Learning latent profiles via cognitive growth charting in psychosis: design and rationale for the PRECOGNITION project

Andre F. Marquand[1*], Barbora Rehák Bučková[1*], Giulia Cattaranusi[2*], Camilla Flaaten[3,4*], Cecilie Busch[3,4*], Cecilie K. Lemvigh[5], Veenu Gupta[6], Charlotte Fraza[1], Lars T. Westlye[3,4], Ole A. Andreassen[3,4], Jaroslav Hlinka[8,9], Bjørn H. Ebdrup[5,7], David Shiers[10,11,12+], Torill Ueland[3,4+], Paola Dazzan[2+]

1: Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Centre, Nijmegen, the Netherlands

2: Department of Psychological Medicine, institute of Psychiatry, Psychology and Neuroscience, King's College London, U.K.

3: Division of Mental Health and Addiction, NORMENT, Oslo University Hospital & Institute of Clinical Medicine, University of Oslo, Oslo, Norway.

4: Department of Psychology, University of Oslo, Oslo, Norway.

5: Center for Neuropsychiatric Schizophrenia Research (CNSR) & Center for Clinical Intervention and Neuropsychiatric Schizophrenia Research (CINS), Mental Health Centre, Glostrup, Copenhagen University Hospital - Mental Health Services CPH, Copenhagen, Denmark

6: Department of Psychology, Durham University, U.K.

7: Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

8: Department of Complex Systems, Institute of Computer Science of the Czech Academy of Sciences, Prague, Czech Republic

9: National Institute of Mental Health, Klecany, Czech Republic

10: Honorary research consultant, Psychosis Research Unit, Greater Manchester Mental

Health NHS Trust, Manchester, U.K.

11: Honorary Reader in early psychosis, University of Manchester, U.K.

12: Honorary Senior Lecturer, School of Medicine, Keele University, U.K.

*: shared first authorship

+: shared senior authorship.

Abstract

Background and Hypothesis: Cognitive impairments are a core feature of psychosis that are often evident before illness onset and have substantial impact on both clinical and real-world functional outcomes. Therefore, these are an excellent target for stratification and early detection in order to facilitate early intervention. While many studies have aimed to characterise the effects of cognition at the group level and others have aimed to detect individual differences by referencing subjects against existing norms, these studies have limited generalisability across clinical populations, demographic backgrounds and instruments and do not fully account for the inter-individual heterogeneity inherent in psychosis.

Study Design: Here, we outline the rationale, design and analysis plan for the PRECOGNITION project which aims to address these challenges.

Study Results: This project is a collaboration between partners in five European countries, The project will not generate any primary data, but by leveraging existing datasets and combining these with novel analytic methods, it will produce multiple contributions including: (i) translating normative modelling approaches pioneered in brain imaging to psychosis data, to yield 'cognitive growth charts' for longitudinal tracking and individual prediction (ii) developing machine learning models for harmonising and stratifying cohorts on the basis of these models; (iii) providing integrated next-generation norms, having broad sociodemographic coverage including different languages and distinct norms for individuals with psychosis and unaffected individuals

Conclusions: This study will enable precision stratification of psychosis cohorts and furnish predictions for a broad range of functional outcome measures. It will be guided throughout by lived experience experts.

Introduction

Cognitive impairments, such as impairments in memory, reasoning and concentration, are a core feature of psychosis. They are often evident before illness onset[1–5] and have substantial impact on functional outcomes and activities of daily living[6]. This suggests that cognitive functions should be a priority target for developing early detection and stratification approaches to facilitate early intervention. However, while cognitive alterations may have high sensitivity early in the course of the illness, they have low specificity in that impairments are present across several cognitive domains[2] and are also evident early in the course of a range of common behavioural and mental disorders. Indeed, a recent meta-analytic report[4] provided evidence that group level impairments are evident across all cognitive domains in first episode psychosis, bipolar disorder and depression. Moreover, differentiation between disorders is principally evident in the magnitude of the effects reported[4] rather than in the nature of the effects, although some differences exist[4]. This suggests that precise quantification of the magnitude, timing and progression of cognitive impairments are essential in order to parse the heterogeneity underlying psychosis.

An additional problem is that studies to date have overwhelmingly focussed on group level analyses, which mask considerable inter-individual variation that may have clinical relevance. This limitation is a major barrier in our ability to predict the onset and course of illness and to the development of personalized early interventions. The heterogeneity of mental disorders is well recognised in theoretical models[7–11] and the importance of modelling individual differences is widely acknowledged[10–13]. However, this theoretical recognition is not reflected in the typical 'case-control' paradigm often used to study psychiatric disorders: nearly all approaches used in practice are implicitly rooted in the 'case-control' paradigm which assumes clinical groups are homogeneous and well-defined (patient/control, responder/remitter).

While this has helped to partly understand the mechanisms of mental disorders, this assumption of clean distinctions between clinical groups is often unrealistic and focusing on group effects masks inter-individual variation that may be crucial for individual prediction. For

example, the clinical group may be diffuse and heterogeneous due to comorbidity and to a convergence of different pathophysiological pathways on the same symptoms. In line with this, there has been considerable effort invested in finding subgroups of patients based on cognition or symptoms over the last 50 years and many putative stratifications have been proposed[14–20], yet none have advanced beyond proof-of-concept. We have suggested that one of the main reasons is that clustering algorithms always yield a result, regardless of whether clearly defined clusters are evident in the data, resulting in ill-defined and poorly replicated subtypes.[21,22]

Whilst clinical neuropsychology has traditionally adopted a single subject perspective by referencing subjects against group-level norms, these approaches also have limitations. For example, in a research context, most studies also apply group level analyses to detect average differences between cohorts based on normed data. Moreover, existing norms are typically generated for specific tests or batteries and based on a single population. These have limited generalisability across different tests, more demographically representative populations and across languages. Also, existing approaches do not fully capture the potentially complex shape of different test scores, which often have non-linear, non-Gaussian and heteroskedastic effects across the lifespan. Finally, it is questionable whether the same norms should be applied to healthy individuals and to individuals with psychosis. Indeed, we have shown that the variability amongst individuals with psychosis is greater than within the healthy population for most cognitive functions and norms estimated on healthy populations are therefore suboptimal for understanding inter-individual differences in psychosis (e.g. due to the presence of floor or ceiling effects).[23] We have also provided proof of concept evidence that the approach we propose here can help to make progress.[24]

Methods

The Wellcome-Trust funded PRECOGNITION project aims to address these problems, by applying the normative modelling approach developed in brain imaging to cognitive data. The normative modelling framework generalises the notion of pediatric growth charts aiming

to model centiles of variation in the population as a function of clinically relevant covariates, whilst providing statistical inferences at the level of the individual person.[21,25] It has proved to be especially useful in understanding inter-individual variability in psychosis.[26–30] More specifically, this project aims to: (i) apply normative modelling to cognitive data, to yield 'cognitive growth charts'; (ii) develop machine learning-based tools to harmonise disparate cognitive data from independent studies to a common reference to enable this to be applied to a wide range of cognitive instruments; (iii) assemble and curate demographically diverse reference datasets derived from large, population level cohorts in order to define comprehensive reference models that provide equitable predictions for a wide range of sociodemographic, racial and clinical backgrounds. This will yield considerably greater generalisability relative to existing norms for cognitive data. Lastly, we will (iv) use machine learning models on the basis of these models to stratify individuals from early psychosis cohorts and predict their clinical and functional outcomes (e.g. educational and vocational function) across short, medium and long timescales. A crucial feature of this project is that it has been co-developed together with experts with lived experience to promote real-world impact and relevance, which has been shown essential for interpreting findings and shaping research questions.[31] Guided by the input from these experts we will refine research questions towards concerns they identify using focus groups, dissemination and collaborative governance to ensure that our approach aligns with their perspectives. Ultimately, we hope that providing a nuanced understanding of cognitive impairments in individuals with psychosis can improve individualised care and aligns with the motivations of the lived experience experts contributing to the project.

The PRECOGNITION project comprises partners from the Netherlands (Donders Institute / Radboud University Medical Centre), the UK (Institute of Psychiatry, Psychology and Neuroscience, King's College London), Norway (Oslo University Hospital, University of Oslo), Denmark (Mental Health Services CPH, Copenhagen University Hospital), and the Czech Republic (National Institute of Mental Health, Klecany). The project is structured

around three integrated, yet complementary work packages (WPs), each aiming to solve key challenges (Figure 1).

[Figure 1 about here]

WP1: Data integration and Harmonisation

Understanding variation amongst individuals with psychosis requires an understanding of how such variation is nested within variation across the wider population and lifespan. This in turn requires large samples and integration across studies to accurately quantify variability across individuals. There are several challenges for this endeavour in psychiatry, the most pertinent are a lack of existing large reference datasets with consistent psychometric instruments and a lack of harmonized assessment across studies. Rather, different studies typically assess cognition using variants of different tests, derived from different cognitive batteries, with poor standardisation across studies. The PRECOGNITION project aims to address this problem and enable multiple datasets to be combined via a novel imputation-based approach to harmonise studies, our proposed high-level framework.[32] Whilst this approach shares similarities with conventional multiple imputation techniques, it is crucial to recognise that conventional data imputation approaches are unsuitable for this task because of strong dependencies within the pattern of missing data across cohorts, which is referred to as 'structured missingness'[33] (discussed in more detail below). Theoretical frameworks and practical tools to accommodate structured missingness are only beginning to be developed within machine learning. Therefore, a key objective of this work package is to develop the appropriate analytical methodology to adequately deal with this task and produce 'complete' datasets that allow the estimation of normative models across heterogeneous samples where not all instruments might have been included in each study. This approach will enable us to reconstruct different cognitive instruments and impute missing data whilst faithfully representing the different distributions of different variables in

the cognitive battery. This is useful, for example, if certain studies only acquired a subset of tasks from a given cognitive battery. In addition, if the different cognitive batteries measure similar cognitive constructs (e.g. different measures of working memory from different cognitive tasks), this approach will also allow us to translate between instruments that measure the same underlying construct, even if the number of instruments and their distributions differ. We will validate all models extensively, using simulations, out of sample data and validation across and within cohorts.

WP2: Develop normative models for cognition ('cognitive growth charting')

The goal of this work package is to estimate normative models for population-level cognitive data across multiple domains (e.g. intellectual functioning, processing speed, working memory and verbal learning). For example, normative modelling has been shown to be useful for mapping variation in brain imaging measures as a function of clinical covariates, while also accounting for structured variation within the population (e.g. study centre, demographic group, diagnosis).[25,34,35] and we have also shown proof of concept evidence that it is also suited to cognitive measures.[24] This allows us to: (i) map centiles of variation across the population that vary smoothly across the lifespan; (ii) make statistical inferences as to where each individual participant can be placed within the population range (e.g. at which centile, and at what level of certainty) and (iii) detect individuals with an atypical or abnormal profile. A key feature of normative modelling techniques is that they enable us to parse heterogeneity at the level of the individual person, without needing to assume that effects overlap across individuals.

It is important to recognise that whilst our approach aligns with the classical motivation for neuropsychological testing in terms of placing individuals within population norms, it also goes considerably beyond conventional statistical approaches for exploiting neuropsychological data for group-level research purposes. In contrast to the conventional approach, which either involves grouping subjects to match particular age bins or, for research purposes, performing a linear regression to adjust for age and then fitting a

parametric distribution to the residuals,[36,37] normative modelling captures the potentially complex shape of different measures, using non-linear centile curves that vary smoothly across the lifespan along with flexible distributions to map variation across individuals and random effect structures to model clinical and demographic structure within the population. This provides multiple benefits in that it: (i) accounts for differences in variance and distributional shape as a function of the input variables and ensures that centiles do not cross across the lifespan;[38] (ii) allows us to efficiently model different clinical and demographic backgrounds in the same model, for example providing a model that provides norms that are properly adjusted for sociodemographic background and that are calibrated for individuals with psychosis, and which (iii) cleanly separates different sources of variance, for example separating inter-subject variation from uncertainty in the model parameters. This provides more precise statistical inferences and greater power for detecting individual deviations, especially in the outer centiles where the data are frequently the sparsest. Finally, (iv) this approach provides coherent inferences in longitudinal studies with respect to a common reference model, for example by modelling the velocity of cognitive changes across the lifespan.[39]

Since many of the datasets in which we will apply this methodology are longitudinal, we will develop analytical methodology to accommodate longitudinal data, where changes in the expected deviation of subjects across longitudinal timepoints can be detected as centile crossings.[40] Preliminary data derived from cohorts of individuals with a diagnosis of schizophrenia in neuroimaging give good evidence that normative modelling techniques become exquisitely sensitive for detecting longitudinal changes.[30] However, we will also develop methodology to model longitudinal change at the individual level, whilst accounting for potentially confounding factors, based on approaches developed within the pediatric growth charting literature.[39,41,42] This framework will also accommodate potential biases specific to longitudinal cognitive studies, for example those that may be due to practice or motivational effects (e.g. where individuals with a mental illness may have a higher or lower motivation to perform the tests). We will give careful attention to these issues; for example,

to differentiate practice effects from clinically relevant variation, we will validate longitudinal trajectories in individuals with a diagnosis of schizophrenia against matched controls.

Finally, we will fit latent variable models (such as structural equation models or canonical correlation analysis) to the individual deviation scores derived from these normative models and use these as a basis for stratifying individuals from psychosis cohorts. These latent variable models –unlike classical clustering methods– do not force individuals to belong to one subtype but rather provide overlapping latent profiles that confer risk or resilience for psychosis, where multiple profiles can be expressed simultaneously within individuals and the same profile can be expressed differently across individuals. We refer to these neuropsychological profiles as 'cognitive fingerprints' (i.e. describing the strengths and weaknesses of individuals across different cognitive domains). This approach will provide more precise stratification of individuals and the ability to identify and map the convergence of different biological, environmental and genetic mechanisms on the same clinical phenotype and to identify and utilize complementary information provided by cognition over and above the information provided by neuroimaging and other types of information. We have shown the value of this approach in multiple publications.[43–45]

WP3: Use neuropsychological profiles to predict outcome and stratify psychosis
In this WP, we will evaluate and comprehensively benchmark the utility of the neuropsychological profiles identified in WP2 to predict functioning in early psychosis, from multiple angles and across short, medium and long timescales. We will also associate the neuropsychological profiles with deviations derived from normative models of neuroimaging features we have already brought online (for example derived from structural MRI and functional connectivity),[46–48] and with genetic markers. In order to assess functioning broadly, we will begin with symptom measures and measures of overall functioning (e.g. Global Assessment of Functioning scales) derived from clinical follow-up (when available) but we also aim to move beyond these and validate against real-world functional outcomes including educational and vocational status. These outcomes will be co-defined together with

the lived experience experts in our team and other consultants. Since this work package is the most extensive and clinically multifaceted, we will subdivide the work according to three clinical objectives:

1. Use cognition to improve prediction of real-world functional outcomes in psychosis

2. Map individual variation in cognition to neurobiology (MRI) to stratify early psychosis

3. Relate premorbid cognition to neuroimaging and genetics using prospective cohorts.

Cohorts and Measures

To achieve these objectives, we will make use of multiple cohorts, including large population-based cohorts to estimate variation across the spectrum of functioning across different cognitive domains. In order to ensure that the normative models we develop are as broadly representative as possible, we will leverage data from large scale international initiatives including the Human Connectome Project Early Psychosis initiative (humanconnectome.org/study/human-connectome-project-for-early-psychosis) and samples that assess premorbid functioning in healthy cohorts enriched for psychopathology, including psychosis, e.g. the Philadelphia Neurodevelopmental Cohort (PNC)[49], the Adolescent Brain Cognitive Development (ABCD) study[50,51] and also the Norwegian Mother and Child Cohort study (MoBa), which is an ongoing population-based prospective birth cohort that follows 114,000 individuals from birth in addition to their fathers and mothers (N>240,000 including both parents).[52] These cohorts are described in Table 1 below.

In addition, we will make use of large cohorts of individuals with early psychosis, derived from four European countries. Specifically, the thematically organised psychosis (TOP) dataset includes over 1,100 individuals diagnosed with schizophrenia, including first-episode psychosis cases, across a wide age range. Similarly, three datasets supplied by King's College London (STRATA, AESOP, GAP) together contain over 1,000 cases of first-episode

schizophrenia, with participants ranging in age from 18 to 65 years. Additionally, the Copenhagen University Hospital and the National Institute of Mental Health of the Czech Republic will contribute over 200 first-episode cases. Where possible, we will also incorporate data from available open-access initiatives, such as the Human Connectome Project Early Psychosis dataset. Full details about these samples are provided in Table 2 below. All these samples have extensive cognitive assessments, using standardized tasks plus matched controls, extensive clinical and functional assessments[53], neuroimaging and genetics.

A unique and compelling feature of our project is that we have access to extensive health, vocational and educational outcomes via Scandinavian registries (for example linked with the MoBa study described above). We will make use of this information to assess functional outcomes as broadly as possible and in a manner that faithfully reflects functioning in the real world, not only over the timeframe of clinical follow-up, but also across the lifespan. Furthermore, as mentioned above, these outcomes will be guided by input from our experts with lived experience.

Another key feature of our project is that most of the cohorts we will use are genotyped and have neuroimaging data in addition to information on environmental stressors allowing us to assess the interplay between cognition, environmental and neurobiological influences.

Table 1: Population-based cohorts

	Dataset	Subjects	Ages	Notes
1	ABCD[50]	11,000	9-20	Longitudinal cohort enriched for risk. With neuroimaging + genetics
2	Philadelphia Neurodevelopmental Cohort[49]	10,000	5-21	Neurodevelopmental sample, enriched for risk, neuroimaging in N=1200
3	ALSPAC[54]	>10,000	18-30 at baseline	Multigenerational longitudinal cohort study with imaging in a subsample
4	HCP[55]	1,113	22-40	High-quality neuroimaging data with genetics
5	HCP Lifespan[55]	1,000	5-90	High-quality neuroimaging and genetics
6	IDLaS[56]	1,000	18-60	Sample with neuroimaging in 200. This sample contains a comprehensive cognitive battery developed at the host institute.
7	HealthyBrain[57]	1,000	30-40	Longitudinal population sample with imaging and genetics and the IDLaS battery.
8	CamCan[58]	648	18-88	Healthy lifespan sample with extensive phenotyping, neuroimaging and genetics
9	MoBa[52]	114,000 (240,000 with parents)	Before birth-age 21	Ongoing prospective birth cohort of adolescents and their parents with links to school, socioeconomic and Norwegian health registries.
10	UK Biobank	500,000	50-80	Population-based cohort of 500,000 participants with genetics, cognition and health data
11	BRAINMINT[59,60]	1,200	8-45	Ongoing longitudinal population-based study on brain, cognition and mental health, including MRI, cognitive, genetics and clinical phenotypes with possibilty to link to Norwegian population and health registries

Table 2: Psychosis cohorts. Abbreviations: FEP=first episode psychois, SCZ= schizophrenia, UHR=ultra-high risk

	Dataset	Subjects	Ages	Notes
1	TOP[26] (Oslo)	1,100 SCZ 1,200 HC	20-65	Cohort study, with imaging and genetics and Norwegian health and population registries.
2	STRATA[61] (KCL)	150 FEP	18-48	Two sub-studies: 1. Cross-sectional comparison of treatment responders and non-responders. 2. Prospective cohort study of FEP individuals recently starting their first medication, followed for 6 weeks
3	AESOP[62] (KCL)	557 FEP	18-60	Prospective cohort study, with 10-year follow up, neuroimaging and genetics
4	GAP[63] (KCL)	410 FEP 370 HC	18-65	Cross-sectional study with clinical follow up, neuroimaging and genetics.
5	Mental Health Services CPH, Copenhagen University Hospital[64–67]	200 FEP 200 HC 50 UHR	18-45	Cohort studies with longitudinal follow-up and links to Danish registries. This includes sub-studies of medication naïve first episode patients, longitudinal treatment cohorts and utra-high risk (UHR) individuals
6	NIMH (Prague)	200 FEP 200 HC		Prospective cohort study with longitudinal follow up at 1 year plus imaging.
7	HCP early psychosis	320 FEP 80 HC	16-35	Imaging protocol matched to healthy HCP samples

Cognitive and outcome measures

Across the different cohorts (summarized in Table 1 and 2), a range of different tests were administered based on a number of standard cognitive batteries including the Wechsler Abbreviated Scale of Intelligence (WASI), Wechsler adult intelligence scale (WAIS),[68] Brief Assessment of Cognition in Schizophrenia (BACS)[69], National Adult Reading Test (NART),[70] Delis-Kaplan Executive Functioning System (D-KEFS),[71] MATRICS Consensus Cognitive Battery (MCCB),[53] the Cambridge Neuropsychological test automated battery (CANTAB)[72] and the National Institute of Health (NIH) Toolbox for the assessment of cognitive function.[73]

Cognitive data for the MoBa sample are provided by the Moba BrainHealth initiative (https://www.fhi.no/studier/moba/undersokelser/hjernehelse/) which invites participants from MoBa to complete online cognitive testing using the Memoro platform (https://memoro.medisin.ntnu.no). BRAINMINT participants have completed a standardized computerized test battery assessing various cognitive domains[60,74]. Note that although the methods we will develop in WP1 will provide item-level predictions, for simplicity, we will group cognitive instruments according to their underlying cognitive domains, namely *intellectual functioning, processing speed, verbal learning and memory, semantic fluency, inhibitory control, working memory and fine-motor speed.*

All clinical studies (Table 2) also have measures of symptom severity including the Positive and Negative Symptom Scales (PANSS) and measures of global functioning including the Global Assessment of Functioning (GAF) scales at baseline and at clinical follow-up. In addition, many of the samples we include allow us to assess functioning in terms of long-term health related outcomes. For example, the UK Biobank has links with primary care health related data and the Scandinavian samples allow us to assess lifetime functioning using health registry outcomes (e.q. quantifying the frequency, timing and duration of interactions with mental health services as well as prescription drugs across the lifespan). More importantly, as outlined above, we are also able to obtain measures of functioning derived, for example, from employment status and other vocational outcomes via the Norwegian Labour and Welfare Administration (NAV) registry (see e.g. https://www.nav.no/en/home/employers/nav-state-register-of-employers-and-employees). We can similarly leverage education registry data (https://www.ssb.no/a/english/mikrodata/datasamling/nudb/nudb_20130607-en.html) to derive proxy measures of educational attainment, as we have done in previous studies.[75]

Analysis Plan

A high-level overview of the analytic workflow we will follow in the project is shown in Figure 2. We follow a similar staged protocol for data harmonisation that we have outlined previously,[32] which involves the following stages:

Expert review

In this stage, the goal is to determine which instruments can be aligned directly and which instruments assess the same underlying construct, but which may elicit different distributions or have different psychometric properties. This involves input from clinicians, methodologists and psychometricians and the goal is to determine which instruments can be directly aligned by simple operations, which require a more nuanced statistical harmonisation approach, and crucially, which instruments ought not to be combined (e.g. because they measure different underlying psychological constructs).

Pre-statistical harmonisation

In this stage, data from directly comparable instruments (e.g. different variants of the same test) are combined using simple operations. For example, this might involve rescaling to accommodate structural differences in the tests. A simple example is shown in Figure 2A, where two measures of symbol coding were combined by adjusting for the length of time allowed under the test to perform the symbol coding. As shown in the figure, this yields a measure that has equivalent psychometric properties under the lifespan.

Statistical harmonisation

In this stage, we will apply an imputation-based approach to combine data from different studies. This is illustrated in Figure 2B, where we show a graphical representation of the cognitive data from the TOP dataset which consists of two partially overlapping cognitive batteries that were administered to different subjects (see [23]). This induces a strong pattern of structured missingness within the data (Figure 2B, left) which will be addressed by custom built imputation techniques to accommodate this structure (right). To achieve this, we will

build on the emerging machine learning discipline of structured missingness[33] which involves building a generative model (e.g. implemented using Bayesian hierarchical models) that aims to model the missing data mechanisms as accurately as possible during the imputation process. In this specific case, structured missingness arises because different tests were administered at different sites, creating systematic gaps in the dataset. To incorporate this structure into the imputation model, we will explicitly model site- and test-specific missingness patterns, ensuring that imputations reflect the conditional dependencies between cognitive measures rather than treating missing values as random. This approach will allow us to account for site-level variability while preserving meaningful within-site relationships.

To validate imputation quality, we will conduct a simulation study to determine optimisation criteria that best capture imputation accuracy under structured missingness. Additionally, we will develop validation procedures for real-world data, assessing imputation performance using cross-validation and distribution matching. These evaluations will guide the selection of the most appropriate imputation strategy, ensuring that reconstructed data maintain both cross-site comparability as well as multivariate dependencies.

Normative modelling

Finally, we will estimate normative models for each harmonised measure. In order to estimate the normative models, we will employ hierarchal Bayesian regression techniques that we have developed for neuroimaging.[34,35,76] These provide several important features for our purposes, namely sufficient flexibility to model site effects, non-Gaussianity and heteroskedasticity. This will also allow us to model clinical variables as random effects, effectively allowing us – for example – to estimate reference models that provide distinct centile curves for individuals with and without psychosis, also accounting for sociodemographic characteristics. We will extensively validate all models out of sample and

across cohorts to ensure generalisability. An example is shown in Figure 2C that illustrates a cognitive measure derived from the NIH toolbox that shows non-linearity across the lifespan.

We will employ a multi-stage strategy to validate the cognitive norms we develop against existing norms. First, we will check their calibration in independent datasets. That is, we will ensure that the group level distributions of both proposed and existing norms are faithful to the distribution observed in the empirical data. We will do this first in cohorts well matched demographically to the reference cohort used to derive the norms, then in additional cohorts with progressively greater demographic diversity and then subsequently across different languages. In addition, we will check the coverage and granularity of existing and proposed norms, which is important to ensure adequate coverage of the full range of functioning (i.e. from good to very poor cognitive performance) and particularly important for languages for which existing norms are defined based on relatively small cohorts. Finally, we will validate both types of norms via their ability to predict external variables (i.e. diagnosis and outcome).

Application to clinical cohorts

We will then apply these models to the clinical cohorts in order to generate deviation scores for individuals with psychosis, which will be validated extensively within and across cohorts. We will then apply latent variable models to the deviation scores in order to obtain latent profiles across cognitive domains, for example Bayesian non-parametric variants of factor analysis that properly accommodate the distribution of different measures.[77] Next, we will determine whether these predict both clinical and functional outcome measures using penalised regression models and against external measures (for example imaging and genetics), which can provide additional validation to support the stratifications we will derive. Finally, we will evaluate different proxy measures for cognition (e.g. secondary school educational attainment), that can be applied to large scale studies, and then use these to search for pre-morbid markers in prospective population cohorts with data acquired before the onset of the first psychosis episode (for example MoBa[52]).

[Figure 2 about here]

Results

We show preliminary results consisting of data derived from existing studies in Figure 3. These illustrate several of the key problems this project aims to solve. In the left panel, we plot the scores from the California Verbal Learning Test (CVLT)[78] and the Hopkins Verbal Learning Test (HVLT)[79] as a function of age. This shows two challenges: first, the CVLT shows clear age-related heteroskedasticity in that the variance increases with age. This should be accounted for in the modelling in order to ensure accurate inferences. Second, these tests should – from a theoretical perspective – measure the same underlying construct (i.e. verbal learning), but the CVLT involves recalling a list of 15 words, relative to 12 words from the HVLT. While it may be tempting to combine these instruments it is clear that this change to the test also changes its psychometric properties over the lifespan in that the CVLT shows age-dependence, whereas the HVLT does not. In the middle panel, we show an instrument, the 'Mazes' score from the MATRICS battery (assessing planning and foresight), that was administered during the second TOP battery, but which does not have a direct correspondent in the first TOP battery (although other tasks measure related abilities). This therefore needs to be addressed via a statistical approach. In the right panel, we show reading age and picture vocabulary scores derived from the NIH toolbox[73] which shows: (i) highly nonlinear effects across the lifespan and (ii) strong cohort effects, despite the same test being administered across different cohorts. Again, these effects must be properly accounted for in the modelling.

[Figure 3 about here]

Lived Experience Involvement

A notable feature of the PRECOGNITION project is engagement throughout with lived experience experts to guide the project throughout its duration. This input is multi-faceted but especially important topics include helping the project to focus on functional outcomes most relevant to individuals with psychosis, where such measures are available, and discussing

the broader societal implications of the project outcomes from the outset. These activities will be coordinated with involvement of lived experience experts within the project team, their network of people with lived experience of psychosis in the UK, in Norway and more broadly through EU Networks we are part of (e.g. the European Brain Research Area and GAMIAN-Europe). At the outset of the project, we conducted a launch event attended by scientists and individuals with lived experience in order to gain lived experience feedback.

Early in the project, we will conduct focus groups and interviews with individuals with lived experience to provide us background information in order for us to define, select and refine functional, clinical and psychosocial outcomes for us to focus on as primary endpoints. The focus groups will be conducted by the study team, guided by input from members of the project team with lived experience of psychosis who have guided the selection of questions to include. More specifically, at study onset, we will conduct one-to-one interviews with people with lived experience of psychosis and focus groups with unpaid carers of people with lived experience of psychosis at each of the main clinical sites (London and Oslo). During the one-to-one interviews, we will aim to understand what are the cognitive deficits that individuals with psychosis consider to be the most impairing. During the focus groups, we will discuss the carers' personal experience of caring for a person with psychosis and cognitive impairments. This will help the study team decide which to consider as endpoints for the prediction models we will develop in WP3. Focus groups and one-to-one interviews data will be transcribed verbatim and subjected to reflexive thematic analysis.[80] Reflexive thematic analysis is a theoretically flexible approach widely used in mental health research to analyze qualitative data.[81–83] Reflexive thematic analysis will include inductive and deductive coding, to ensure that the analysis is shaped by participants' priorities and concerns.[84].

We will use the outcomes from the one-to-one interviews and the focus groups in several ways. First, the results of the qualitative analysis will give us a deeper understanding of the personal experience of cognitive impairment in people with psychosis, as well as of their carers' experience. This is crucial, as it will ultimately help to move the field forward

towards personalized medicine approaches. Second, we aim to understand which aspects of cognitive impairment are principally responsible for the resulting functional impairment. Moreover, lived experience can provide insights into the evolving nature of specific impairments that accumulate collectively over time to impact on functional impairments. This is important because most measures of functioning that are used in the literature are quite broad, both formal functional measures (e.g. Quality of Life scales) and proxy vocational measures (e.g. 'employed' or 'not employed') and these may not capture the everyday experiences relevant to people with psychosis experiencing cognitive impairment, nor its dynamic nature and implications for future functioning. Third, this will enable us to prioritise specific functional targets in the cohorts where these measures are available; for example, giving priority to specific items of psychometric functioning scales over aggregate scores, or prioritising vocational outcomes over psychometric functioning scores if these scores do not capture the core attributes driving impairment. Fourth, this will help to guide the selection and development of more fine-grained measures of functioning for future studies and providing guidance on how these relate to cognition. Finally, this could also provide information about what areas of functioning we can expect cognitive interventions to have an impact on, which could in the longer term be used to help tailor cognitive remediation programmes.

At the end of the project, we will organise seminars and webinars with people with lived experience of psychosis to present the results of the models derived in WPs 1 and 2, discuss risks and benefits of using these predictive models, and include recommendations regarding future stages of research. host webinars to communicate our findings to the lay public, healthy personnel and people with lived experience. Again, we will conduct these activities in tandem with lived experience experts.

In addition, we will rely on lived experience experts within our project team and external lived experience experts to help us with other tasks, including discussing and interpreting study findings and in providing input into articles specifically directed at mental

health organisations to describe our research. Finally, we will rely on our lived experience experts to guide the dissemination of our results, to ensure that the communication of our findings at scientific fora, to service user communities and to the lay public is, clear, understandable, ethically informed and sensitive to the potential issues of inadvertent stigmatization of groups of individuals. As noted above, an important aspect of the work will include clarifying the ethical and user relevant aspects of applying prediction models in psychiatry, as well as clarifying the clinical and societal aspects of the project outcomes.

Ethical Considerations

Addressing ethical considerations in the PRECOGNITION project necessitates a structured approach to data privacy and the responsible application of machine learning (ML) in clinical decision-making. With regard to data privacy, the collection and analysis of sensitive health data (such as diagnosis and outcome) raise concerns regarding patient confidentiality and the potential for misuse. In this project, we will adhere to all relevant regulatory frameworks such as the General Data Protection Regulation (GDPR), which mandates explicit consent for data processing, ensures transparency in data usage, and upholds individuals' rights over their personal information. In addition, all data will be analysed within protected computing environments using decentralised ('federated') ML techniques where applicable.[34,76] Particular care will be taken for the Nordic registry data used for validation, which will be analysed *in situ* according to national and local regulatory requirements.

Additional concerns relate to bias and fairness of ML models, both at the algorithmic and data levels. For instance training on biased datasets can perpetuate existing healthcare disparities, leading to unequal outcomes across different patient demographics. While this is difficult to mitigate against entirely because most existing datasets are biased toward so-called 'WEIRD' (white, educated, industrialised, rich and democratic) populations, we are developing tools to deal with this explicitly within a normative modelling context.[85] Indeed, one of the key objectives from this project is to generate cogntiive norms that are more generalisable across countries, langauges and demographic contexts. In addition, we will

employing techniques to identify and correct biases within ML algorithms, in line with current recommendations.[86]

An additional concern relates to the "black box" nature of some ML models, which can lead to a lack of transparency, making it difficult for clinicians to understand and trust algorithmic decisions. We will employ 'explainable AI' techniques, cross-cohort validation and active monitoring of model performance over time to mitigate this. Finally, over-reliance on ML systems without adequate human oversight can lead to ethical dilemmas, especially when algorithms make erroneous or biased recommendations. We emphasize that in this project we do not aim to develop any automated decision systems, rather we will develop ML tools to support and enhance human decision-making, rather than replace it, ensuring that final clinical judgments rest with qualified healthcare professionals. This is particularly important considering that individuals may not wish to be informed of predictions of outcome (for instance if such predictions are unfavourable).

By systematically addressing these ethical considerations, the PRECOGNITION project aims to responsibly integrate ML into psychosis research, thereby enhancing predictive modeling and patient stratification without compromising ethical standards or patient trust.

Discussion

We have outlined the rationale, design and analysis plan for the PRECOGNITION project, which aims to develop cognitive growth charting approaches for psychosis. The PRECOGNITION project will not generate any new primary data, but by leveraging existing data in novel ways, it will produce multiple distinct contributions including (i) guidance informed by lived experience as to functional outcomes that can inform this and future studies; (ii) novel methodological approaches for normative modelling of large scale cognitive data, including approaches for longitudinal tracking and prediction at the individual level and for combining data from different tasks that measure the same underlying construct (iii) machine learning models for stratifying cohorts on the basis of these models; (iv)

comprehensive reference models that provide broad sociodemographic coverage including different languages and distinct norms for individuals with psychosis and unaffected individuals. We intend that these models will supersede current norms currently used in the field[37] and also provide domain specific norms; (v) software tools to enable these that we will distribute via our online platform[48] and finally (vi) stratification models and predictions for functional outcome measures in psychosis cohorts.

We acknowledge several limitations to this project: whilst we anticipate that this project will generate models that have greater demographic and cross-language generalisability than existing norms, we do not anticipate that these will (yet) be universally applicable because we require some overlapping cognitive instruments between studies in order to apply the structured missingness and normative modelling techniques we will develop. Moreover, we will still be limited by the diversity of the samples for which we have data and we cannot expect that our models will generalise to completely new demographic contexts. In addition, it is likely that we will identify key functional outcomes via the focus groups for which we do not have sufficiently fine-grained data in all existing cohorts to target our prediction models towards. However, this study will nevertheless provide valuable information to guide the development of future studies.

In the longer term, if successful, the techniques and models developed during this project could be widely deployed, for example via mobile or tablet technology. This will provide a platform to maximise the value of cognitive instruments in predicting the course and functional outcome of psychosis.

References

1. Keefe RSE, Fenton WS. How Should DSM-V Criteria for Schizophrenia Include Cognitive Impairment? *Schizophr Bull*. 2007;33(4):912-920. doi:10.1093/schbul/sbm046

2. Mesholam-Gately RI, Giuliano AJ, Goff KP, Faraone SV, Seidman LJ. Neurocognition in first-episode schizophrenia: a meta-analytic review. *Neuropsychology*. 2009;23(3):315-336. doi:10.1037/a0014708

3. Lin A, Wood SJ, Nelson B, et al. Neurocognitive predictors of functional outcome two to 13 years after identification as ultra-high risk for psychosis. *Schizophr Res*. 2011;132(1):1-7. doi:10.1016/j.schres.2011.06.014

4. Orygen. Cognitive Impairment in Psychosis: Landscape Report.

5. Barch DM, Sheffield JM. Cognitive impairments in psychotic disorders: common mechanisms and measurement. *World Psychiatry*. 2014;13(3):224-232. doi:10.1002/wps.20145

6. Bowie CR, Reichenberg A, Patterson TL, Heaton RK, Harvey PD. Determinants of Real-World Functional Performance in Schizophrenia Subjects: Correlations With Cognition, Functional Capacity, and Symptoms. *Am J Psychiatry*. 2006;163(3):418-425. doi:10.1176/appi.ajp.163.3.418

7. Cicchetti D, Rogosch FA. Equifinality and multifinality in developmental psychopathology. *Dev Psychopathol*. 1996;8(4):597-600. doi:10.1017/S0954579400007318

8. Cannon TD, Keller MC. Endophenotypes in the genetic analyses of mental disorders. *Annu Rev Clin Psychol*. 2006;2:267-290. doi:10.1146/annurev.clinpsy.2.022305.095232

9. Insel T, Cuthbert B, Garvey M, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *Am J Psychiatry*. 2010;167(7):748-751. doi:10.1176/appi.ajp.2010.09091379

10. Insel TR, Cuthbert BN. Brain disorders? Precisely. *Science*. 2015;348(6234):499-500. doi:10.1126/science.aab2358

11. Schumann G, Binder EB, Holte A, et al. Stratified medicine for mental disorders. *Eur Neuropsychopharmacol*. 2014;24(1):5-50. doi:10.1016/j.euroneuro.2013.09.010

12. Foulkes L, Blakemore SJ. Studying individual differences in human adolescent brain development. *Nat Neurosci*. 2018;21(3):315-323. doi:10.1038/s41593-018-0078-4

13. Seghier ML, Price CJ. Interpreting and Utilising Intersubject Variability in Brain Function. *Trends Cogn Sci*. 2018;22(6):517-530. doi:10.1016/j.tics.2018.03.003

14. Carruthers SP, Van Rheenen TE, Gurvich C, Sumner PJ, Rossell SL. Characterising the structure of cognitive heterogeneity in schizophrenia spectrum disorders. A systematic review and narrative synthesis. *Neurosci Biobehav Rev*. 2019;107:252-278. doi:10.1016/j.neubiorev.2019.09.006

15. Farmer AE, McGuffin P, Spitznagel EL. Heterogeneity in schizophrenia: a cluster-analytic approach. *Psychiatry Res*. 1983;8(1):1-12. doi:10.1016/0165-1781(83)90132-4

16. Murray V, McKee I, Miller PM, et al. Dimensions and classes of psychosis in a population cohort: a four-class, four-dimension model of schizophrenia and affective psychoses. *Psychol Med*. 2005;35(4):499-510. doi:10.1017/s0033291704003745

17. Clementz BA, Sweeney JA, Hamm JP, et al. Identification of Distinct Psychosis Biotypes Using Brain-Based Biomarkers. *Am J Psychiatry*. 2016;173(4):373-384. doi:10.1176/appi.ajp.2015.14091200

18. Miskowiak KW, Kjærstad HL, Lemvigh CK, et al. Neurocognitive subgroups among newly diagnosed patients with schizophrenia spectrum or bipolar disorders: A hierarchical cluster analysis. *J Psychiatr Res*. 2023;163:278-287. doi:10.1016/j.jpsychires.2023.05.025

19. Kristensen TD, Mager FM, Ambrosen KS, et al. Cognitive profiles across the psychosis continuum. *Psychiatry Res*. 2024;342:116168. doi:10.1016/j.psychres.2024.116168

20. Ambrosen KS, Skjerbæk MW, Foldager J, et al. A machine-learning framework for robust and reliable prediction of short- and long-term treatment response in initially antipsychotic-naïve schizophrenia patients based on multimodal neuropsychiatric data. *Transl Psychiatry*. 2020;10(1):276. doi:10.1038/s41398-020-00962-8

21. Marquand AF, Wolfers T, Mennes M, Buitelaar J, Beckmann CF. Beyond Lumping and Splitting: A Review of Computational Approaches for Stratifying Psychiatric Disorders. *Biol Psychiatry Cogn Neurosci Neuroimaging*. 2016;1(5):433-447. doi:10.1016/j.bpsc.2016.04.002

22. Liu Y, Hayes DN, Nobel A, Marron JS. Statistical Significance of Clustering for High-Dimension, Low–Sample Size Data. *J Am Stat Assoc*. 2008;103(483):1281-1293. doi:10.1198/016214508000000454

23. Haatveit B, Westlye LT, Vaskinn A, et al. Intra- and inter-individual cognitive variability in schizophrenia and bipolar spectrum disorder: an investigation across multiple cognitive domains. *Schizophrenia*. 2023;9(1):89. doi:10.1038/s41537-023-00414-4

24. Kjelkenes R, Wolfers T, Alnæs D, et al. Mapping Normative Trajectories of Cognitive Function and Its Relation to Psychopathology Symptoms and Genetic Risk in Youth. *Biol Psychiatry Glob Open Sci*. Published online February 2022:S266717432200012X. doi:10.1016/j.bpsgos.2022.01.007

25. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. *Mol Psychiatry*. 2019;24(10):1415-1424. doi:10.1038/s41380-019-0441-1

26. Wolfers T, Doan NT, Kaufmann T, et al. Mapping the Heterogeneous Phenotype of Schizophrenia and Bipolar Disorder Using Normative Models. *JAMA Psychiatry*. 2018;75(11):1146-1155. doi:10.1001/jamapsychiatry.2018.2467

27. Wolfers T, Rokicki J, Alnaes D, et al. Replicating extensive brain structural heterogeneity in individuals with schizophrenia and bipolar disorder. *Hum Brain Mapp*. 2021;42(8):2546-2555. doi:10.1002/hbm.25386

28. Lv J, Di Biase M, Cash RFH, et al. Individual deviations from normative models of brain structure in a large cross-sectional schizophrenia cohort. *Mol Psychiatry*. 2021;26(7):3512-3523. doi:10.1038/s41380-020-00882-5

29. Worker A, Berthert P, Lawrence AJ, et al. Extreme deviations from the normative model reveal cortical heterogeneity and associations with negative symptom severity in first-episode psychosis from the OPTiMiSE and GAP studies. *Transl Psychiatry*. 2023;13(1):373. doi:10.1038/s41398-023-02661-6

30. Berthet P, Haatveit BC, Kjelkenes R, et al. A 10-Year Longitudinal Study of Brain Cortical Thickness in People with First-Episode Psychosis Using Normative Models. *Schizophr Bull*. Published online July 5, 2024:sbae107. doi:10.1093/schbul/sbae107

31. Fusar-Poli P, Estradé A, Stanghellini G, et al. The lived experience of psychosis: a bottom-up review co-written by experts by experience and academics. *World Psychiatry*. 2022;21(2):168-188. doi:10.1002/wps.20959

32. Neidhart M, Kjelkenes R, Jansone K, et al. A protocol for data harmonization in large cohorts. *Nat Ment Health*. 2024;2(10):1134-1137. doi:10.1038/s44220-024-00315-0

33. Mitra R, McGough SF, Chakraborti T, et al. Learning from data with structured missingness. *Nat Mach Intell*. 2023;5(1):13-23. doi:10.1038/s42256-022-00596-z

34. Kia SM, Huijsdens H, Rutherford S, et al. Closing the life-cycle of normative modeling using federated hierarchical Bayesian regression. *PloS One*. 2022;17(12):e0278776. doi:10.1371/journal.pone.0278776

35. De Boer AAA, Bayer JMM, Kia SM, et al. Non-Gaussian Normative Modelling With Hierarchical Bayesian Regression. *Imaging Neurosci*. Published online March 29, 2024. doi:10.1162/imag_a_00132

36. Heaton RK, Avitable N, Grant I, Matthews CG. Further Crossvalidation of Regression-Based Neuropsychological Norms with an Update for the Boston Naming Test. *J Clin Exp Neuropsychol*. 1999;21(4):572-582. doi:10.1076/jcen.21.4.572.882

37. Kern RS, Nuechterlein KH, Green MF, et al. The MATRICS Consensus Cognitive Battery, Part 2: Co-Norming and Standardization. *Am J Psychiatry*. 2008;165(2):214-220. doi:10.1176/appi.ajp.2007.07010043

38. Borghi E, De Onis M, Garza C, et al. Construction of the World Health Organization child growth standards: selection of methods for attained growth curves. *Stat Med*. 2006;25(2):247-265. doi:10.1002/sim.2227

39. Cole TJ. Growth charts for both cross-sectional and longitudinal data. *Stat Med*. 1994;13(23-24):2477-2492. doi:10.1002/sim.4780132311

40. Rehák Bučková B, Fraza C, Rehák R, et al. Using normative models pre-trained on cross-sectional data to evaluate longitudinal changes in neuroimaging data. Published online June 9, 2023. doi:10.1101/2023.06.09.544217

41. Cole TJ. 3-in-1 weight-monitoring chart. *The Lancet*. 1997;349(9045):102-103. doi:10.1016/S0140-6736(05)60886-0

42. Van Buuren S. Evaluation and prediction of individual growth trajectories. *Ann Hum Biol*. 2023;50(1):247-257. doi:10.1080/03014460.2023.2190619

43. Holz NE, Floris DL, Llera A, et al. Age-related brain deviations and aggression. *Psychol Med*. Published online April 22, 2022:1-10. doi:10.1017/S003329172200068X

44. Wolfers T, Arenas AL, Onnink AMH, et al. Refinement by integration: aggregated effects of multimodal imaging markers on adult ADHD. *J Psychiatry Neurosci*. 2017;42(6):386-394. doi:10.1503/jpn.160240

45. Doan NT, Kaufmann T, Bettella F, et al. Distinct multivariate brain morphological patterns and their added predictive value with cognitive and polygenic risk scores in mental disorders. *NeuroImage Clin*. 2017;15:719-731. doi:10.1016/j.nicl.2017.06.014

46. Rutherford S, Fraza C, Dinga R, et al. Charting brain growth and aging at high spatial precision. *eLife*. 2022;11:e72904. doi:10.7554/eLife.72904

47. Rutherford S, Barkema P, Tso IF, et al. Evidence for embracing normative modeling. *eLife*. 2023;12:e85082. doi:10.7554/eLife.85082

48. Barkema P, Rutherford S, Lee HC, et al. Predictive Clinical Neuroscience Portal (PCNportal): instant online access to research-grade normative models for clinical neuroscientists. *Wellcome Open Res*. 2023;8:326. doi:10.12688/wellcomeopenres.19591.2

49. Calkins ME, Merikangas KR, Moore TM, et al. The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J Child Psychol Psychiatry*. 2015;56(12):1356-1369. doi:10.1111/jcpp.12416

50. Casey BJ, Cannonier T, Conley MI, et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci*. 2018;32:43-54. doi:10.1016/j.dcn.2018.03.001

51. Volkow ND, Koob GF, Croyle RT, et al. The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev Cogn Neurosci*. 2018;32:4-7. doi:10.1016/j.dcn.2017.10.002

52. Magnus P, Irgens LM, Haug K, Nystad W, Skjærven R, Stoltenberg C. Cohort profile: The Norwegian Mother and Child Cohort Study (MoBa). *Int J Epidemiol*. 2006;35(5):1146-1150. doi:10.1093/ije/dyl170

53. Nuechterlein KH, Green MF, Kern RS, et al. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. *Am J Psychiatry*. 2008;165(2):203-213. doi:10.1176/appi.ajp.2007.07010042

54.   Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013;42(1):97-110. doi:10.1093/ije/dys066

55.   Barch DM, Burgess GC, Harms MP, et al. Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*. 2013;80:169-189. doi:10.1016/j.neuroimage.2013.05.033

56.   Hintz F, Dijkhuis M, van 't Hoff V, McQueen JM, Meyer AS. A behavioural dataset for studying individual differences in language skills. *Sci Data*. 2020;7(1):429. doi:10.1038/s41597-020-00758-x

57.   Healthy Brain Study consortium, Aarts E, Akkerman A, et al. Protocol of the Healthy Brain Study: An accessible resource for understanding the human brain and how it dynamically and individually operates in its bio-social context. D'Acquisto F, ed. *PLOS ONE*. 2021;16(12):e0260952. doi:10.1371/journal.pone.0260952

58.   Shafto MA, Tyler LK, Dixon M, et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) study protocol: a cross-sectional, lifespan, multidisciplinary examination of healthy cognitive ageing. *BMC Neurol*. 2014;14:204. doi:10.1186/s12883-014-0204-1

59.   Voldsbekk I, Kjelkenes R, Frogner ER, Westlye LT, Alnaes D. Testing the sensitivity of diagnosis-derived patterns in functional brain networks to symptom burden in a Norwegian youth sample. *Hum Brain Mapp*. 2024;45(3):e26631. doi:10.1002/hbm.26631

60.   Moskal V, Voldsbekk I, Leonardsen EH, Holthe LCF, Kjelkenes R, Westlye LT. Cognitive age prediction and psychological distress in adolescence. Published online May 21, 2024. doi:10.31234/osf.io/egwtu

61.   Millgate E, Kravariti E, Egerton A, et al. Cross-sectional study comparing cognitive function in treatment responsive versus treatment non-responsive schizophrenia: evidence from the STRATA study. *BMJ Open*. 2021;11(11):e054160. doi:10.1136/bmjopen-2021-054160

62.   Morgan C, Lappin J, Heslin M, et al. Reappraising the long-term course and outcome of psychotic disorders: the AESOP-10 study. *Psychol Med*. 2014;44(13):2713-2726. doi:10.1017/S0033291714000282

63.   Murray RM, Mondelli V, Stilo SA, et al. The influence of risk factors on the onset and outcome of psychosis: What we learned from the GAP study. *Schizophr Res*. 2020;225:63-68. doi:10.1016/j.schres.2020.01.011

64.   Fagerlund B, Pantelis C, Jepsen JRM, et al. Differential effects of age at illness onset on verbal memory functions in antipsychotic-naïve schizophrenia patients aged 12-43 years. *Psychol Med*. 2021;51(9):1570-1580. doi:10.1017/S0033291720000409

65.   Klærke LR, Baandrup L, Fagerlund B, et al. Diagnostic stability and long-term symptomatic and functional outcomes in first-episode antipsychotic-naïve patients with schizophrenia. *Eur Psychiatry J Assoc Eur Psychiatr*. 2019;62:130-137. doi:10.1016/j.eurpsy.2019.07.001

66. Nielsen MØ, Kristensen TD, Borup Bojesen K, Glenthøj BY, Lemvigh CK, Ebdrup BH. Differential Effects of Aripiprazole and Amisulpride on Negative and Cognitive Symptoms in Patients With First-Episode Psychoses. *Front Psychiatry*. 2022;13:834333. doi:10.3389/fpsyt.2022.834333

67. Glenthøj LB, Mariegaard LS, Fagerlund B, et al. Cognitive remediation plus standard treatment versus standard treatment alone for individuals at ultra-high risk of developing psychosis: Results of the FOCUS randomised clinical trial. *Schizophr Res*. 2020;224:151-158. doi:10.1016/j.schres.2020.08.016

68. Wechsler D. Wechsler Adult Intelligence Scale--Third Edition. Published online February 11, 2019. doi:10.1037/t49755-000

69. Keefe RSE, Goldberg TE, Harvey PD, Gold JM, Poe MP, Coughenour L. The Brief Assessment of Cognition in Schizophrenia: reliability, sensitivity, and comparison with a standard neurocognitive battery. *Schizophr Res*. 2004;68(2-3):283-297. doi:10.1016/j.schres.2003.09.011

70. Nelson, Hazel E. *National Adult Reading Test (NART)*. Nfer-Nelson Windsor; 1991.

71. Delis DC, Kaplan E, Kramer JH. Delis-Kaplan Executive Function System. Published online October 8, 2012. doi:10.1037/t15082-000

72. Robbins TW, James M, Owen AM, Sahakian BJ, McInnes L, Rabbitt P. Cambridge Neuropsychological Test Automated Battery (CANTAB): a factor analytic study of a large sample of normal elderly volunteers. *Dement Basel Switz*. 1994;5(5):266-281. doi:10.1159/000106735

73. Gershon RC, Wagster MV, Hendrie HC, Fox NA, Cook KF, Nowinski CJ. NIH Toolbox for Assessment of Neurological and Behavioral Function. *Neurology*. 2013;80(Issue 11, Supplement 3):S2-S6. doi:10.1212/WNL.0b013e3182872e5f

74. Richard G, Kolskår K, Sanders AM, et al. Assessing distinct patterns of cognitive aging using tissue-specific brain age prediction based on diffusion tensor imaging and brain morphometry. *PeerJ*. 2018;6:e5908. doi:10.7717/peerj.5908

75. Askeland RB, Hannigan LJ, Ask H, et al. Early manifestations of genetic risk for neurodevelopmental disorders. *J Child Psychol Psychiatry*. 2022;63(7):810-819. doi:10.1111/jcpp.13528

76. Kia SM, Huijsdens H, Dinga R, et al. Hierarchical Bayesian Regression for Multi-Site Normative Modeling of Neuroimaging Data. *ArXiv200512055 Cs Stat*. Published online May 25, 2020. Accessed August 2, 2021. http://arxiv.org/abs/2005.12055

77. Ruiz FJR, Valera I, Blanco C, Perez-Cruz F. Bayesian nonparametric comorbidity analysis of psychiatric disorders. Published online 2014. doi:10.48550/ARXIV.1401.7620

78. Delis DC, Kramer JH, Kaplan E, Ober BA. California Verbal Learning Test--Second Edition. Published online November 14, 2016. doi:10.1037/t15072-000

79. Benedict RHB, Schretlen D, Groninger L, Brandt J. Hopkins Verbal Learning Test--Revised. Published online June 10, 2013. doi:10.1037/t10851-000

80.   Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. 2006;3(2):77-101. doi:10.1191/1478088706qp063oa

81.   Diamond R, Waite F, Boylan AM, Hicks A, Kabir T, Freeman D. Framework for understanding movement and physical activity in patients diagnosed with psychosis. *BMJ Ment Health*. 2024;27(1):e300878. doi:10.1136/bmjment-2023-300878

82.   Mahmood R, Wallace V, Wiles N, Kessler D, Button KS, Fairchild G. The lived experience of withdrawal from Selective Serotonin Reuptake Inhibitor (SSRI) antidepressants: A qualitative interview study. *Health Expect Int J Public Particip Health Care Health Policy*. 2024;27(1):e13966. doi:10.1111/hex.13966

83.   Clarkson C, Scott HR, Hegarty S, et al. "You get looked at like you're failing": A reflexive thematic analysis of experiences of mental health and wellbeing support for NHS staff. *J Health Psychol*. 2023;28(9):818-831. doi:10.1177/13591053221140255

84.   Byrne D. A worked example of Braun and Clarke's approach to reflexive thematic analysis. *Qual Quant*. 2022;56(3):1391-1412. doi:10.1007/s11135-021-01182-y

85.   Rutherford S, Wolfers T, Fraza C, et al. To which reference class do you belong? Measuring racial fairness of reference classes with normative modeling. Published online 2024. doi:10.48550/ARXIV.2407.19114

86.   MacIntyre MR, Cockerill RG, Mirza OF, Appel JM. Ethical considerations for the use of artificial intelligence in medical decision-making capacity assessments. *Psychiatry Res*. 2023;328:115466. doi:10.1016/j.psychres.2023.115466
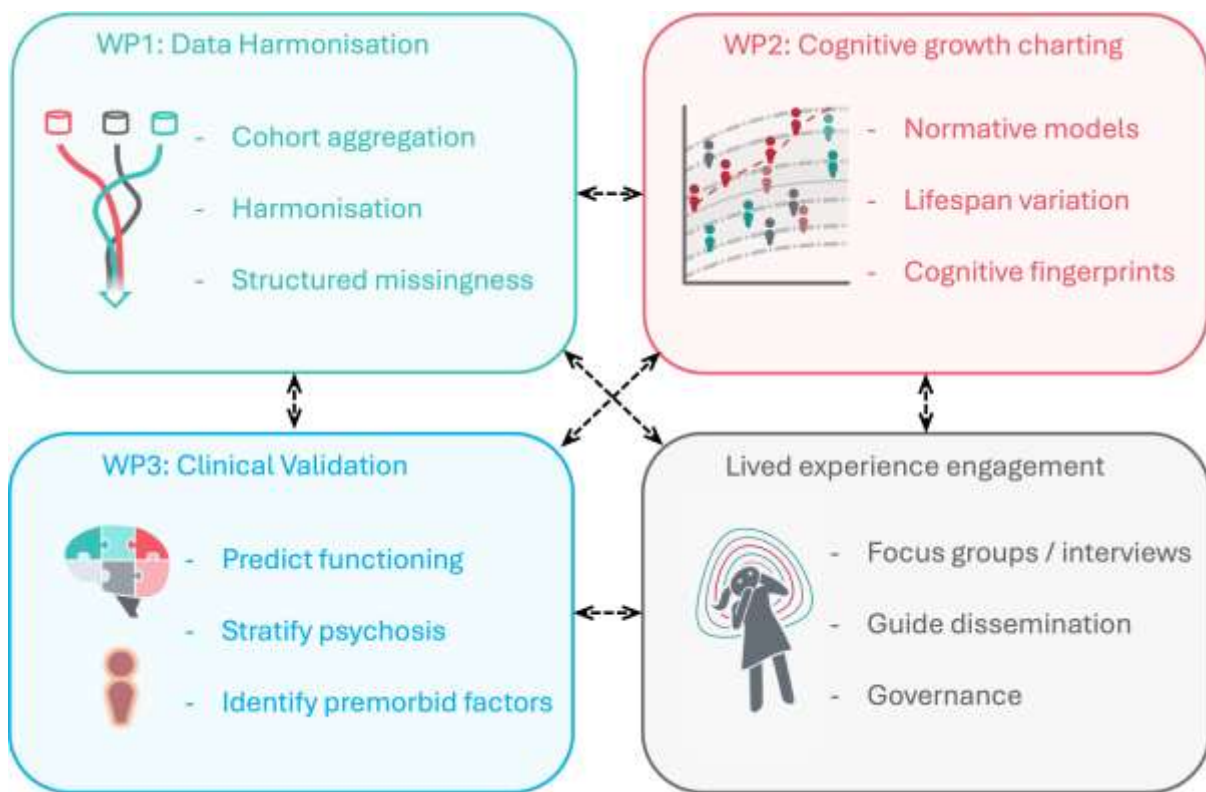
Figure Legends:

Figure 1: The PRECOGNITION project is structured into three mutually interacting workpackages (WPs), with lived experience engagement throughout the project. Key themes for each WP are identified.

Figure 2: A high-level overview of our analytical workflow including pre-statistical harmonisation (A), imputation (B), followed by fitting normative models (C). In panel A, two measures are shown that are equivalent, but derived from different tests and having slight variations. These can be accommodated by simple mathematical operations such as rescalings. In panel B, we show a graphical representation of data derived from the thematically organised pyschosis (TOP) cohort, which contains two distinct cognitive batteries applied to different subjects, inducing a strong pattern of structured missingness in the data, requiring custom-built imputation techniques to complete (right). In panel C, we show an example of an identical measure derived from the NIH toolbox for multiple studies from the Human Connectome Project lifespan datasets. In this case cohort effects and non-linearity across the lifespan can be accommodated in the normative modelling step.

Figure 3: Illustrative data assembled from some of the cohorts used in the PRECOGNITION project. Left panel: Delayed recall scores from the California Verbal Learning Test and Hopkins Verbal Learning Test, derived from the different Thematically Organised Psychosis (TOP) batteries. Although the tests are nearly identical, they have different psychometric properties across the lifespan. Middle panel: MATRICS 'Mazes' score (assessing planning and foresight) used in the second TOP battery shows clear heteroskedasticity across the lifespan and does not have an equivalent in the first TOP battery. This means it must be reconstructed using imputation. Right panel: Reading age score from the Human Connectome project (HCP) lifespan data. This shows a non-linear relationship with age and strong cohort effects, which must be accommodated during the modelling.
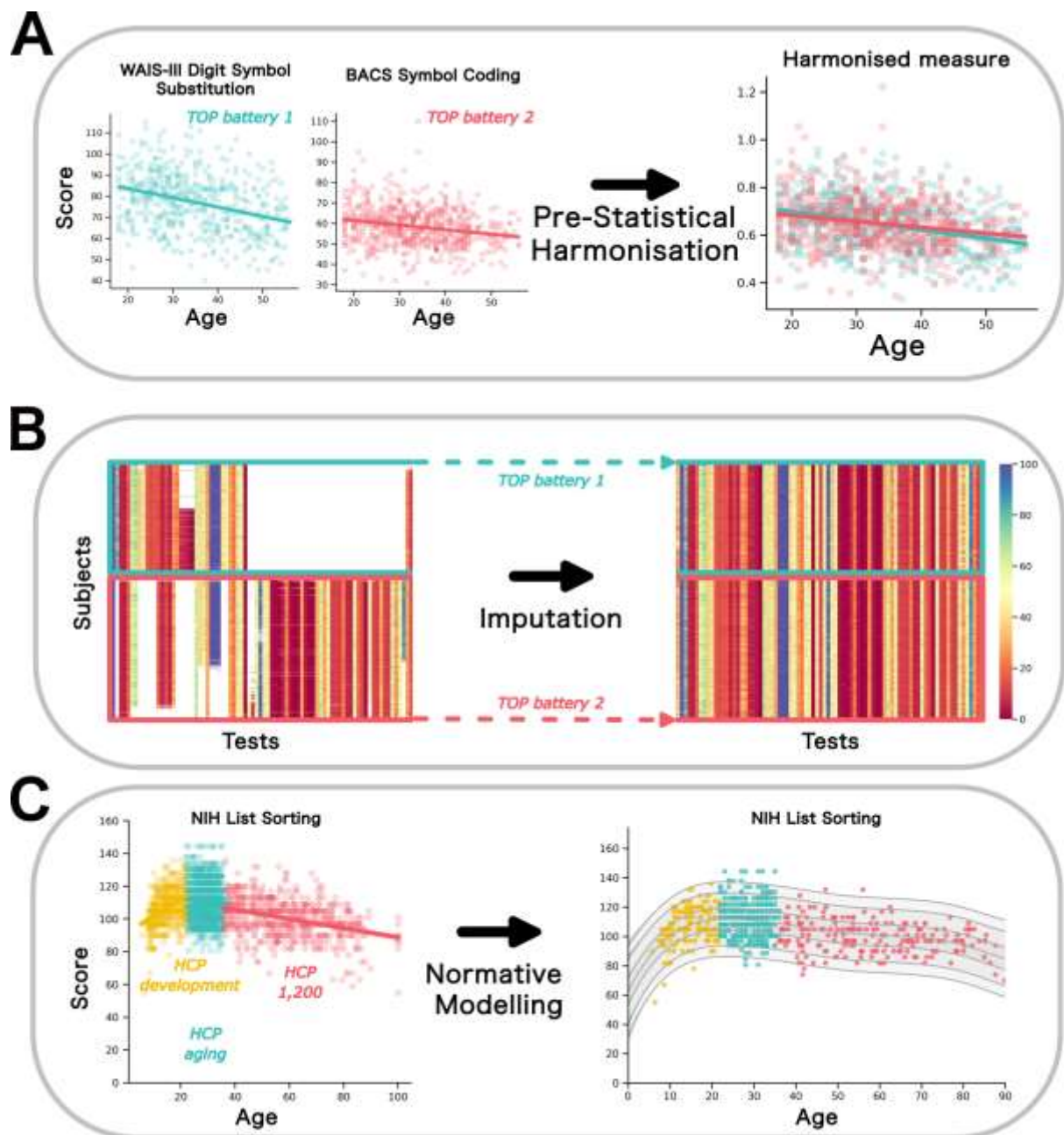
Figure 1

Figure 2

Figure 3