Machine Learning to Determine Deviance in Neuroanatomical Maturity Associated With Future Psychosis in Youths at Clinically High Risk," *JAMA Psychiatry*, vol. 75, no. 9, pp. 960–968, Sep. 2018, doi: 10.1001/jamapsychiatry.2018.1543.

[33] P. L. Ballester *et al.*, "Brain age in mood and psychotic disorders: a systematic review and meta-analysis," *Acta Psychiatr. Scand.*, vol. 145, no. 1, pp. 42–55, Jan. 2022, doi: https://doi.org/10.1111/acps.13371.

[34] P. D. McGorry *et al.*, "Randomized Controlled Trial of Interventions Designed to Reduce the Risk of Progression to First-Episode Psychosis in a Clinical Sample With Subthreshold Symptoms," *Arch. Gen. Psychiatry*, vol. 59, no. 10, pp. 921–928, Oct. 2002, doi: 10.1001/archpsyc.59.10.921.

[35] A. R. Yung *et al.*, "Randomized controlled trial of interventions for young people at ultra high risk for psychosis: 6-month analysis," *J. Clin. Psychiatry*, vol. 71, no. 4, p. 16654, 2010.

[36] G. E. Berger *et al.*, "Neuroprotective Effects of Low-dose Lithium in Individuals at Ultra-high Risk for Psychosis. A Longitudinal MRI/MRS Study," *Current Pharmaceutical Design*, vol. 18, no. 4. pp. 570–575, 2012, doi: http://dx.doi.org/10.2174/138161212799316163.

[37] A. R. Yung, P. D. McGorry, C. A. McFarlane, H. J. Jackson, G. C. Patton, and A. Rakkar, "Monitoring and Care of Young People at Incipient Risk of Psychosis," *Schizophr. Bull.*, vol. 22, no. 2, pp. 283–303, Jan. 1996, doi: 10.1093/schbul/22.2.283.

[38] A. R. Yung *et al.*, "Psychosis prediction: 12-month follow up of a high-risk ('prodromal') group," *Schizophr. Res.*, vol. 60, no. 1, pp. 21–32, 2003, doi: https://doi.org/10.1016/S0920-9964(02)00167-6.

[39] K. N. Thompson *et al.*, "Stress and HPA-axis functioning in young people at ultra high risk for psychosis," *J. Psychiatr. Res.*, vol. 41, no. 7, pp. 561–569, 2007, doi: https://doi.org/10.1016/j.jpsychires.2006.05.010.

[40] L. J. Phillips *et al.*, "Randomized Controlled Trial of Interventions for Young People at Ultra-High Risk of Psychosis: Study Design and Baseline Characteristics," *Aust. New Zeal. J. Psychiatry*, vol. 43, no. 9, pp. 818–829, Jan. 2009, doi: 10.1080/00048670903107625.

[41] N. C. Andreasen, *Scale for the assessment of negative symptoms (SANS)*. Iowa City: University of Iowa, 1983.

[42] J. E. Overall and D. R. Gorham, "The brief psychiatric rating scale," *Psychol. Rep.*, vol. 10, no. 3, pp. 799–812, 1962.

[43] M. Hamilton, "A rating scale for depression," *J. Neurol. Neurosurg. Psychiatry*, vol. 23, no. 1, pp. 56–62, Feb. 1960, doi: 10.1136/jnnp.23.1.56.

[44] D. W. Heinrichs, T. E. Hanlon, and W. T. Carpenter Jr., "The Quality of Life Scale: An Instrument for Rating the Schizophrenic Deficit Syndrome," *Schizophr. Bull.*, vol. 10, no. 3, pp. 388–398, Jan. 1984, doi: 10.1093/schbul/10.3.388.

[45] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*, 4th ed. Washington, D.C.: American Psychiatric Association, 1994.

[46] B. Fischl, "FreeSurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012, doi: https://doi.org/10.1016/j.neuroimage.2012.01.021.

[47] X. Han *et al.*, "Reliability of MRI-derived measurements of human cerebral cortical thickness: The effects of field strength, scanner upgrade and manufacturer," *Neuroimage*, vol. 32, no. 1, pp. 180–194, 2006, doi: https://doi.org/10.1016/j.neuroimage.2006.02.051.

[48] M. Reuter, N. J. Schmansky, H. D. Rosas, and B. Fischl, "Within-subject template estimation for unbiased longitudinal image analysis," *Neuroimage*, vol. 61, no. 4, pp. 1402–1418, 2012, doi: https://doi.org/10.1016/j.neuroimage.2012.02.084.

[49] J.-P. Fortin *et al.*, "Harmonization of multi-site diffusion tensor imaging data," *Neuroimage*, vol. 161, pp. 149–170, 2017, doi: https://doi.org/10.1016/j.neuroimage.2017.08.047.

[50] F. Orlhac *et al.*, "A guide to ComBat harmonization of imaging biomarkers in multicenter studies," *J. Nucl. Med.*, p. jnumed.121.262464, Sep. 2021, doi: 10.2967/jnumed.121.262464.

[51] H. Hotelling, "Analysis of a complex of statistical variables into principal components.," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933, doi: 10.1037/h0071325.

[52] E. C. H. R. for P. W. Group, "Association of Structural Magnetic Resonance Imaging Measures With Psychosis Onset in Individuals at Clinical High Risk for Developing Psychosis: An ENIGMA Working Group Mega-analysis," *JAMA Psychiatry*, vol. 78, no. 7, pp. 753–766, Jul. 2021, doi: 10.1001/jamapsychiatry.2021.0638.

[53] A. Lin *et al.*, "Neurocognitive predictors of functional outcome two to 13years after identification as ultra-high risk for psychosis," *Schizophr. Res.*, vol. 132, no. 1, pp. 1–7, 2011, doi: https://doi.org/10.1016/j.schres.2011.06.014.

[54] D. Wechsler, "The psychometric tradition: Developing the wechsler adult intelligence scale," *Contemp. Educ. Psychol.*, vol. 6, no. 2, pp. 82–85, 1981, doi: https://doi.org/10.1016/0361-476X(81)90035-7.

[55] D. Wechsler, *Wechsler Abbreviated Scale of Intelligence (WASI)*. San Antonio, TX: Psychological Corporation, 1999.

[56] D. Wechsler, *Wechsler Intelligence Scale for Children*, 3rd ed. San Antonio, TX: Psychological Corporation, 1991.

[57] A. Rey, *L'examen clinique en psychologie. [The clinical examination in psychology.]*. Presses Universitaries De France, 1958.

[58] K. Allott *et al.*, "Longitudinal Cognitive Performance in Individuals at Ultrahigh Risk for Psychosis: A 10-year Follow-up," *Schizophr. Bull.*, vol. 45, no. 5, pp. 1101–1111, Sep. 2019, doi: 10.1093/schbul/sby143.

[59] E. Studerus, M. Papmeyer, and A. Riecher-Rössler, "Neurocognition and Motor Functioning in the Prediction of Psychosis," in *Key Issues in Mental Health*, vol. 181, 2016, pp. 116–132.

[60] S. M. Smith *et al.*, "Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations," *Elife*, vol. 9, p. e52677, 2020, doi: 10.7554/eLife.52677.

[61] K. Franke, G. Ziegler, S. Klöppel, and C. Gaser, "Estimating the age of healthy subjects from T1-weighted MRI scans using kernel methods: Exploring the influence of various parameters," *Neuroimage*, vol. 50, no. 3, pp. 883–892, 2010, doi: https://doi.org/10.1016/j.neuroimage.2010.01.005.

[62] T. Kaufmann *et al.*, "Common brain disorders are associated with heritable patterns of apparent aging of the brain," *Nat. Neurosci.*, vol. 22, no. 10, pp. 1617–1623, 2019, doi: 10.1038/s41593-019-0471-7.

[63] V. Drobinin, H. Van Gestel, C. A. Helmick, M. H. Schmidt, C. V Bowen, and R. Uher, "The Developmental Brain Age Is Associated With Adversity, Depression, and Functional Outcomes Among Adolescents," *Biol. Psychiatry Cogn. Neurosci. Neuroimaging*, Jan. 2022, doi: 10.1016/j.bpsc.2021.09.004.

[64] J. H. Cole, "Multimodality neuroimaging brain-age in UK biobank: relationship to biomedical, lifestyle, and cognitive factors," *Neurobiol. Aging*, vol. 92, pp. 34–42, 2020, doi: https://doi.org/10.1016/j.neurobiolaging.2020.03.014.

[65] L. K. M. Han *et al.*, "Brain aging in major depressive disorder: results from the ENIGMA major depressive disorder working group," *Mol. Psychiatry*, vol. 26, no. 9, pp. 5124–5139, 2021, doi: 10.1038/s41380-020-0754-0.

[66] T. T. Le *et al.*, "A Nonlinear Simulation Framework Supports Adjusting for Age When Analyzing BrainAGE," *Frontiers in Aging Neuroscience*, vol. 10. 2018, [Online]. Available: https://www.frontiersin.org/article/10.3389/fnagi.2018.00317.

[67] D. R. Cox, "Regression Models and Life-Tables," *J. R. Stat. Soc. Ser. B*, vol. 34, no. 2, pp. 187–202, Jan. 1972, doi: https://doi.org/10.1111/j.2517-6161.1972.tb00899.x.

[68] Q. Chen, H. Nian, Y. Zhu, H. K. Talbot, M. R. Griffin, and F. E. Harrell Jr., "Too many covariates and too few cases? – a comparative study," *Stat. Med.*, vol. 35, no. 25, pp. 4546–4558, Nov. 2016, doi: https://doi.org/10.1002/sim.7021.

[69] P. Peduzzi, J. Concato, A. R. Feinstein, and T. R. Holford, "Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates," *J. Clin. Epidemiol.*, vol. 48, no. 12, pp. 1503–1510, 1995, doi: https://doi.org/10.1016/0895-4356(95)00048-8.

[70] J. Concato, P. Peduzzi, T. R. Holford, and A. R. Feinstein, "Importance of events per independent variable in proportional hazards analysis I. Background, goals, and general strategy," *J. Clin. Epidemiol.*, vol. 48, no. 12, pp. 1495–1501, 1995, doi: https://doi.org/10.1016/0895-4356(95)00510-2.

[71] E. Vittinghoff and C. E. McCulloch, "Relaxing the Rule of Ten Events per Variable in Logistic and Cox Regression," *Am. J. Epidemiol.*, vol. 165, no. 6, pp. 710–718, Mar. 2007, doi: 10.1093/aje/kwk052.

[72] F. E. Harrell Jr., K. L. Lee, R. M. Califf, D. B. Pryor, and R. A. Rosati, "Regression modelling strategies for improved prognostic prediction," *Stat. Med.*, vol. 3, no. 2, pp. 143–152, Apr. 1984, doi: https://doi.org/10.1002/sim.4780030207.

[73] P. Fusar-Poli, "The Clinical High-Risk State for Psychosis (CHR-P), Version II," *Schizophr. Bull.*, vol. 43, no. 1, pp. 44–47, Jan. 2017, doi: 10.1093/schbul/sbw158.

[74] F. E. Harrell Jr, K. L. Lee, and D. B. Mark, "Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors," *Stat. Med.*, vol. 15, no. 4, pp. 361–387, 1996.

[75] R Core Team, "R: A Language and Environment for Statistical Computing." Vienna, Austria, 2021, [Online]. Available: https://www.r-project.org/.

[76] F. E. Harrell Jr, "rms: Regression Modeling Strategies." 2021, [Online]. Available: https://cran.r-project.org/package=rms.

[77] N. Simon, J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent," *J. Stat. Softw.*, vol. 39, no. 5, pp. 1–13, Mar. 2011, doi: 10.18637/jss.v039.i05.

[78] N. T. Doan *et al.*, "Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders," *NeuroImage Clin.*, vol. 15, pp. 719–731, 2017, doi: https://doi.org/10.1016/j.nicl.2017.06.014.

[79] E. Zarogianni, A. J. Storkey, E. C. Johnstone, D. G. C. Owens, and S. M. Lawrie, "Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features," *Schizophr. Res.*, vol. 181, pp. 6–12, 2017, doi: https://doi.org/10.1016/j.schres.2016.08.027.

[80] C. Flint *et al.*, "Systematic misestimation of machine learning performance in neuroimaging studies of depression," *Neuropsychopharmacology*, vol. 46, no. 8, pp. 1510–1517, 2021, doi: 10.1038/s41386-021-01020-7.

[81] H. G. Schnack and R. S. Kahn, "Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters," *Frontiers in Psychiatry*, vol. 7. p. 50, 2016, [Online]. Available: https://www.frontiersin.org/article/10.3389/fpsyt.2016.00050.

[82] C. Pantelis *et al.*, "Neuroanatomical abnormalities before and after onset of psychosis: a cross-sectional and longitudinal MRI comparison," *Lancet*, vol. 361, no. 9354, pp. 281–288, 2003, doi: https://doi.org/10.1016/S0140-6736(03)12323-9.

[83] D. Sun *et al.*, "Progressive brain structural changes mapped as psychosis develops in 'at risk' individuals," *Schizophr. Res.*, vol. 108, no. 1, pp. 85–92, 2009, doi: https://doi.org/10.1016/j.schres.2008.11.026.

[84] S. Richter *et al.*, "Validation of cross-sectional and longitudinal ComBat harmonization methods for magnetic resonance imaging data on a travelling subject cohort," *Neuroimage: Reports*, vol. 2, no. 4, p. 100136, 2022, doi: https://doi.org/10.1016/j.ynirp.2022.100136.

[85] N. Maikusa *et al.*, "Comparison of traveling-subject and ComBat harmonization methods for assessing structural brain characteristics," *Hum. Brain Mapp.*, vol. 42, no. 16, pp. 5278–5287, Nov. 2021, doi: https://doi.org/10.1002/hbm.25615.

Figure/Table Legends

Figure 1: Study plan using nested Cox proportional hazards regression models, internal validation, and nested models evaluation

Study plan using nested Cox proportional hazards regression models with a base model including Global Assessment of Functioning (GAF), duration of symptoms prior to service entry, and ultra high risk (UHR) group as clinical variables as well as "full" models that contain the three clinical variables and one of the additional modalities: neuroimaging (MRI), brain age gap, or cognition. We also fitted a model containing the three clinical variables plus the CAARMS subscales disorders of thought content and conceptual disorganization to control for the effect of adding two variables to the base model. The full models are subsequently compared to the base model using the likelihood-ratio test (LRT), the fraction of new information, and the concordance index (CI).

Table 1: Baseline and follow-up descriptive information about neuroimaging and cognition sample

	Cognition, N = 94[1]	MRI, N = 212[1]
Intervention treatment	24% (23/94)	37% (78/212)
Female	51% (48/94)	49% (104/212)
Age at baseline (years)	21.1+/-3.5	19.9+/-3.5
Time between symptom onset and first contact with PACE (days)	564.2+/-1,056.4	430.9+/-632.5
UHR subgroup		
any BLIPS	24% (23/94)	16% (34/212)
APS or APS+Vulnerability	59% (55/94)	68% (144/212)
Vulnerability	17% (16/94)	16% (34/212)
Clinical measures		
BPRS Total	43.4+/-7.5	46.2+/-9.0
SANS Total	18.3+/-13.6	19.3+/-11.9
GAF	62.3+/-14.5	59.2+/-11.7
QLS Total	73.3+/-22.3	75.1+/-20.3
CAARMS Disorders of Thought Content, severity	2.1+/-1.1	2.0+/-1.1
CAARMS Perceptual Abnormalities, severity	1.9+/-1.5	2.1+/-1.5
CAARMS Conceptual Disorganisation, severity	2.1+/-1.1	1.9+/-1.0
Cognition		
Coding	9.3+/-2.5	
Arithmetic	8.7+/-3.1	
RAVLT total	28.6+/-6.3	
Brain Age Gap		-0.3+/-7.4
Follow-up		
Transition to psychosis	41% (39/94)	31% (65/212)
Follow-up time (days)	4,055.3+/-320.6	3,052.8+/-1,066.0
Time to transition (days)	217.4+/-528.0	167.8+/-460.9
SOFAS at follow-up	62.9+/-17.0)	69.0+/-16.0

[1]% (n/N) ; Mean+/-SD (N)

Table 2: Hazard ratio (HR) and P value after internal validation (bootstrapping N = 1,000) for clinical and neuroimaging predictor variables in the multivariate Cox proportional hazards regression analysis of transition to psychosis in PACE 400 (n=212).

	Base		Base + Disorders of Thought Content		Base + Conceptual disorganization		Base + MRI surface PCA		Base + MRI curve PCA		Base + MRI volume PCA		Base + MRI thickness PCA		Base + Thickness Paracentral left		Base + Thickness Paracentral Right		Base + Thickness superior temporal right		Base + Thickness fusiform left	
	HR	P	HR	P	HR	P	HR	P	HR	P	HR	P	HR	P	HR	P	HR	P	HR	P	HR	P
Duration of symptoms prior to service entry – log transformed	**1.68**	**0.015***	**1.68**	**0.021***	**1.64**	**0.025***	**1.68**	**0.019***	**1.67**	**0.021***	**1.66**	**0.018***	**1.65**	**0.019***	**1.66**	**0.018***	**1.67**	**0.017***	**1.65**	**0.024***	**1.70**	**0.019***
GAF	**0.51**	**0.001***	**0.53**	**0.002***	**0.52**	**0.001***	**0.51**	**0.001***	**0.51**	**0.001***	**0.50**	**0.001***	**0.50**	**0.001***	**0.50**	**0.002***	**0.50**	**0.001***	**0.50**	**<0.001***	**0.51**	**<0.001***
UHR subgroup		**0.017***		0.050		**0.024***		**0.015***		**0.014***		**0.014***		**0.017***		**0.031***		**0.027***		**0.017***		**0.024***
BLIPS vs. Vulnerability	**3.13**		2.56		**3.08**		**3.12**		**3.12**		**3.10**		**3.04**		**3.10**		**2.87**		**3.05**		**2.99**	
Attenuated Psychosis vs. Vulnerability	**1.28**		1.11		**1.35**		**1.28**		**1.28**		**1.26**		**1.26**		**1.25**		**1.21**		**1.29**		**1.26**	
CAARMS - Disorders of thought content			1.37	0.292																		
CAARMS - Conceptual disorganization					1.14	0.389																
1. PC left hemisphere MRI variable							10.1399	0.893954	1.1705	0.666819	0.5393	0.522728	1.2912	0.579491	0.78	0.164	0.77	0.140	0.87	0.468	0.76	0.131
LR χ2		35.00		36.38		36.07		35.00		35.07		35.16		35.54		37.44		37.28		35.67		37.21
Fraction of new information		-		0.04		0.03		0.00		0.00		0.01		0.02		0.07		0.06		0.02		0.06
Concordance index (CI)		0.68		0.68		0.68		0.68		0.68		0.68		0.68		0.69		0.69		0.68		0.68

For each full model, the likelihood-ratio (LR) test was obtained by comparing log likelihoods of the base and full models. The fraction of new information is the proportion of total predictive information in clinical plus MRIthat was added by MRI. It was calculated as follows: 1 – Base LR χ2/Full LR χ2. HR, Hazard ratio; PC, principal component; LR χ2, likelihood ratio Chi-squared statistics; *significance level for LR test P values: $P < 0.05$

Table 3: Hazard ratio (HR) and P value after internal validation (bootstrapping N = 1,000) for clinical and brain age gap predictor variables in the multivariate Cox proportional hazards regression analysis of transition to psychosis in PACE 400 (n=212).

	Base		Base + Disorders of thought content		Base + Conceptual disorganization		Base + Brain Age Gap		Base + Age		Base + Brain Age Gap + Age	
	HR	P	HR	P	HR	P	HR	P	HR	P	HR	P
Duration of symptoms prior to service entry – log transformed	**1.68**	**0.015***	**1.68**	**0.021***	**1.64**	**0.025***	**1.67**	**0.020***	**1.71**	**0.022***	**1.70**	**0.024***
GAF	**0.51**	**0.001***	**0.53**	**0.002***	**0.52**	**0.001***	**0.51**	**0.001***	**0.49**	**<0.001***	**0.49**	**0.001***
UHR subgroup		**0.017***		0.050		**0.024***		**0.017***		**0.003***		**0.004***
BLIPS vs. Vulnerability	**3.13**		2.56		**3.08**		**3.14**		**3.19**		**3.21**	
Attenuated Psychosis vs. Vulnerability	**1.28**		1.11		**1.35**		**1.28**		**1.13**		**1.13**	
CAARMS - Disorders of thought content			1.37	0.292								
CAARMS – Conceptual disorganization					1.14	0.389						
Brain age gap							1.03	0.879			1.04	0.847
Age									0.74	0.151	0.74	0.180
LR χ2		35.00		36.38		36.07		35.03		37.42		37.47
Fraction of new information		-		0.04		0.03		0.00		0.07		0.07
Concordance index (CI)		0.68		0.68		0.68		0.68		0.69		0.69

For each full model, the likelihood-ratio (LR) test was obtained by comparing log likelihoods of the base and full models. The fraction of new information is the proportion of total predictive information in clinical plus brain age gap that was added by brain age gap. It was calculated as follows: 1 – Base LR χ2/Full LR χ2. HR, Hazard ratio; PC, principal component; LR χ2, likelihood ratio Chi-squared statistics; *significance level for LR test P values: $P < 0.05$

Table 4: Hazard ratio (HR) and P value after internal validation (bootstrapping N = 1,000) for clinical and cognition predictor variables in the multivariate Cox proportional hazards regression analysis of transition to psychosis in PACE 400 (n=127).

	Base		Base + Disorders of thought content		Base + Conceptual disorganization		Base + Digit Symbol Coding		Base + Arithmetic		Base + RAVLT total	
	HR	P	HR	P	HR	P	HR	P	HR	P	HR	P
Duration of symptoms prior to service entry – log transformed	1.47	0.145	1.49	0.153	1.47	0.150	1.44	0.180	1.52	0.134	1.47	0.166
GAF	**0.33**	**0.001***	**0.35**	**0.002***	**0.33**	**0.001***	**0.31**	**0.001***	**0.35**	**0.002***	**0.33**	**0.001***
UHR subgroup		0.712		0.839		0.684		0.677		0.845		0.760
BLIPS vs. Vulnerability	1.63		1.34		1.71		1.72		1.42		1.63	
Attenuated Psychosis vs. Vulnerability	1.24		1.04		1.33		1.26		1.21		1.24	
CAARMS - Disorders of thought content			1.21	0.338								
CAARMS - Conceptual disorganization					1.09	0.716						
Cognition variable							1.165	0.480550	0.61	0.178	1.01	0.977
LR χ2		26.94		27.99		27.21		27.41		29.46		26.94
Fraction of new information		-		0.04		0.01		0.02		0.09		0.00
Concordance index (CI)		0.69		0.69		0.61		0.68		0.69		0.69

For each full model, the likelihood-ratio (LR) test was obtained by comparing log likelihoods of the base and full models. The fraction of new information is the proportion of total predictive information in clinical plus cognition that was added by cognition. It was calculated as follows: 1 – Base LR χ2/Full LR χ2. HR, Hazard ratio; PC, principal component; LR χ2, likelihood ratio Chi-squared statistics; *significance level for LR test P values: $P < 0.05$