Schizophrenia

ARTICLE OPEN

Check for updates

Data analysis strategies for the Accelerating Medicines Partnership® Schizophrenia Program

Nora Penzel[1,2,51], Pablo Polosecki[3,51], Jean Addington[4], Celso Arango[5], Ameneh Asgari-Targhi[6], Tashrif Billah[2], Sylvain Bouix[2,7], Monica E. Calkins[8], Dylan E. Campbell[2], Tyrone D. Cannon[9], Eduardo Castro[3], Kang Ik K. Cho[2], Michael J. Coleman[2], Cheryl M. Corcoran[10], Dominic Dwyer[11,12], Sophia Frangou[10,13], Paolo Fusar-Poli[14,15], Robert J. Glynn[16,17], Anastasia Haidar[2], Michael P. Harms[18], Grace R. Jacobs[2], Joseph Kambeitz[19], Tina Kapur[6], Sinead M. Kelly[2], Nikolaos Koutsouleris[14,20], K. R. Abhinandan[21], Saryet Kucukemiroglu[22], Jun Soo Kwon[23,24], Kathryn E. Lewandowski[25,26], Qingqin S. Li[27], Valentina Mantua[28], Daniel H. Mathalon[29,30], Vijay A. Mittal[31], Spero Nicholas[30,32], Gahan J. Pandina[33], Diana O. Perkins[34], Andrew Potter[35], Abraham Reichenberg[10], Jenna Reinen[3], Michael S. Sand[36], Johanna Seitz-Holland[1,2], Jai L. Shah[37,38], Vairavan Srinivasan[33,39], Agrima Srivastava[10], William S. Stone[40], John Torous[40], Mark G. Vangel[41,42], Jijun Wang[43], Phillip Wolff[44], Beier Yao[25,26], Alan Anticevic[45], Daniel H. Wolf[8], Hao Zhu[35], Carrie E. Bearden[46,47], Patrick D. McGorry[11,12], Barnaby Nelson[11,12], John M. Kane[48,49], Scott W. Woods[45,50], René S. Kahn[10], Martha E. Shenton[1,2,6], Accelerating Medicines Partnership® Schizophrenia (AMP® SCZ)*, Guillermo Cecchi[3,52] and Ofer Pasternak[1,2,6,52] ✉

The Accelerating Medicines Partnership® Schizophrenia (AMP® SCZ) project assesses a large sample of individuals at clinical high-risk for developing psychosis (CHR) and community controls. Subjects are enrolled in 43 sites across 5 continents. The assessments include domains similar to those acquired in previous CHR studies along with novel domains that are collected longitudinally across a period of 2 years. In parallel with the data acquisition, multidisciplinary teams of experts have been working to formulate the data analysis strategy for the AMP SCZ project. Here, we describe the key principles for the data analysis. The primary AMP SCZ analysis aim is to use baseline clinical assessments and multimodal biomarkers to predict clinical endpoints of CHR individuals. These endpoints are defined for the AMP SCZ study as transition to psychosis (i.e., conversion), remission from CHR syndrome, and persistent CHR syndrome (non-conversion/non-remission) obtained at one year and two years after baseline assessment. The secondary aim is to use longitudinal clinical assessments and multimodal biomarkers from all time points to identify clinical trajectories that differentiate subgroups of CHR individuals. The design of the analysis plan is informed by reviewing legacy data and the analytic approaches from similar international CHR studies. In addition, we consider properties of the newly acquired data that are distinct from the available legacy data. Legacy data are used to assist analysis pipeline building, perform benchmark experiments, quantify clinical concepts and to make design decisions meant to overcome the challenges encountered in previous studies. We present the analytic design of the AMP SCZ project, mitigation strategies to address challenges related to the analysis plan, provide rationales for key decisions, and present examples of how the legacy data have been used to support design decisions for the analysis of the multimodal and longitudinal data. Watch Prof. Ofer Pasternak discuss his work and this article: https://vimeo.com/1023394132?share=copy#t=0.

Schizophrenia (2025)11:53 ; https://doi.org/10.1038/s41537-025-00561-w

INTRODUCTION

Individuals at clinical high-risk for psychosis (CHR) (see Addington, J et al.[1], within this special issue for definition of CHR criteria) show variable clinical trajectories, resulting in a variety of short- and long-term outcomes[2]. Based on a recent meta-analysis[3], the rate of CHR individuals who develop psychosis (also termed 'converters') is about 0.15 (95% confidence interval [CI], 0.13–0.15; $n = 3408$) within one year and 0.19 (95% CI, 0.17–0.22, $n = 7351$) within two years of assessment. From the group of those who do not develop psychosis (also termed 'non-converters') about 50% remit, with positive symptoms reducing below CHR threshold[4]. The remaining non-converters continue to experience subthreshold psychotic symptoms[5] along with psychosocial impairments and compromised quality of life[6,7]. Group-based analyses have provided important insights into the patterns and abnormalities within CHR populations compared to control groups, such as community controls. However, actionable and reliable quantitative models are needed for capturing unique variations among individuals at CHR and for predicting clinical outcomes. Such models could offer improved prognosis and treatment planning (see Fig. 1) and are important to improve the design of clinical trials to test interventions aiming to improve symptoms, prevent further deterioration, and potentially prevent psychosis onset[8]. For example, trial efficiency and efficacy can be enhanced by enriching cohorts to include participants who are more likely to develop psychosis. In addition, modeling the various trajectories leading to the clinical outcomes could identify subgroups of CHR individuals with similar pathologies, and would potentially provide a better mechanistic understanding of the transition from CHR to remission, psychosis, and other disorders. In turn, a better mechanistic understanding might facilitate future treatment development[9].

The Accelerating Medicines Partnership® Schizophrenia (AMP® SCZ) study[10,11] is a large project aimed to generate quantitative models for the prediction of outcomes and the identification of clinical subgroups in individuals at CHR (see Fig. 1). This

A full list of author affiliations appears at the end of the paper.

Published in partnership with the Schizophrenia International Research Society

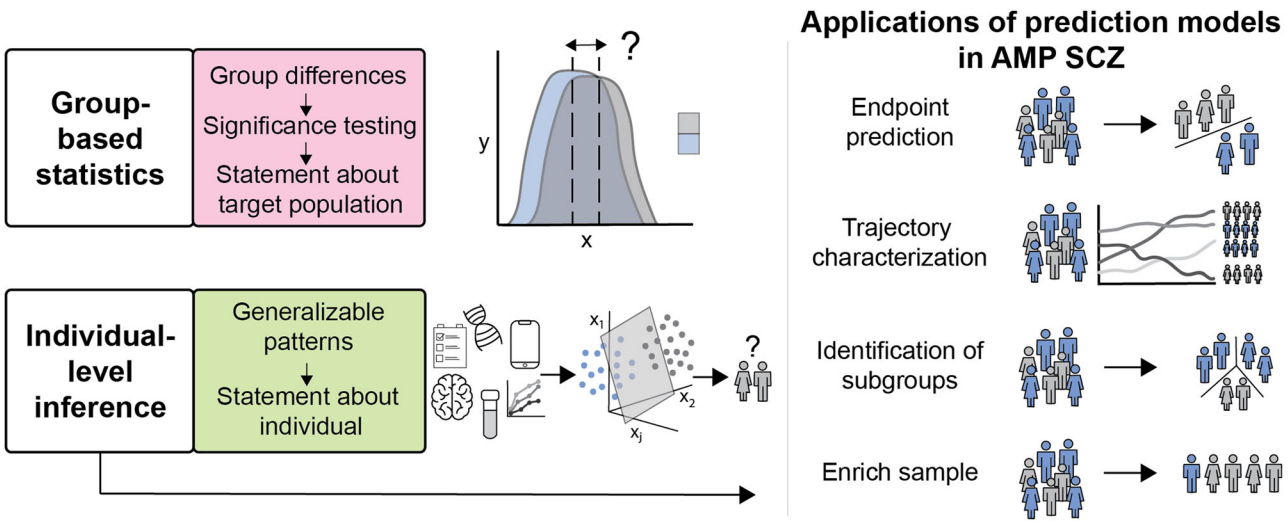

Fig. 1 Potential utility of individual-level inference models for the AMP SCZ study. A central goal of AMP SCZ analyses is the development and validation of quantitative models with straightforward translatability to clinically useful applications. Group-based statistical approaches quantify group differences in observed variables to make statements about the distribution of those variables in a target population[99]. In contrast, models for individual-level inferences combine observations at baseline, or over time, to make statements about previously unseen individuals[48]. Here, several applications of individual-level inference models relevant to the AMP SCZ study are shown.

publication outlines the rationale and general principles guiding the analytic approach for AMP SCZ. We first describe the state-of-the-art of quantitative modeling in CHR research, highlight complexities of the expected AMP SCZ dataset and present the analytic aims of the AMP SCZ study, emphasizing the applications of quantitative modeling. We then lay out the analysis plan, including the main statistical tools to be used, and a validation plan. With examples of two analyses, we demonstrate how legacy data from previous CHR studies informs decisions made for the AMP SCZ project. Specifically, we present a plan to integrate multimodal and longitudinal measurements into our analyses. We also discuss the technical challenges expected in our analyses and present a variety of mitigation strategies. By transparently outlining the analytic design of AMP SCZ and its rationale, we hope to provide a roadmap for other researchers in the scientific community interested in conducting similar studies.

State-of-the-art quantitative models in CHR research

Studies of quantitative models in CHR research have mostly focused on prediction models, forecasting who will develop psychosis[12–14]. However, recognizing that most individuals at CHR will not develop psychosis, more recent studies also aim to predict who will remit and who will develop more broadly defined adverse outcomes including persistent CHR syndrome[8], non-psychotic disorders and functional deficits[15,16]. Prediction in these studies is typically facilitated with survival analyses (e.g., Cox proportional hazard regression[12]) and/or machine-learning-based prediction algorithms (e.g., support vector machine[17], random forest[18,19], regularized regression[20]). Predictive value has been found for sociodemographic data, clinical characteristics, cognitive data, structural and functional neuroimaging, and electrophysiology (see review, ref. [21]), as well as language processing[22] and genetic and fluid biomarkers[17,23–31]. Multimodal models combining several data modalities have outperformed unimodal predictions[17,32,33]. Overall, the predictive power across studies is moderate, with a recent meta-analysis[34] showing limited predictive power with an average of 67% sensitivity and 78% specificity for the prediction of psychosis. Small datasets might contribute to this limited predictive power. In addition, the reported predictive power is likely overestimated, since the suggested models have limited generalizability where studies typically investigate individuals that are from similar geographical catchment areas[35] and are not appropriately validated. Strict validation calls for a replication using external data, while less strict, yet still robust approaches require internal cross-validation approaches[36]. However, many of the reported studies lack these validation approaches, leading to higher risk of overfitting. Indeed, when testing the applicability of models on new data, a sizable decrease in predictive power is almost always observed[14].

Large multi-site studies have addressed the urgent need to increase sample sizes and to broaden catchment areas including different countries from the respective continents to develop more generalizable prediction models with stronger predictive power. Consortia such as the European Network of Gene-Environment Interactions in Schizophrenia (EU-GEI)[37], the North American Prodrome Longitudinal Study (NAPLS)[38], the Shanghai At Risk for Psychosis (SHARP) project[39], and the Personalized Prognostic Tools for Early Psychosis Management study (PRONIA)[15] have significantly advanced the field, while revealing important multi-site related complications that stem from heterogeneities across sites, such as heterogeneity in sociodemographic characteristics and imaging measures.

The Harmonization of At-Risk Multisite Observational Networks for Youth (HARMONY) project was established to enable joint analyses with the ability to validate results across projects[13] and continents. However, differences across consortia, including variable criteria defining CHR status and inconsistent acquisition and processing procedures[13], impede retrospective data harmonization.

One of the main advantages of the AMP SCZ project is the pre-harmonization efforts to reduce heterogeneity of assessment across sites (harmonized protocols can be found in the following Methods papers [Addington, J et al.[1]; Mathalon, D et al.[40]; Harms, M et al.[41]; Allott, K et al.[42]; Perkins, D et al.[43]; Wigman, J et al.[44], Bilgrami, Z et al.[45]]. Nevertheless, and as we show below, we use legacy data from some of the aforementioned studies to inform decisions made for AMP SCZ.

METHODS

In the following section, we describe the data modalities acquired in the AMP SCZ project, highlighting properties of the data that might challenge data analyses. Detailed acquisition protocols and

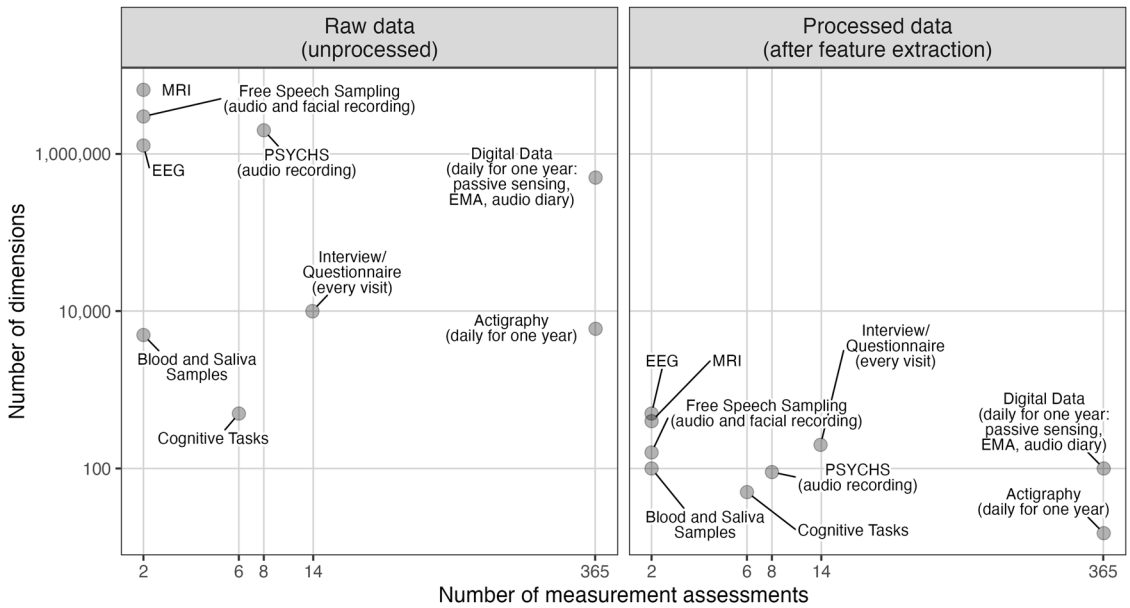

Fig. 2 Feature dimensionality and frequency in the AMP SCZ study. The figure captures the dimensionality and the frequency of the different data domains within the AMP SCZ study for raw data (left) and after preprocessing and feature extraction (right). Measurement frequency and number of dimensions are presented on a logarithmic scale for clarity. Abbreviations: magnetic resonance imaging (MRI), electroencephalography (EEG), ecological momentary assessment (EMA), Positive SYmptoms and Diagnostic Criteria for the CAARMS Harmonized with the SIPS (PSYCHS).

Table 1. Legacy data: sample size and basic demographics of legacy datasets analyzed.

Group	PRONIA		NAPLS-3		SHARP		NYSPI
	CHR	HC	CHR	HC	CHR	HC	CHR
Number of individuals at clinical high-risk for psychosis	173	334	698	96	147	97	102
Number of individuals at clinical high-risk for psychosis who developed a psychotic disorder (% of whole sample)	23 (13%)	–	71 (10%)	–	29 (20%)	–	27 (26%)
Number of females at birth (%)	88 (51%)	185 (55%)	319 (46%)	48 (50%)	72 (49%)	47 (50%)	24 (24%)
Age in years (Mean [Standard deviation])	23.6 (5.1)	25.7 (6.1)	18.7 (4.1)	19.1 (4.2)	18.9 (5.0)	18.6 (4.5)	20.1 (3.8)
Time to conversion in days (Mean [Standard deviation])	237 (36.4)	–	262 (280)	–	395 (342)	–	381 (383)
Latest assessment timepoint after baseline assessment in years/months	~2 years/1 month	–	~4 years/3 months	~1 year/1 month	~5 years/2 months	~4 years	~4 years/6 months
Assessment period	02/2014-06/2017		01/2015-01/2019		08/2012-01/2016		05/2003-12/2012
Number of sites for recruitment	7		9		1		1

Abbreviations: *CHR* clinical high-risk for developing psychosis, *HC* healthy controls, *NAPLS* North American Prodrome Longitudinal Study, *NYSPI* Columbia University/New York State Psychiatric Institute, *PRONIA* Personalized Prognostic Tools for Early Psychosis Management, *SHARP* Shanghai At Risk for Psychosis.

standard operating procedures for each data domain are available in other publications within this special issue and on the AMP SCZ website (https://www.ampscz.org).

AMP SCZ data

Overall, the AMP SCZ study will include 1977 individuals at CHR for psychosis and 640 community controls at baseline, across 43 recruitment sites, spanning 13 countries across 5 continents, while using harmonized CHR inclusion criteria and operating procedures[11]. All participants undergo a comprehensive baseline assessment and major follow-up assessments after 1, 2, 3, 6, 12, 18, and 24 months. In addition, CHR individuals and a part of the control group undergo monthly follow-up assessments during the first year after study inclusion. Each site follows specified recruitment targets that consider balanced recruitment of minority groups, and sociodemographic variables. The protocol consists of interview-based measurements [Addington, J et al.[1], within this special issue], cognitive assessments [Allott, K et al.[42], within this special issue], magnetic resonance imaging (MRI) [Harms, M et al.[41], within this special issue], electroencephalography (EEG) [Mathalon, D et al.[40], within this special issue], blood and saliva samples [Perkins, D et al.[43], within this special issue], audio-visual recordings and speech transcriptions [Bilgrami, Z et al.[45], within this special issue], actigraphy, and smartphone-based assessments including active and passive sensing [Wigman, J et al.[44], within this special issue].

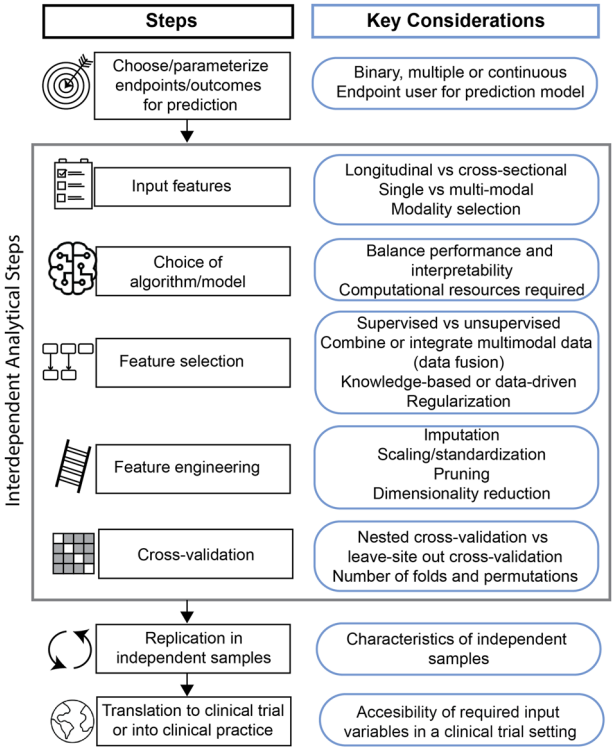

Fig. 3 Key considerations in making an analysis pipeline. A typical analysis pipeline involves several different steps. Each step depends on key considerations and requires optimization of performance that is specific for the study needs.

Some models will rely on minimally processed data (Fig. 2: "Raw data"), whereas others will be built from preselected variables based on a priori knowledge (Fig. 2: "Processed data"). However, as depicted in Fig. 2 there is a high variability in the measurement frequency (timescale) and in the number of features (dimensionality) across both the raw and the processed data domains used in AMP SCZ.

To inform the design of AMP SCZ analyses, while data from AMP SCZ is still being collected, we utilized legacy data with properties similar to the data expected in AMP SCZ. Specifically, the legacy data assists analysis pipeline building, performing benchmark experiments, quantifying clinical concepts, testing new cutting-edge approaches, and making design decisions meant to overcome the challenges encountered in previous studies. The legacy data for the AMP SCZ observational study include data from the PRONIA[15], the NAPLS-3[38] and the SHARP[39] studies, and a study that acquired data from CHR individuals at Columbia University/New York State Psychiatric Institute (NYSPI; Table 1)[46]. A thorough description of these studies and the included samples can be found in their primary publications. All studies were approved at the respective recruiting sites by the Institutional Review Board committees and all participants provided written, informed consent (personally or through a legal guardian if below age of 18 years).

Analysis aims of AMP SCZ

While the overall aims and the general approach of the AMP SCZ study have been described in detail elsewhere[10,11], in this paper we focus particularly on the analytic aims, design, and associated challenges. The primary analytic aim is to use a set of clinical assessments and multimodal biomarkers obtained during the first two months of assessments to predict clinical endpoints of CHR individuals at one and two years (see selection of endpoints below). Achieving the primary analytic aim will enable the development of tools for the selection of enriched patient populations in clinical trials for individuals at CHR for psychosis. The secondary analytic aim is to use the longitudinal clinical assessments and multimodal biomarkers from all time points to characterize trajectories of CHR individuals towards different clinical outcomes. Upon achieving the secondary analytic aim, we will be able to identify trajectories of clinical outcome measures, digital measures, and potential biomarkers that are associated with the endpoints so they can inform future clinical trials, help to construct risk calculators, act as intermediate endpoints in shorter study designs and provide some mechanistic understanding as to how individuals evolve to reach a particular clinical endpoint. Importantly, the developed models will focus on single subject-level predictions rather than inferring information at the population/group level (see Fig. 1).

Technical considerations and challenges

As part of the analytic design for the AMP SCZ study, we outlined several technical considerations and challenges based on experience from previous studies, and on some unique aspects of the AMP SCZ study. See Fig. 3 for an overview of the different steps involved in the analytic planning and some of the main decisions that were required. In Table 2 we summarize some of the most important challenges identified and present different possible mitigation strategies along with their advantages and disadvantages. In the remainder of this paper, we focus on the following main challenges towards accomplishing our analytic aims:

(1) The need to define and to quantify endpoints that go beyond conversion to psychosis and are clinically meaningful for the CHR population. Several clinical endpoints could be of importance for individuals at CHR, such as the remission from CHR status, poor global functioning, or cognitive deficits. While the definitions of these endpoints are important and widely recognized, they vary across studies and their stability over longer time-courses is not yet well defined.
(2) The methods to be developed need to be generalizable. This means that appropriate validation procedures need to be in place such that when the developed models are applied to new data, they should maintain the accuracy obtained using the original training data, i.e., AMP SCZ data.
(3) The high number of assessments produces a large number of features with different relevance that require feature engineering for dimensionality reduction.
(4) The methods to be developed should accommodate fusing multiple modalities. Previous studies have already highlighted that multimodal algorithms perform better than unimodal algorithms in predicting transition to psychosis and global functioning[15,17,32,33]. Employing multimodal models for the large AMP SCZ assessment battery is promising since AMP SCZ includes a broader number of domains than most previous studies. However, the large variety among domains (e.g., in the number of features per domain, see Fig. 2) demands appropriate data fusion techniques.
(5) The methods to be developed should accommodate longitudinal measures to leverage the rich longitudinal measurements in AMP SCZ. This is complicated by the fact that the different domains have a different frequency of sampling (Fig. 2). Longitudinal information collected within a short period of time close to baseline may aid prediction models. Also, including the complete longitudinal data may help to characterize clinical trajectories and their relationship with clinical endpoints, which could provide enhanced mechanistic understanding, as well as establish a naturalistic baseline which can then inform change with treatment in clinical trials.

Table 2. Technical considerations and challenges.

Examples of challenges	Mitigation strategies	Additional information; Advantages (+) and Disadvantages (−)
AMP-SCZ acquires several different data domains that might be combined for predictive modeling	Stacking or meta-learning (see e.g., Koutsouleris et al.[17])	Late fusion techniques which combines unimodal outputs using a weighted voting scheme + simple + accommodates missing modalities + allows step-wise combination useful for clinical implementation − doesn't capture relationships between modalities
	Multiple kernel learning[33,49,50]	See example in section "Design of multimodal approaches" + simple and interpretable + allows modalities to inform each other − less powerful than neural networks
	Early fusion[12,67]	Concatenates matrices of different domains before model building + simple − works for preselected features, difficult to balance different modalities
	Neural Network	Learns the relevant features by successive application of non-linear transformations + flexibility − complex implementation and training − low model interpretability − requires very large training data
AMP-SCZ study combines measures with several (different) time points	Joint-modeling[71–73]	Integrates change of measurements via linear mixed-effects models into survival analysis to predict outcome + integrates modeling of longitudinal measures − relatively simplified approach that mainly measures linear relationships − difficulties with high numbers of longitudinal variables
	Dynamic time warping[79–82]	See example in section "Design of longitudinal approaches" + accommodates different sampling frequencies + accommodates trajectories of different lengths + accommodates multiple longitudinal variables − not predictive in itself
	Mere combining (concatenating) data from different time points	Include mere baseline and longitudinal data in the feature space + very simple approach − loss of temporal order information
	Latent trajectory analysis (e.g., latent growth curve modeling[2], Bayesian nonparametric trajectory mixture modeling approach[75])	Identifies underlying patterns or latent classes within longitudinal data to uncover distinct trajectories or patterns of change over time + modeling longitudinal trajectories as endpoints might provide more stable and clinically relevant outcomes − large sample sizes and sufficient number of time points needed for reliable estimates
The high number of assessments produces many features with different relevance	Model Regularization[65]	Several algorithms integrate feature selection mechanisms within the learning phase of the model + supervised selection of features that might reflect associations between features − there are several regularization approaches, each with their own assumptions.
	Dimensionality reduction such as independent component analysis[62], principal component analysis[61], non-negative matrix factorization[63]	Project data into a lower-dimensional space, capturing essential components + enables simpler prediction models + eliminates redundancies − unsupervised; might discard predictive information − might sacrifice interpretability
	Uniform manifold approximation and projection[95]	

Table 2 continued

Examples of challenges	Mitigation strategies	Additional information; Advantages (+) and Disadvantages (−)
	Hypothesis-driven feature selection[12]	Researchers preselect features that should be included in the models + might reduce the number of spurious findings + increase interpretability − high researcher degree of freedom − previously unexplored features will not be included
	Wrapper-based feature selection[66]	Integration of feature selection in the model building process. Features are added to the feature space stepwise and evaluated based on their predictive performance until either a specific number of features has been tested or stop-criterion defined by the researcher has been fulfilled. + has proven high predictive accuracies in previous studies − high computational demand − might easily lead to overfitting
Expected missing data in several data modalities	Prediction algorithms that allow for missing data in their input (e.g., random forest)	Some algorithms treat missing data as informative data itself + no additional imputation required + missingness itself might be informative for outcome predictions − several algorithms that require complete data cannot be used (e.g., support vector machine)
	Imputation methods, e.g., mean/median imputation, k-Nearest Neighbor imputation[96], Multiple Imputation by Chained Equations (MICE)[97]	Missing values are replaced with best estimates based on different methods + increases sample size + enables the use of algorithms not adapted for missing data − assumes a model for imputation that may bias the prediction
Variables might be confounded with the outcomes of interest, e.g., age, sex, race, ethnicity, socio-economic status, intelligence quotient	Including potential confounds as features	As several of the potentially confounding variables might carry predictive power for the outcome, they can directly be integrated in the models + if appropriately testing their predictive power, we can learn about their contribution to the outcome prediction − might reflect cohort specific associations − may require a control sample
	Regressing out their effect	Statistically removing the influence of confounding variables + applicable to all domains − potential non-linear effects will not be captured − might also remove effects relevant for the disorder, e.g., if males are more likely to transition to psychosis
	Training models on subgroups of individuals	Segmenting the dataset based on specific characteristics or criteria + captures subgroup specific association − sample sizes might become too small for complex modeling
	Propensity scores and confound-isolating cross-validation	Systematically leaving out specific subgroups + allows quantification of performance without confounds − equivalent to discarding certain samples
AMP-SCZ is a world-wide study and models should perform across countries and diverse clinical/research facilities (generalizability)	Split sample based on a predefined cut-off in discovery and replication sample	See section "Selection of a validation plan" for description of validation in AMP-SCZ study + unbiased test of model performance − reduction in sample size, which is especially relevant for model training
	Leave-site out cross validation[48]	See section "Selection of a validation plan" for description of validation in AMP-SCZ study + training across different sites and direct testing whether model performs in left-out site − might lead to high variance of estimated model performance due to site-specific characteristics

Table 2 continued

Examples of challenges	Mitigation strategies	Additional information; Advantages (+) and Disadvantages (−)
	Pre-harmonization by standardization of operating procedures such as protocols for MRI or clinical assessments[52]	Protocols are harmonized across sites + better comparability of the data across sites + assuring that differences across sites are related with the catchment area and not based on protocol differences − translation of findings to cohorts with deviating protocols is more difficult
	Post-harmonization using statistical tools (e.g., harmonization of diffusion weighted imaging data[98], mean-centering of clinical variables[13])	Statistically harmonize the data + modeling differences across sites allows model application in diverse settings relying on post-harmonization procedures + enhances integration of consortium-level data − most approaches require adequate sample sizes of controls to adjust for disorder-specific effects

RESULTS

Analysis design for AMP SCZ

To achieve its analysis aims, AMP SCZ is intending to use quantitative models based on state-of-the-art machine learning (ML) approaches for individual-level inferences[47] and risk monitoring. Specifically, we are selecting ML methods that address some of the aforementioned challenges, following three principles: (1) Robustness—includes reliable cross-validation and independent replication[48]; (2) Flexibility—capable of handling diverse features from different domains that may be independent or non-linearly related, including missing and unmatched entries[49,50]; and (3) Methods that are clinically informed—that can be guided by informative features preselected by experts, and the predictions can be mapped back to individual features[12]. Importantly, we prioritize tools that will balance optimizing accuracies with providing interpretable models (instead of "black box" analyses). Therefore, we will make use of explainable artificial intelligence[51] that identifies which features and potential biomarkers contribute to the individual-level inferences.

This section describes the design decisions taken for AMP SCZ along with examples of benchmark experiments performed on legacy data supporting the decisions.

Selection and quantification of clinical endpoints. The primary endpoint of the AMP SCZ study is the development of a psychotic disorder (conversion), assessed by the Positive SYmptoms and Diagnostic Criteria for the CAARMS Harmonized with the SIPS (PSYCHS)[52] within 12 and 24 months after initial assessment. Secondary endpoints of AMP SCZ are sustained remission from the CHR syndrome and persistent CHR syndrome (non-conversion/non-remission). Sustained remission is defined as the absence of CHR criteria for at least six months and until the last available follow-up[4]. Persistent CHR syndrome is defined as individuals with absence of conversion or sustained remission at 12 and 24 months after the initial assessment. These endpoints were selected from several other clinical outcomes that could serve as endpoints (see [Addington, J et al.[1], within this special issue] for detailed description of these other clinical outcomes). We selected the primary and secondary endpoints based on their relevance for the classification of the CHR population and robust quantification. The definition of conversion is robust and reliable across studies, whereas the criteria for the definition of remission are more heterogenous across studies, with variable definitions of the type of remission (clinical or clinical and functional) and the period required for the definition of sustained remission. The definition of remission for AMP SCZ was therefore based on comparing the rate and stability of various remission definitions in legacy data (NAPLS-3).

Selection of a validation plan. *Generalizability* is crucial to assure robustness of quantitative models applicable for drug development, clinical research, and clinical practice. *Generalizable* models will maintain their performance when applied to individuals who have not been involved in the model building phase, like those from new studies or clinical trial populations[48]. Achieving generalizability requires appropriate validation. Diversity in the training sample is also important to assure that the model will perform well on diverse ethnic and racial groups, and geographical locations. Through its large size and international coverage, the AMP SCZ study uniquely offers the opportunity to implement an extensive validation plan—an unmet need of most preceding studies.

The AMP SCZ model validation plan involves several layers. The ultimate test of model robustness will be performed on an independent sample of $N = 200$ individuals at CHR entirely excluded from the model building step. The size of this leave-out sample was determined using a sample size analysis that replaces standard power analyses, which cannot be directly applied to prediction model problems. Here, we set the class imbalance ratio (fraction of converters) and a desired performance level using area under the ROC curve (AUC) as a metric of model performance (analogous to an effect size). By varying the sample size, we found the smallest sample size needed for a statistically significant AUC at a selected significance level (p-value). The significance level is calculated for rejecting the null hypothesis H_0: AUC = 0.5, i.e., that the AUC of the predictor is at chance level[53,54]. We considered several combinations of AUC levels and imbalance ratios (Fig. 4), and focused on a combination of a conversion rate of 0.1 and AUC of 0.7, which reflects worse than average results reported in the literature[55]. The sample size analysis determined that for these values a held-out set of $N = 100$ would be statistically significant at the $p = 0.05$ level, and $N = 200$ (see Fig. 4) would be statistically significant with $p = 0.003$. This sample size analysis result matches that of a standard power analysis for AUC[56] yielding a required sample of $N = 202$ for AUC of 0.75, power of 0.8, and 0.003 significance level. To allow room for redundancy and to account for possible attrition within the held out sample we decided on $N = 200$, where any predictive models with AUC better than 0.7 or higher ratio of converters than 10% would be statistically significant at more stringent significance thresholds (Fig. 4). Accordingly, in our study, an AUC of 0.7 is

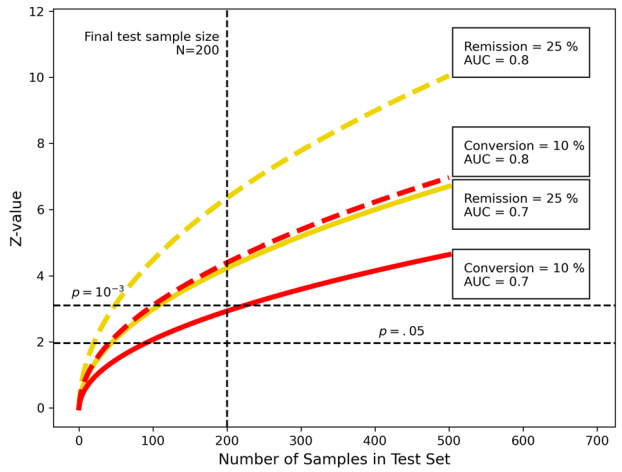

Fig. 4 Sample size analysis for predictive models. The curves in the figure show the dependency of significance on sample size for 4 model scenarios differing by AUC value and conversion rate. Significance, quantified by a z-score, represents the test with a null hypothesis that the AUC is the same as chance level AUC (0.5). The associated significant levels of $p = 0.05$ and $p = 10^{-3}$ are marked. The model scenarios are characterized by a minimal AUC (analogous to the effect size in traditional power calculations) and by a minimal conversion rate. For example, the red line describes a model with an AUC of 0.7, and conversion rate of 10%.

designated as the minimal goal or acceptance criteria for model performance. With a randomly left-out sample there is always the chance that it does not represent well the heterogeneities of the sample from which it was drawn. To mitigate this chance, the left-out sample will be across all sites, maintaining a distribution of positive symptoms similar to the expected distribution across the entire sample, and includes samples collected throughout the entire study term. Note that the left-out sample will not be shared via the NIMH Data Archives (NDA) until the analysis stage of the study is completed.

Prior to external validation, we will apply internal validation approaches during the model building steps. For hyperparameter optimization, when the number of hyperparameters is low relative to training sample size, we plan to use repeated nested *k*-fold cross-validation to estimate their optimal combination without information leakage between training and testing phases[17,48]. Instead of randomly dividing individuals into folds, a systematic cross-validation design will inspect the transportability of learned models by separating individuals into different folds based on specified characteristics of interest. Previous studies have, for example, explored whether models generalize across sites[17,57], i.e., whether models trained on some sites generalize to a completely held-out site, which represents an important step towards potential model implementation into clinical practice. Similarly, in AMP SCZ, we plan folds that include testing for model transferability across ethnic or racial groups, sex and gender, and countries. We also plan to use bootstrapping techniques to estimate confidence intervals of model performance in the population sample, and for finding patterns in the data[58].

Feature engineering and selection for unimodal models. Feature selection and other dimensionality reduction strategies of the input data are important to simplify the analysis tasks. We plan to use a combination of different strategies to reduce dimensionality at varying analysis stages. One such strategy is to select hypothesis-driven features based on a priori expert knowledge[12]. Features of interest in the AMP SCZ study have been identified by dedicated domain specific AMP SCZ workgroups and many can be found in the domain specific manuscripts in this special issue. Deriving composite scores is another domain specific approach, where based on previous studies and expert knowledge we will combine several scores into meaningful overall scores. Examples include composite scores for interview-based features (see [Addington, J et al.[1], within this special issue]), and for cognitive measures (see [Allott, K et al.[42] within this special issue]). In neuroimaging, hypothesis-driven feature selection includes extracting information across anatomical regions of interest, for example gray matter regions[59] or specific white matter tracts[60]. Additional data-driven analytical approaches for dimensionality reduction such as principal component analysis (PCA)[61], independent component analysis[62], non-negative matrix factorization[63] and multidimensional scaling[64] aim to represent the original data in a new feature space, thereby uncovering the underlying data structure. Several of the ML algorithms we plan to use integrate a feature selection mechanism within their model learning phase. For example, the elastic net[65] is a stable regularization method that reduces dimensionality by encouraging sparsity as in the LASSO regularization. Another example we plan to use during the model optimization stage, when model complexity allows it, is wrapper-based feature selection[66], where the optimization is performed while individual features are added or removed consecutively to the feature space. In each step, the predictive performance of the model is evaluated and only features that improve model performance are included, until a pre-set number of features or other pre-set stop-criteria have been fulfilled.

Design of multimodal approaches. Multimodal prediction models combine information obtained from different data sources which may be complementary or display higher-order relationships. In previous studies, multimodal prediction models have outperformed unimodal prediction models in forecasting long-term outcomes in individuals at CHR[17,32,33]. However, technical difficulties limit the construction of multimodal models, thus, most studies still focus on a single data domain or on a small number of domains[34]. Several fusion approaches exist and can roughly be categorized by when in the analysis stream (early, intermediate, late) data is integrated or fused across modalities (see Table 2 for a non-exhaustive overview). *Early fusion* is a simple fusion approach that concatenates features from all modalities and uses them as input for a multivariate model. The approach is effective for small feature sets from a limited number of modalities[12,67] but less so for varying dimensionalities across modalities as expected in AMP SCZ (Fig. 2). In the case of varying dimensionalities, modalities with many features will contribute more than those with a small number of features. *Late fusion* trains unimodal models for each modality and then combines their output using a weighted voting scheme. This approach is effective for correlated modalities, in which case averaging across modalities can reduce noise and improve accuracy[15,17]. However, this approach is less effective when modalities contain complementary information (i.e., higher-order relationships) with synergies (and redundancies) that can be exploited during learning, which late fusion ignores. Finally, *intermediate fusion* approaches apply unimodal layers on each modality to represent it with a small and set number of composite variables (similar to late fusion). Then, it combines the composite variables across modalities (similar to early fusion)[49,68]. During learning, signals from all modalities inform each other even at the first unimodal layer. Given the large number of domains in AMP SCZ, the large variability in dimensionality across domains, and the expected complementary aspects across domains, intermediate fusion was deemed most suitable.

AMP SCZ plans to utilize multiple kernel learning (MKL), which is an intermediate fusion approach, as the main data fusion approach, drawing from our experience using MKL in previous psychosis prediction tasks[33], and on legacy data.

MKL extends other popular kernel-based classification algorithms such as support vector machines (SVMs)[69]. Kernels define a

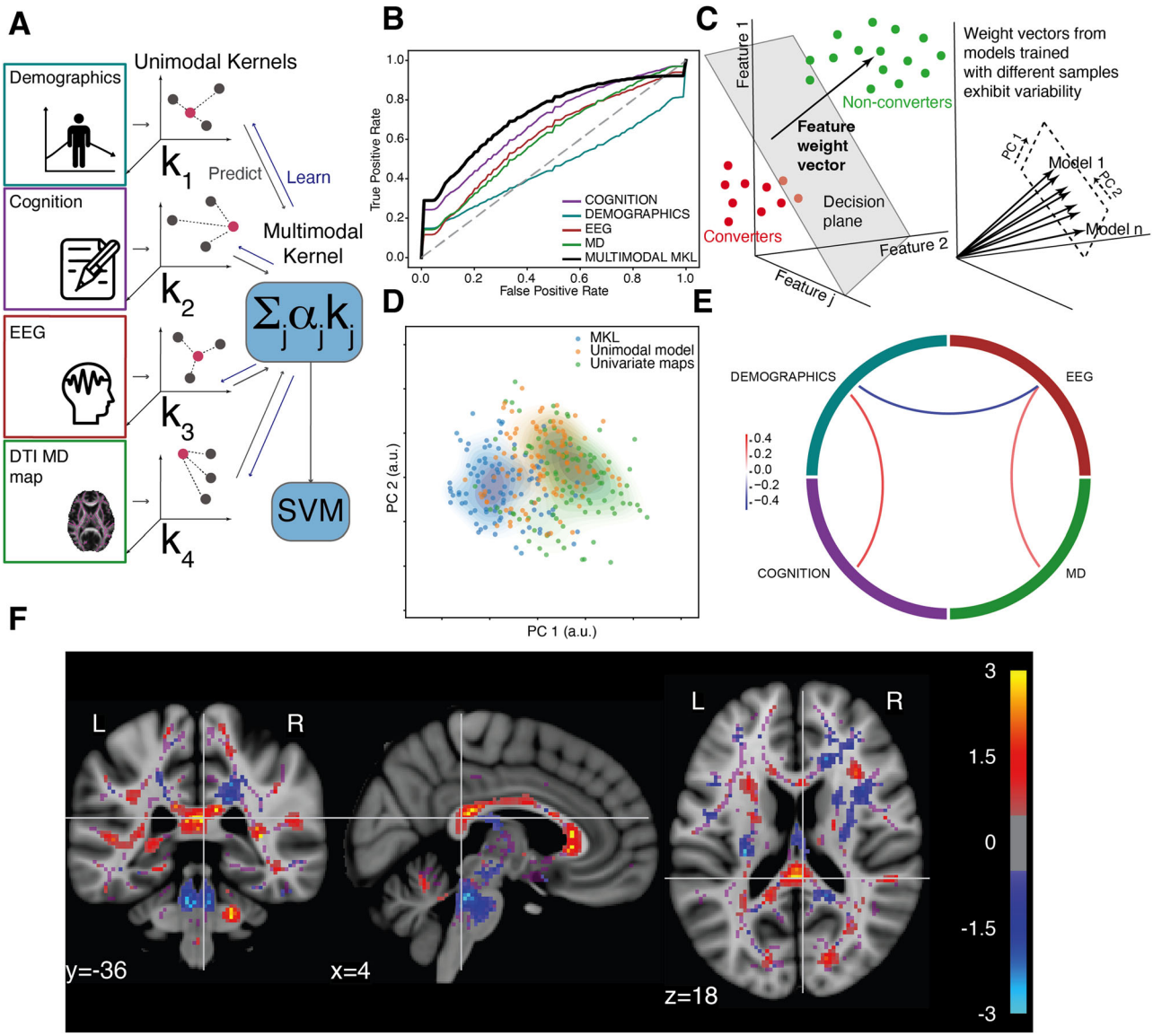

Fig. 5 Multimodal fusion of four modalities to predict development of psychosis. A We used multiple kernel learning (MKL) to combine signals from four modalities, to predict development of psychosis from baseline data in individuals at CHR from the SHARP study. Our analysis fused demographic information with EEG, cognitive scores, and mean diffusivity maps obtained from diffusion weighted imaging (DWI-MD). **B** Using 25 repetitions of 4-fold validation we observed an advantage of MKL (AUC = 0.73) over individual modalities (logistic regression with elastic net, AUC = 0.69 for the best modality, which was Cognition). **C** Linear MKL provides a characteristic feature weight vector for each modality that is used to project individual samples. Stability is quantified by the variability in the feature weight vector across 100 training folds. **D** Weight vectors for DWI-MD from MKL show more stability (vary less) and additional information (less overlap) compared with weight vectors from DWI-MD unimodal models, and DWI-MD t-statistics vectors derived from univariate group comparisons. **E** Visualization of the correlation between prediction signals from different modalities (within converters). **F** Visualization of DWI-MD weight feature vectors derived from MKL provides interpretability (e.g., locations on the brain). Units are z-scores across training folds.

similarity score for any pair of samples. Once a kernel is defined, a similarity matrix is calculated for all pairs of samples and is then used to perform classification. For example, SVM uses the similarity matrix to identify support vectors that define a separation plane between two classes. MKL extends SVM by learning a multimodal kernel as a weighted sum of unimodal kernels[49,50]. The use of a similarity score to perform classification in the multimodal kernel approach ensures no bias towards any modality (even in the presence of differing dimensionalities) while incorporating information from all modalities during model training. Here we present, by way of example, results from the application of MKL on the SHARP dataset (Table 1) to predict conversion to psychosis (Fig. 5).

We tested MKL by fusing four modalities from the SHARP data ($N = 69$, converters: $N = 14$, 20%) and compared the results with unimodal predictions. First, we ran unimodal models based on baseline data and selected demographic information (age, sex, education, and marital status) along with 3 modalities showing at least moderate (AUC > 0.6) predictive power for predicting conversion: cognition, skeletonized mean diffusivity maps from diffusion weighted MRI (mean diffusivity, DWI-MD) and EEG (Oddball task event-related potential, ERP) (Fig. 5A). Comparing the AUC using 25 repetitions of 4-fold validation, MKL (AUC 0.73) outperformed the best unimodal model (Cognitive scores; logistic regression with elastic net, AUC 0.69; Fig. 5B). We further investigated stability and information content. MKL provides a characteristic feature weight vector for each modality, indicating the importance/contribution of each feature. We quantified stability by the variability in the feature weight vector across 100

training folds (Fig. 5C). Using DWI-MD as an example, we compared the variability of the weight vectors obtained for MKL with those obtained for the unimodal model and with the variability of a t-statistics vector obtained by performing a univariate group comparison for each feature. We used the two principal components obtained from PCA to visualize the variability in 2D. MKL had less variability compared with unimodal models and with t-statistics, indicating that the inclusion of more modalities stabilizes the fit. Also, the unimodal variability and the t-statistics variability were overlapping, while the MKL variability was less overlapping, indicating that MKL injects additional information to learned patterns. MKL also allows us to study the redundancy of predictive information across modalities by computing the correlation across samples between the projection to feature maps of each modality (i.e., how correlated the decision signal picked from each modality is in test samples). Focusing only on individuals who developed psychosis, we observed a positive correlation between predictive signals coming from the EEG and DWI-MD measures, as well as between cognitive scores and demographics (see red lines in Fig. 5E), whereas the predictive signals of EEG and demographics were anticorrelated. Another advantage of the MKL approach is that the prediction patterns themselves can be interpreted. For instance, in DWI-MD, the prediction is driven by increased MD around the corpus callosum, especially in the forceps major (Fig. 5F), consistent with previous findings in CHR studies[70]. Together, our proof-of-concept analysis showcases how MKL can increase predictive power while maintaining robustness (internal cross-validation), flexibility (accommodates any kind of modality) and interpretability (provides unique insights about predictive signals).

Design of longitudinal approaches. As outlined in the "Analysis aims of AMP SCZ" section, longitudinal analyses serve two main purposes, leveraging domains with varying timescales throughout the two-year study protocol. First, the prediction of endpoints by analyzing data from the first two months of assessments longitudinally. Second, longitudinal analysis of all time points to identify subgroups based on characterizing trajectories of CHR individuals and their association with the endpoints.

Recent important efforts to accommodate longitudinal data in psychosis prediction rely on joint modeling[71–73]. This technique extends Cox regression survival analyses on baseline data to include additional variables derived from time-varying predictors using linear mixed effect modeling. However, current implementations of joint modeling are limited regarding the amount of longitudinal information that can be added to the model, requiring pre-selection of the most useful longitudinal parameters. This complicates cross-validation approaches, and limits generalizability. As a result, inconsistent findings are reported when applied on CHR data[71–74], and in the largest study joint modeling did not improve prediction accuracy[74].

The identification of longitudinal subtypes typically relies on various clustering algorithms, aiming to uncover latent homogeneous subgroups. Identification of such subgroups holds potential for informing drug development and for enrollment selection in future clinical trials. A key challenge lies in finding clustering approaches adept at capturing information embedded in longitudinal clinical trajectories with multiple assessments at varying intervals. Latent growth models and variations of it[2], and Bayesian Nonparametric Model for Disease Subtyping[75,76] are examples of such approaches.

The AMP SCZ team identified two main challenges regarding the inclusion of longitudinal data: integrating longitudinal data with baseline data in prediction models and combining signals from asynchronously measured domains (e.g., visit-based assessments and phone-based reports). More broadly, synchronization becomes a concern when participants are sampled longitudinally at different stages of illness development, thereby impacting

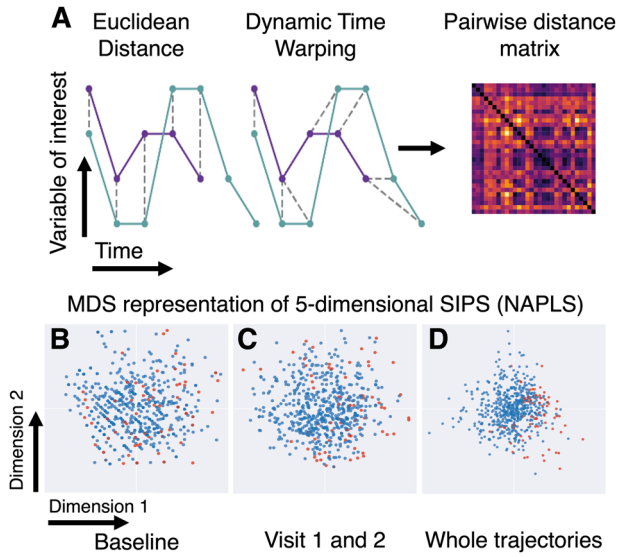

Fig. 6 Longitudinal analyses. An example using dynamic time warping in the NAPLS-3 dataset. **A** Dynamic time warping (DTW) is a method that quantifies the dissimilarity (distance) of pairs of time series. Here, we provide a schematic diagram about DTW: Its key feature is that it is sensitive to the overall shape of trajectories, accommodating temporal delays between them or differences in number of time points. It uses a "stretching" and "compressing" mechanism to match corresponding time points across time series. Given a list of time courses, DTW produces a matrix of pairwise distances. **B** Multidimensional scaling (MDS) representation, a dimensionality reduction method based on preservation of pairwise distances, for individuals at CHR (blue: non-converters within years, red: converters) of SIPS-based positive symptoms at the baseline visit in NAPLS. **C** MDS representations of SIPS at baseline and month 2 visits in the same individuals, showing nascent differentiation between the groups, with converters away from the center. **D** MDS representation of whole SIPS trajectories using DTW shows most pronounced grouping of converters.

techniques that are not designed to accommodate such temporal differences.

Our team identified several techniques to address the synchronization issue, including dynamic time warping (DTW), cross-correlation analyses[77], and hidden Markov models[78]. Below we focus on DTW[79–82] since it seamlessly extends the MKL approach to longitudinal analyses. DTW measures the distance between two time series samples (uni- or multivariate) based on the shape of trajectories, accounting for time delays and assessment frequency differences through a "stretching" and "compressing" mechanism (Fig. 6A). The matrix of pairwise DTW distances across subjects/samples can then be included in prediction algorithms, or in clustering approaches to identify trajectory-based subtypes.

To assess the effectiveness of DTW, we applied it to the five Structured Interview for Prodromal Syndromes (SIPS)-based positive symptoms of all NAPLS-3 CHR participants ($N = 621$) with at least two visits, comparing it with baseline data. Using DTW to incorporate longitudinal information from the baseline and 2-months follow-up, we generated a pairwise similarity matrix, followed by multidimensional scaling (MDS) to visualize underlying structures and Kolmogorov-Smirnov test to compare individuals who develop psychosis and those who did not. While SIPS symptom distribution differentiated groups significantly at baseline ($D = 0.26$, $p < 0.001$), inclusion of follow-up data via DTW enhanced this distinction ($D = 0.37$, $p < 0.001$) (Fig. 6C versus Fig. 6B). In addition, applying DTW and MDS to the entire longitudinal trajectories of the 5 SIPS positive symptom dimensions, excluding conversion or post-conversion time-points, which is possible since

DTW can accommodate samples with different lengths of follow-up assessments, further improved the separability of the two groups ($D = 0.39$, $p < 0.001$) (Fig. 6D versus Fig. 6C). These results underscore the value of longitudinal information, even at two months from baseline, in predicting psychosis development.

This benchmark analysis demonstrates the value of DTW in generating embeddings from longitudinal signals and is the basis for the selection of DTW as a main tool in the AMP SCZ longitudinal analysis. We plan to extend DTW to all longitudinal domains and use the DTW embeddings to create kernels that could be included in methods like MKL (Fig. 5) to combine longitudinal and baseline data.

In addition, we will consider state space models to integrate multimodal longitudinal information. For example, the Kalman Filter (KF)[83] dynamically estimates system states to, for instance, estimate whether or not an individual belongs to a particular family of trajectories, by optimally combining noisy measurements iteratively over time with predictions based on a model of the system's dynamics. KF was historically used in the Apollo space program by National Aeronautics and Space Administration (NASA) and in Global Positioning System (GPS) systems, but has also been used in clinical contexts to estimate clinical trajectories[84–86].

DISCUSSION

AMP SCZ, with its ongoing recruitment, has grown to be the largest and most diverse dataset in CHR research, providing unparalleled depth of information and countless analysis prospects. The analysis plan for AMP SCZ described here aims to facilitate the unique opportunity of this observational study to generate a platform that could improve quantitative modeling, facilitate clinical trials, and potentially improve the clinical outcomes of affected individuals.

An important aspect in the design of the analysis plan is the inclusion of multidisciplinary experts from academia, people with lived experiences, industry, drug companies, funding agencies, and other stakeholders. These collaborative efforts have been directing the analysis plan towards a need to establish a transformative impact on individual treatment paths through clinical trials. As a result, the analytical approaches selected for the AMP SCZ observational study establish a foundation for future clinical trials in several important ways: (1) Predictors developed through AMP SCZ will aid in population enrichment for clinical high-risk studies[87]. This would be particularly beneficial for trials aimed at addressing the onset of psychosis. Given the comparatively low conversion rate, conducting a clinical trial would necessitate a substantial number of CHR participants. However, population enrichment strategies, such as pre-screening subjects and excluding individuals more likely to remit, could elevate the conversion rate among the remaining subjects, thereby reducing the required number of study participants. (2) Once developed, the multimodal predictive models and associated risk calculators can be disseminated for validation and application across various clinical populations[88,89]. (3) By characterizing the trajectories of individuals at CHR and exploring their association with endpoints, we can pinpoint outcome measures more likely to be observed within the shorter timeframe of a clinical trial[5,8,11]. For example, several relevant indicators could be considered, based on the findings within AMP SCZ, including additional outcomes such as persistent cognitive or functional impairments, as well as persistent positive/negative symptoms. (4) Identifying CHR subgroups based on their assessment trajectories over time[2,75] can enhance our mechanistic understanding, and may identify individuals more likely to benefit from specific treatments. (5) Incorporating longitudinal data into prediction models will enable dynamic modeling approaches[90,91] that could inform clinical decisions at each evaluation point in future clinical trials. (6) Selecting the MKL approach as our primary prediction algorithm will allow us to explore the association between domains and the endpoints. This facilitates the identification of key measures crucial for prediction, while also identifying redundancy between measures. Such insights are invaluable for designing a more cost-effective assessment protocol for future trials. Exemplifying this, we show insightful associations across the different domains, included within the model, such as the correlation between predictive signal from EEG or DWI-MD. While derived from a relatively small sample, these findings can illustrate how multimodal modeling that integrates complementary data types can improve predictive accuracies, thus improving the probability to identify those at highest risk for developing poor mental health outcomes early, and further, how these models can provide mechanistic insight.

Another important aspect is the addition of novel smartphone and actigraphy-based assessments. Incorporating these newly available assessments into AMP SCZ required developing new data aggregation and analysis pipelines [Wigman, J et al.[44], within this special issue]. These assessments operate on different timescales compared to traditional measures, necessitating adjustments in our analysis plans. While the current literature lacks approaches for the integration of such data, the importance of dynamic predictions is acknowledged in the field[92]. To address this, we introduced dynamic time warping to handle the diverse assessment time points of different modalities, demonstrating its potential to enhance predictive performance. However, the field of digital biomarkers, including smartphone-based assessments, natural language processing and audio-video material, is rapidly evolving. We anticipate that through collaborative efforts within AMP SCZ we will develop additional novel analysis approaches for generating informative summary scores and predictors. We believe that our developments will benefit the wider community by establishing analysis pipelines for these innovative measures. In addition, we remain adaptable to incorporate other emerging analyses in this dynamic field. We recognize the significance of these new measures in tracking dynamic changes in symptomatology, well-being, and behavior in real-time. This is especially relevant in the context of planned future clinical trials, where such real-time tracking can offer valuable insight into treatment effects, including multiple treatment trajectories/outcomes, and adverse events.

The algorithms developed for AMP SCZ will utilize features from a wide array of assessment tools. For optimal application of these algorithms in future data, consistency in assessment tools is crucial. However, once prediction algorithms are established and trajectories are characterized, we aim to assess the contribution of different assessment tools to identify redundancy and calculate the cost-benefit tradeoff of including or excluding certain assessments. Such analysis is crucial not only for alleviating the burden on participants caused by intensive assesments[93], which could improve attrition rates[94], but also for making the AMP SCZ protocol more accessible to the broader community. For instance, by prioritizing more accessible measures to clinicians in the community or in lower-income countries, we can optimize the inclusion of assessments, enhancing accessibility and making them more applicable and beneficial to regions with limited resources.

A particular strength of the AMP SCZ project lies in its worldwide participant recruitment, encompassing various cultures, ethnicities, healthcare systems, and environments, with representation of minority groups and different socio-economic backgrounds. This recruitment strategy will contribute to a broader transferability of the newly developed predictors. In addition, models will be assessed for potential biases, and for their potential transferability across key subgroups, e.g., across sex, race, or ethnicity. By integrating these important aspects, AMP SCZ hopefully contributes not only to the development of models

able to forecast important mental health outcomes in individuals at clinical high-risk for psychosis but which also are appliable to a wider population, fostering equity in mental health research and care.

In summary, this paper outlines the analytic aims of AMP SCZ, its core analysis principles, and addresses several key challenges. These analyses are set to establish an important infrastructure for future studies involving individuals at CHR and will support the initiation of clinical trials. Many additional analyses, supporting the primary and secondary aims, or capitalizing on the comprehensive data of the AMP SCZ study, are planned among various investigators within the AMP SCZ consortium. Furthermore, it is important to note that the data collected in the AMP SCZ study is made available through the NDA, and the developed algorithms will be made available through the AMP SCZ website. Consequently, we anticipate that the broader research community will utilize this unique contribution to implement novel analyses, thus further advancing research and potentially lead to transformative impact on individual treatment paths.

DATA AVAILABILITY

The data collected for the AMP SCZ study are available via scheduled releases at the NIMH Data Archive (NDA) AMP SCZ Data Repository (https://nda.nih.gov/ampscz).

Received: 30 August 2024; Accepted: 8 October 2024;
Published online: 03 April 2025

REFERENCES

1. Addington, J. et al. Sample Ascertainment and Clinical Outcome Measures in the Accelerating Medicines Partnership® Schizophrenia Program. *Schizophrenia* https://doi.org/10.1038/s41537-025-00556-7 (2025).
2. Allswede, D. M. et al. Characterizing covariant trajectories of individuals at clinical high risk for psychosis across symptomatic and functional domains. *Am. J. Psychiatry* **177**, 164–171 (2020).
3. Salazar de Pablo, G. et al. Probability of transition to psychosis in individuals at clinical high risk: an updated meta-analysis. *JAMA Psychiatry* **78**, 970–978 (2021).
4. Salazar De Pablo, G. et al. Clinical outcomes in individuals at clinical high risk of psychosis who do not transition to psychosis: a meta-analysis. *Epidemiol. Psychiatr. Sci.* **31**, e9 (2022).
5. Addington, J. et al. Clinical and functional characteristics of youth at clinical high-risk for psychosis who do not transition to psychosis. *Psychol. Med.* **49**, 1670–1677 (2019).
6. Beck, K. et al. Clinical and functional long-term outcome of patients at clinical high risk (CHR) for psychosis without transition to psychosis: a systematic review. *Schizophr. Res.* **210**, 39–47 (2019).
7. Fusar-Poli, P. et al. Disorder, not just state of risk: meta-analysis of functioning and quality of life in people at high risk of psychosis. *Br. J. Psychiatry J. Ment. Sci.* **207**, 198–206 (2015).
8. Worthington, M. A. et al. Individualized prediction of prodromal symptom remission for youth at clinical high risk for psychosis. *Schizophr. Bull.* **48**, 395–404 (2022).
9. Hauser, T. U., Skvortsova, V., De Choudhury, M. & Koutsouleris, N. The promise of a model-based psychiatry: building computational models of mental ill health. *Lancet Digit. Health* **4**, e816–e828 (2022).
10. Brady, L. S., Larrauri, C. A. & AMP SCZ Steering Committee. Accelerating Medicines Partnership ® Schizophrenia (AMP ® SCZ): developing tools to enable early intervention in the psychosis high risk state. *World Psychiatry* **22**, 42–43 (2023).
11. Wannan, C. M. J. et al. Accelerating Medicines Partnership® Schizophrenia (AMP® SCZ): rationale and study design of the largest global prospective cohort study of clinical high risk for psychosis. *Schizophr. Bull.* **50**, 496–512 (2024).
12. Cannon, T. D. et al. An individualized risk calculator for research in prodromal psychosis. *Am. J. Psychiatry* **173**, 980–988 (2016).
13. Koutsouleris, N. 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* **90**, 632–642 (2021).
14. Rosen, M. et al. Towards clinical application of prediction models for transition to psychosis: a systematic review and external validation study in the PRONIA sample. *Neurosci. Biobehav. Rev.* **125**, 478–492 (2021).
15. Koutsouleris, N. 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* **75**, 1156 (2018).
16. Antonucci, L. A. et al. Using combined environmental–clinical classification models to predict role functioning outcome in clinical high-risk states for psychosis and recent-onset depression. *Br. J. Psychiatry* **220**, 229–245 (2022).
17. Koutsouleris, N. et al. Multimodal machine learning workflows for prediction of psychosis in patients with clinical high-risk syndromes and recent-onset depression. *JAMA Psychiatry* **78**, 195 (2021).
18. Li, Y. et al. A random forest model for predicting social functional improvement in Chinese patients with schizophrenia after 3 months of atypical antipsychotic monopharmacy: a cohort study. *Neuropsychiatr. Dis. Treat.* **17**, 847–857 (2021).
19. Smucny, J., Davidson, I. & Carter, C. S. Are we there yet? Predicting conversion to psychosis using machine learning. *Am. J. Psychiatry* **180**, 836–840 (2023).
20. Ciarleglio, A. J. et al. A predictive model for conversion to psychosis in clinical high-risk patients. *Psychol. Med.* **49**, 1128–1137 (2019).
21. Andreou, C. et al. Predictors of transition in patients with clinical high risk for psychosis: an umbrella review. *Transl. Psychiatry* **13**, 286 (2023).
22. Corcoran, C. M. et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. *World Psychiatry* **17**, 67–75 (2018).
23. Perkins, D. O. et al. Polygenic risk score contribution to psychosis prediction in a target population of persons at clinical high risk. *Am. J. Psychiatry* **177**, 155–163 (2020).
24. Perkins, D. O. et al. Towards a psychosis risk blood diagnostic for persons experiencing high-risk symptoms: preliminary results from the NAPLS project. *Schizophr. Bull.* **41**, 419–428 (2015).
25. Byrne, J. F. et al. Proteomic biomarkers for the prediction of transition to psychosis in individuals at clinical high risk: a multi-cohort model development study. *Schizophr. Bull.* **50**, 579–588 (2024).
26. Byrne, J. F. et al. Plasma complement and coagulation proteins as prognostic factors of negative symptoms: an analysis of the NAPLS 2 and 3 studies. *Brain. Behav. Immun.* **119**, 188–196 (2024).
27. Cullen, A. E. et al. Stressor-cortisol concordance among individuals at clinical high-risk for psychosis: novel findings from the NAPLS cohort. *Psychoneuroendocrinology* **115**, 104649 (2020).
28. Walker, E. F. et al. Cortisol levels and risk for psychosis: initial findings from the North American prodrome longitudinal study. *Biol. Psychiatry* **74**, 410–417 (2013).
29. Chan, M. K. et al. Development of a blood-based molecular biomarker test for identification of schizophrenia before disease onset. *Transl. Psychiatry* **5**, e601–e601 (2015).
30. Worthington, M. A. et al. Incorporating cortisol into the NAPLS2 individualized risk calculator for prediction of psychosis. *Schizophr. Res.* **227**, 95–100 (2021).
31. Mongan, D. et al. Development of proteomic prediction models for transition to psychotic disorder in the clinical high-risk state and psychotic experiences in adolescence. *JAMA Psychiatry* **78**, 77 (2021).
32. Zarogianni, E., Storkey, A. J., Johnstone, E. C., Owens, D. G. C. & Lawrie, S. M. Improved individualized prediction of schizophrenia in subjects at familial high risk, based on neuroanatomical data, schizotypal and neurocognitive features. *Schizophr. Res.* **181**, 6–12 (2017).
33. Reinen, J. M. et al. Multimodal fusion of brain signals for robust prediction of psychosis transition. *Schizophrenia* **10**, 54 (2024).
34. Sanfelici, R., Dwyer, D. B., Antonucci, L. A. & Koutsouleris, N. Individualized diagnostic and prognostic models for patients with psychosis risk syndromes: a meta-analytic view on the state of the art. *Biol. Psychiatry* **88**, 349–360 (2020).
35. Schnack, H. G. & Kahn, R. S. Detecting neuroimaging biomarkers for psychiatric disorders: sample size matters. *Front. Psychiatry* **7**, 50 (2016).
36. Poldrack, R. A., Huckins, G. & Varoquaux, G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry* **77**, 534 (2020).
37. European Network of National Networks studying Gene-Environment Interactions in Schizophrenia (EU-GEI). Identifying gene-environment interactions in schizophrenia: contemporary challenges for integrated, large-scale investigations. *Schizophr. Bull.* **40**, 729–736 (2014).
38. Addington, J. et al. North American prodrome longitudinal study (NAPLS 3): methods and baseline description. *Schizophr. Res.* **243**, 262–267 (2022).
39. Zhang, T. et al. Validating the predictive accuracy of the NAPLS-2 psychosis risk calculator in a clinical high-risk sample from the SHARP (Shanghai at risk for psychosis) program. *Am. J. Psychiatry* **175**, 906–908 (2018).
40. Mathalon, D. et al. The Electroencephalography Protocol for the Accelerating Medicines Partnership® Schizophrenia Program: Reliability and Stability of Measures. *Schizophrenia*, https://doi.org/10.1038/s41537-025-00607-z (2025).
41. Harms, M. et al. The MR Neuroimaging Protocol for the Accelerating Medicines Partnership® Schizophrenia Program. *Schizophrenia,* https://doi.org/10.1038/s41537-025-00581-6 (2025).
42. Allott, K. et al. Cognitive assessment in the Accelerating Medicines Partnership® Schizophrenia Program: Harmonization priorities and strategies in a diverse

international sample. *Schizophrenia,* https://doi.org/10.1038/s41537-025-00578-1 (2025).
43. Perkins, D. et al. Body Fluid Biomarkers and Psychosis Risk in the Accelerating Medicines Partnership® Schizophrenia Program: Design Considerations. *Schizophrenia,* https://doi.org/10.1038/s41537-025-00610-4 (2025).
44. Wigman, J. et al. Digital Health Technologies in the the Accelerating Medicines Partnership® Schizophrenia Program. *Schizophrenia,* https://doi.org/10.1038/s41537-025-00599-w (2025).
45. Bilgrami, Z. et al. Collecting Language, Speech Acoustics, and Facial Expression to Predict Psychosis and Other Clinical Outcomes: Strategies from the Accelerating Medicines Partnership® Schizophrenia Program Initiative. *Schizophrenia,* (2025).
46. Thompson, J. L. et al. Childhood trauma and prodromal symptoms among individuals at clinical high risk for psychosis. *Schizophr. Res.* **108**, 176–181 (2009).
47. Bzdok, D., Altman, N. & Krzywinski, M. Statistics versus machine learning. *Nat. Methods* **15**, 233–234 (2018).
48. Dwyer, D. B., Falkai, P. & Koutsouleris, N. Machine learning approaches for clinical psychology and psychiatry. *Annu. Rev. Clin. Psychol.* **14**, 91–118 (2018).
49. Gönen, M. & Alpaydin, E. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **12**, 2211–2268 (2011).
50. Polosecki, P. et al. Resting-state connectivity stratifies premanifest Huntington's disease by longitudinal cognitive decline rate. *Sci. Rep.* **10**, 1252 (2020).
51. Dhurandhar, A. et al. Explanations based on the missing: towards contrastive explanations with pertinent negatives. *Adv. neural. inf. process. syst.* **31**, https://doi.org/10.48550/arXiv.1802.07623 (2018).
52. Woods, S. W. et al. Development of the PSYCHS: positive symptoms and diagnostic criteria for the CAARMS harmonized with the SIPS. *Early Interv. Psychiatry* https://doi.org/10.1111/eip.13457 (2023).
53. Mason, S. J. & Graham, N. E. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: statistical significance and interpretation. *Q. J. R. Meteorol. Soc.* **128**, 2145–2166 (2002).
54. Eyigoz, E., Mathur, S., Santamaria, M., Cecchi, G. & Naylor, M. Linguistic markers predict onset of Alzheimer's disease. *EClinicalMedicine* **28**, 100583 (2020).
55. Collins, M. A. et al. Accelerated cortical thinning precedes and predicts conversion to psychosis: the NAPLS3 longitudinal study of youth at clinical high-risk. *Mol. Psychiatry* **28**, 1182–1189 (2023).
56. Zhou, X.-H., Obuchowski, N. A. & McClish, D. K. *Statistical Methods in Diagnostic Medicine*. Vol. 712 (John Wiley & Sons, 2011).
57. König, I. R., Malley, J. D., Weimar, C., Diener, H. -C. & Ziegler, A. Practical experiences on the necessity of external validation. *Stat. Med.* **26**, 5499–5511 (2007).
58. Mooney, C. & Duval, R. *Bootstrapping* (SAGE Publications, Inc., 1993).
59. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980 (2006).
60. Hua, K. et al. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *NeuroImage* **39**, 336–347 (2008).
61. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **374**, 20150202 (2016).
62. Calhoun, V. D., Adali, T., Pearlson, G. D. & Pekar, J. J. A method for making group inferences from functional MRI data using independent component analysis. *Hum. Brain Mapp.* **14**, 140–151 (2001).
63. Sotiras, A., Resnick, S. M. & Davatzikos, C. Finding imaging patterns of structural covariance via Non-Negative Matrix Factorization. *NeuroImage* **108**, 1–16 (2015).
64. Anowar, F., Sadaoui, S. & Selim, B. Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne). *Comput. Sci. Rev.* **40**, 100378 (2021).
65. Zou, H. & Hastie, T. Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67**, 301–320 (2005).
66. Saeys, Y., Inza, I. & Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**, 2507–2517 (2007).
67. Cornblatt, B. A. et al. Psychosis prevention: a modified clinical high risk perspective from the recognition and prevention (RAP) program. *Am. J. Psychiatry* **172**, 986–994 (2015).
68. Wang, Y. et al. Multi-modal intermediate integrative methods in neuropsychiatric disorders: a review. *Comput. Struct. Biotechnol. J.* **20**, 6149–6162 (2022).
69. Boser, B. E., Guyon, I. M. & Vapnik, V. N. A training algorithm for optimal margin classifiers. in *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* 144–152 (ACM, 1992).
70. Di Biase, M. A. et al. White matter changes in psychosis risk relate to development and are not impacted by the transition to psychosis. *Mol. Psychiatry* **26**, 6833–6844 (2021).
71. Zhang, T. et al. Multivariate joint models for the dynamic prediction of psychosis in individuals with clinical high risk. *Asian J. Psychiatry* **81**, 103468 (2023).
72. Yuen, H. P. et al. Dynamic prediction of transition to psychosis using joint modelling. *Schizophr. Res.* **202**, 333–340 (2018).
73. Yuen, H. P., Mackinnon, A. & Nelson, B. Dynamic prediction systems of transition to psychosis using joint modelling: extensions to the base system. *Schizophr. Res.* **216**, 207–212 (2020).
74. Cannon, T. D. et al. Notice of Retraction: Worthington MA et al. Dynamic prediction of outcomes for youth at clinical high risk for psychosis: a joint modeling approach. *JAMA Psychiatry*. 2023;80(10):1017–1025. *JAMA Psychiatry* **81**, 109 (2024).
75. Ross, J. C. et al. A Bayesian nonparametric model for disease subtyping: application to emphysema phenotypes. *IEEE Trans. Med. Imaging* **36**, 343–354 (2017).
76. Lee, C. & Van Der Schaar, M. Temporal phenotyping using deep predictive clustering of disease progression. In *International Conference on Machine Learning* 5767–5777 (2020).
77. Horvatic, D., Stanley, H. E. & Podobnik, B. Detrended cross-correlation analysis for non-stationary time series with periodic trends. *EPL Europhys. Lett.* **94**, 18007 (2011).
78. Boeker, M., Hammer, H. L., Riegler, M. A., Halvorsen, P. & Jakobsen, P. Prediction of schizophrenia from activity data using hidden Markov model parameters. *Neural Comput. Appl.* **35**, 5619–5630 (2023).
79. Sakoe, H. & Chiba, S. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* **26**, 43–49 (1978).
80. Wenzel, J. et al. Ecological momentary assessment (EMA) combined with unsupervised machine learning shows sensitivity to identify individuals in potential need for psychiatric assessment. *Eur. Arch. Psychiatry Clin. Neurosci.* https://doi.org/10.1007/s00406-023-01668-w (2023).
81. Hebbrecht, K. et al. Understanding personalized dynamics to inform precision medicine: a dynamic time warp analysis of 255 depressed inpatients. *BMC Med.* **18**, 400 (2020).
82. Permanasari, Y., Harahap, E. H. & Ali, E. P. Speech recognition using Dynamic Time Warping (DTW). *J. Phys. Conf. Ser.* **1366**, 012091 (2019).
83. Kalman, R. E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **82**, 35–45 (1960).
84. Lavieri, M. S., Puterman, M. L., Tyldesley, S. & Morris, W. J. When to treat prostate cancer patients based on their PSA dynamics. *IIE Trans. Healthc. Syst. Eng.* **2**, 62–77 (2012).
85. Garcia, G.-G. P. et al. Using Kalman filtering to forecast disease trajectory for patients with normal tension glaucoma. *Am. J. Ophthalmol.* **199**, 111–119 (2019).
86. Eberle, C. & Ament, C. The Unscented Kalman Filter estimates the plasma insulin from glucose measurement. *Biosystems* **103**, 67–72 (2011).
87. Temple, R. Enrichment of clinical study populations. *Clin. Pharmacol. Ther.* **88**, 774–778 (2010).
88. Haas, S. S. et al. Evidence of discontinuity between psychosis-risk and non-clinical samples in the neuroanatomical correlates of social function. *Schizophr. Res. Cogn.* **29**, 100252 (2022).
89. McGorry, P. D., Hartmann, J. A., Spooner, R. & Nelson, B. Beyond the "at risk mental state" concept: transitioning to transdiagnostic psychiatry. *World Psychiatry* **17**, 133–142 (2018).
90. Bouneffouf, D., Rish, I. & Aggarwal, C. Survey on applications of multi-armed and contextual bandits. In *2020 IEEE Congress on Evolutionary Computation (CEC)* 1–8 (IEEE, 2020).
91. Bouneffouf, D., Rish, I., Cecchi, G. & Féraud, R. Context Attentive Bandits: Contextual Bandit with Restricted Context. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 1468–1475 (2017).
92. Nelson, B., McGorry, P. D., Wichers, M., Wigman, J. T. W. & Hartmann, J. A. Moving from static to dynamic models of the onset of mental disorder: a review. *JAMA Psychiatry* **74**, 528 (2017).
93. Lingler, J. H., Schmidt, K. L., Gentry, A. L., Hu, L. & Terhorst, L. A. A new measure of research participant burden: brief report. *J. Empir. Res. Hum. Res. Ethics* **9**, 46–49 (2014).
94. Leanza, L. et al. Predictors of study drop-out and service disengagement in patients at clinical high risk for psychosis. *Soc. Psychiatry Psychiatr. Epidemiol.* **55**, 539–548 (2020).
95. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
96. Fix, E. & Hodges, J. L. Discriminatory analysis. Nonparametric discrimination: consistency properties. *Int. Stat. Rev. Rev. Int. Stat.* **57**, 238 (1989).
97. Azur, M. J., Stuart, E. A., Frangakis, C. & Leaf, P. J. Multiple imputation by chained equations: what is it and how does it work? *Int. J. Methods Psychiatr. Res.* **20**, 40–49 (2011).
98. Cetin Karayumak, S. et al. Retrospective harmonization of multi-site diffusion MRI data acquired with different acquisition parameters. *NeuroImage* **184**, 180–200 (2019).
99. Hunter, D. J. & Holmes, C. Where medical statistics meets artificial intelligence. *N. Engl. J. Med.* **389**, 1211–1219 (2023).

ACKNOWLEDGEMENTS

The Accelerating Medicines Partnership® Schizophrenia (AMP SCZ) is a public-private partnership managed by the Foundation for the National Institutes of Health. The AMP SCZ research program is supported by contributions from the AMP SCZ public and private partners, which include NIMH (U24MH124629, U01MH124631, and U01MH124639) and Wellcome (220664/Z/20/Z and 220664/A/20/Z). A full list of AMP SCZ members and affiliations can be found at https://www.ampscz.org/members/ and within the Supplementary File.

AUTHOR CONTRIBUTIONS

Conception and/or design of the work: all authors. Acquisition, analysis, and/or interpretation of the work: all authors. Drafted the manuscript: N.P., P.P., G.C., and O.P. Critically revised the work: all authors. Approved the submitted version and took responsibility for contributions and integrity of the work: all authors.

COMPETING INTERESTS

A.A. is a cofounder, serves as a member of the Board of Directors, as a scientific adviser, and holds equity in Manifest Technologies, Inc.; and is a coinventor on the following patent: Anticevic A., Murray J.D., Ji J.L.: Systems and Methods for NeuroBehavioral Relationships in Dimensional Geometric Embedding, PCT International Application No. PCT/US2119/022110, filed Mar 13, 2019. C.A. has been a consultant to or has received honoraria or grants from Acadia, Angelini, Biogen, Boehringer, Gedeon Richter, Janssen Cilag, Lundbeck, Medscape, Menarini, Minerva, Otsuka, Pfizer, Roche, Sage, Servier, Shire, Schering Plough, Sumitomo Dainippon Pharma, Sunovion and Takeda. D.D. has received honorary funds for one educational seminar for CSL Sequiris. G.J.P. is a full-time employee of Janssen Research & Development LLC, and a Johnson & Johnson stockholder. J.M.K. is Consultant to or receives honoraria and/or travel support and/or speakers bureau: Alkermes, Allergan, Boehringer-Ingelheim, Cerevel, Dainippon Sumitomo, H. Lundbeck, HealthRhythms, HLS Therapeutics, Indivior, Intracellular Therapies, Janssen Pharmaceutical, Johnson & Johnson, Karuna Therapeutics/Bristol Meyer-Squibb, LB Pharmaceuticals, Mapi, Maplight, Merck, Minerva, Neurocrine, Newron, Novartis, NW PharmaTech, Otsuka, Roche, Saladax, Sunovion, Teva. J.T. is Advisor to Percison Mental Wellness. Research support from Otsuka. J.K. has received speaking or consulting fees from Janssen, Boehringer Ingelheim, ROVI and Lundbeck. P.F.P. has received research funds or personal fees from Lundbeck, Angelini, Menarini, Sunovion, Boehringer Ingelheim, Proxymm Science, Otsuka, outside the current study. Q.S.L. is an employee of Janssen Research & Development, LLC and a shareholder in Johnson & Johnson, the parent company of the Janssen companies. R.S.K. is consulting: Alkermes, Boehringer-Ingelheim. R.J.G.: Grants to Brigham and Women's Hospital from Amgen, AstraZeneca, Kowa, and Novartis. S.W.W. has received speaking fees from the American Psychiatric Association and from Medscape Features. He has been granted US patent no. 8492418 B2 for a method of treating prodromal schizophrenia with glycine agonists. He owns stock in NW PharmaTech. S.V. is a full-time employee of Janssen Research & Development LLC, and a Johnson & Johnson stockholder. All other authors do not report any conflict of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41537-025-00561-w.

Correspondence and requests for materials should be addressed to Ofer Pasternak.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [2]Department of Psychiatry, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [3]IBM T.J. Watson Research Center, Yorktown Heights, NY, USA. [4]Department of Psychiatry, Hotchkiss Brain Institute, University of Calgary, Calgary, AB, Canada. [5]Department of Child and Adolescent Psychiatry, Institute of Psychiatry and Mental Health, Hospital General Universitario Gregorio Marañón, IiSGM, CIBERSAM, Instituto de Salud Carlos III, School of Medicine, Universidad Complutense, Madrid, Spain. [6]Department of Radiology, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [7]Department of Software Engineering and Information Technology, École de technologie supérieure, Montreal, QC, Canada. [8]Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [9]Departments of Psychology and Psychiatry, Yale University, New Haven, CT, USA. [10]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [11]Centre for Youth Mental Health, The University of Melbourne, Parkville, VIC, Australia. [12]Orygen, Parkville, VIC, Australia. [13]University of British Columbia, Vancouver, British Columbia, Canada. [14]Department of Psychosis Studies, King's College, London, UK. [15]Department of Brain and Behavioral Sciences, University of Pavia, Pavia, Italy. [16]Department of Biostatistics, Harvard T.H. Chan School of Public Health and Harvard Medical School, Boston, MA, USA. [17]Department of Medicine, Harvard Medical School, Boston, MA, USA. [18]Department of Psychiatry, Washington University Medical School, St. Louis, MO, USA. [19]Department of Psychiatry, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany. [20]Department of Psychiatry and Psychotherapy, Ludwig Maximilian University of Munich, Munich, Germany. [21]Data and Analytics, Otsuka Pharmaceutical Development & Commercialization, Inc, Princeton, USA. [22]Center for Devices and Radiological Health, US Food and Drug Administration, Silver Spring, MD, USA. [23]Department of Psychiatry, Seoul National University College of Medicine, Seoul, Korea. [24]Department of Neuropsychiatry, Seoul National University Hospital, Seoul, Korea. [25]Psychotic Disorders Division, McLean Hospital, Belmont, MA, USA. [26]Department of Psychiatry, Harvard Medical School, Boston, MA, USA. [27]JRD Data Science, Janssen Research & Development, LLC, Titusville, NJ, USA. [28]Division of Psychiatry, Center for Drug Evaluation and Research, US Food and Drug Administration, Silver Spring, MD, USA. [29]Department of Psychiatry and Behavioral Sciences and Weill Institute for Neurosciences, University of California San Francisco, San Francisco, CA, USA. [30]Veterans Affairs San Francisco Health Care System, San Francisco, CA, USA. [31]Department of Psychology, Northwestern University, Evanston, IL, USA. [32]Northern California Institute for Research and Education, San Francisco, CA, USA. [33]Johnson & Johnson Innovative Medicine, Titusville, NJ, USA. [34]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [35]Food and Drug Administration, Silver Spring, MD, USA. [36]S2 Consulting LLC, Danbury, CT, USA. [37]Douglas Research Centre, Montreal, QC, Canada. [38]Department of Psychiatry, McGill University, Montreal, QC, Canada. [39]Department of Health Informatics, IOPPN, KCL, London, UK. [40]Department of Psychiatry, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA, USA. [41]Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [42]Biostatistics Center, Massachusetts General Hospital, Boston, MA, USA. [43]Shanghai Mental Health Center, Shanghai Jiaotong University School of Medicine, Shanghai 200030, China. [44]Department of Psychology, Emory University, Atlanta, USA. [45]Department of Psychiatry, Yale University School of Medicine, New Haven, CT, USA. [46]Departments of Psychiatry and Biobehavioral Sciences & Psychology, Semel Institute for Neuroscience and Human Behavior, University of California, Los Angeles, CA, USA. [47]Department of Psychology, University of California, Los Angeles, CA, USA. [48]Department of Psychiatry, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA. [49]Institute of Behavioral Science, Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, USA. [50]Connecticut Mental Health Center, New Haven, CT, USA. [51]These authors contributed equally: Nora Penzel, Pablo Polosecki. [52]These authors jointly supervised this work: Guillermo Cecchi, Ofer Pasternak. *A list of authors and their affiliations appears at the end of the paper. ✉email: ofer@bwh.harvard.edu

ACCELERATING MEDICINES PARTNERSHIP® SCHIZOPHRENIA (AMP® SCZ)

Nora Penzel[1,2,51], Pablo Polosecki[3,51], Jean Addington[4], Celso Arango[5], Ameneh Asgari-Targhi[6], Tashrif Billah[2], Sylvain Bouix[2,7], Monica E. Calkins[8], Dylan E. Campbell[2], Tyrone D. Cannon[9], Eduardo Castro[3], Kang Ik K. Cho[2], Michael J. Coleman[2], Cheryl M. Corcoran[10], Dominic Dwyer[11,12], Sophia Frangou[10,13], Paolo Fusar-Poli[14,15], Robert J. Glynn[16,17], Anastasia Haidar[2], Michael P. Harms[18], Grace R. Jacobs[2], Joseph Kambeitz[19], Tina Kapur[6], Sinead M. Kelly[2], Nikolaos Koutsouleris[14,20], K. R. Abhinandan[21], Saryet Kucukemiroglu[22], Jun Soo Kwon[23,24], Kathryn E. Lewandowski[25,26], Qingqin S. Li[27], Valentina Mantua[28], Daniel H. Mathalon[29,30], Vijay A. Mittal[31], Spero Nicholas[30,32], Gahan J. Pandina[33], Diana O. Perkins[34], Andrew Potter[35], Abraham Reichenberg[10], Jenna Reinen[3], Michael S. Sand[36], Johanna Seitz-Holland[1,2], Jai L. Shah[37,38], Vairavan Srinivasan[33,39], Agrima Srivastava[10], William S. Stone[40], John Torous[40], Mark G. Vangel[41,42], Jijun Wang[43], Phillip Wolff[44], Beier Yao[25,26], Alan Anticevic[45], Daniel H. Wolf[8], Hao Zhu[35], Carrie E. Bearden[46,47], Patrick D. McGorry[11,12], Barnaby Nelson[11,12], John M. Kane[48,49], Scott W. Woods[45,50], René S. Kahn[10], Martha E. Shenton[1,2,6], Guillermo Cecchi[3,52] and Ofer Pasternak[1,2,6,52] ✉

A full list of members and their affiliations appears in the Supplementary Information.