Digital behavioural tests as diagnostic aid for psychosis

Piotr Słowiński[1], Alexander White[2], Sian Lison[3], Sarah Sullivan[4], Tobit Emmens[3], Philip Self[3], Jane Wileman[5], Anke Karl[2], Krasimira Tsaneva-Atanasova[1]

[1] Translational Research Exchange @ Exeter, Living Systems Institute, Department of Mathematics and Statistics, Faculty of Environment, Science and Economy, University of Exeter, UK
[2] Department of Psychology, Faculty of Health and Life Sciences, University of Exeter, UK
[3] Research & Development Department, Devon Partnership NHS Trust, Exeter, UK
[4] Faculty of Health Sciences, Bristol Medical School, University of Bristol, UK
[5] Specialist Team for Early Psychosis, Devon Partnership NHS Trust, Exeter, UK

Corresponding author: Piotr Słowiński, p.m.slowinski@exeter.ac.uk

Abstract:

Background:
Timely interventions have a proven benefit for people experiencing psychotic illness. One bottleneck to accessing timely interventions is the referral process to the specialist team for early psychosis (STEP). Many general practitioners lack awareness or confidence in recognising psychotic symptoms or state. Additionally, referrals for people without apparent psychotic symptoms, although beneficial at a population level, lead to excessive workload for STEPs. There is a clear unmet need for accurate stratification of STEPs users and healthy cohorts that this study aimed to investigate by applying digital behavioural tests.

Methods:
To discriminate between the STEPs users (SU; n=32) and controls (n=32, age and sex matched), we employed naïve Bayes classifier, and applied it to objective, quantitative and interpretable features of a 'mirror game' (MG) and trail making task (TMT). The MG is a movement coordination task shown to be a potential socio-motor biomarker of schizophrenia, while TMT is a neuropsychiatric test of cognitive function.

Findings:
The proposed classifier shows an excellent performance, AUC = 0.92 (95%CI 0.75-1), Sensitivity = 0.88 (95%CI 0.62-1), Specificity = 1 (95%CI 0.75-1), evaluated on 25% hold-out and 1000 folds. The study demonstrates that cheap off-the-shelf equipment (laptop computer and a leap motion sensor) records clinically relevant behavioural data sufficiently. We also find that MG and TMT are unsuitable in isolation to successfully differentiate between SU with and without at-risk-mental-state or first episode psychosis with sufficient level of performance.

Interpretation:
Including digital behavioural tests into healthcare practice could allow improvements in care for people at risk of developing psychotic illness. Our findings show that introduction of standardised battery of digital behavioural tests would benefit both clinical and research practice. It would allow standardization of referrals, while the high specificity of digital behavioural tests would benefit research on prognostic instruments for psychosis by enriching and homogenising clinical high-risk populations. Here we demonstrate that digital behavioural tests can be successfully used and could help to address this need.

Funding:
EPSRC Impact Acceleration Account, Impact & Knowledge Exchange Award, Jean Golding Institute seed corn, Avon & Wiltshire Mental Health Partnership NHS Trust Research Capability Funding.
PS was generously supported by the Wellcome Trust Institutional Strategic Support Award 204909/Z/16/Z. KTA gratefully acknowledges the financial support of the EPSRC via grant EP/T017856/1.

RESEARCH IN CONTEXT

Evidence before this study:
We searched PubMed (without any date or language restrictions) for articles published up to 11 January 2023 that contained following combinations of terms: (psychosis OR schizophrenia) AND (motor coordination) 396 publications; (psychosis OR schizophrenia) AND (trail making task) 93; (psychosis OR schizophrenia) AND (mirror game) 3 publications; (psychosis OR schizophrenia) AND (digital tests battery) 13. The large number of publications related to the first search is due to well-known association of motor abnormalities (that include deficits in motor coordination) with psychosis and schizophrenia. Similarly, trail making task (TMT) is a well-established instrument for assessment of cognitive function, with performance worsened in psychosis and schizophrenia. Of interest are two recent publications, a meta-analysis showing that performance in TMT differs in clinical high-risk population between people who did and did not transition to psychosis and a longitudinal study showing that deficits in motor coordination could help predict the long-term severity of psychosis. The search including mirror game (MG) returned our own work demonstrating that MG could be used as a base for socio-motor biomarkers in schizophrenia; proposed socio-motor biomarkers include measures based on motor coordination. Search for digital test returned two studies of interest on using digital/ web-based test to assess cognitive function in schizophrenia. One study showed that some digital/ web-based test are affected by practice effects. The other showed that web-based battery of tests had low performance in terms of discriminating between schizophrenia patients and healthy controls (AUROC>0.70).

Added value of this study:
This study demonstrates that combining digital tests with data analytics has potential for simplifying neuropsychiatric assessment. It shows that using measurements from TMT and MG allows to differentiate between people accepted for assessment in specialist team for early psychosis (STEP) and controls with excellent performance (AUROC>0.9), while achieving 100% specificity (no false positive detections). The study shows feasibility of using cheap, portable equipment, assembled from off-the-shelf components, for collection of clinically relevant data that could be used to inform clinical decision making.

Implications of all the available evidence:
Our study, with its state-of-the-art performance and interpretable results, demonstrate high clinical potential of implementing digital batteries of behavioural tests in clinical practice. Future work should focus on identification of optimal set of tests for establishing standardised digital batteries of behavioural tests as well as on longitudinal studies to understand how socio-motor function changes in the course of psychosis. Such developments would not only help to stratify STEPs users but would facilitate rapid assessment for all people seeking care in mental health early intervention services improving their quality of life and wellbeing.

INTRODUCTION

Psychosis is a severe mental illness characterised by loss of contact with reality and symptoms such as hallucinations, delusions and thought disorders. It can be one of the first symptoms of a range of serious and long-term mental disorders such as schizophrenia, affective and other psychoses. Developing psychosis in young adulthood is devastating and often disrupts the trajectory into healthy and independent adulthood; the mean age of onset is 22 years for men and 26 years for women.[1] People with serious mental illness die 15 to 20 years earlier than the general population.[2]

Serious mental disorders are extremely expensive to treat, with presence of psychotic or affective symptoms being one of patient characteristics driving increase in hospital costs.[3] Their direct healthcare toll to NHS England have been estimated at £2.82 billion annually in 2019.[3] While the most recent estimate puts the overall annual economic impact of schizophrenia and psychosis in England at £11.8 billion.[4] The higher overall economic impact includes reduced labour supply, premature mortality, reduced health-related quality of life, lost output, lost tax revenue, transfer payments, and unpaid care by family or friends.

Most risk factors for a poor outcome, such as gender or low socio-economic status, are difficult or impossible to alter. But people with psychosis have better outcomes if they are treated as soon as possible after their first symptom.[5] Early interventions can reduce the rate of relapse, risk of suicide, and number of hospital admissions.[1,6] They also significantly improve quality of life by enabling people to finish education and develop supportive networks outside the family of origin.[1]

Early interventions are typically delivered by a specialist team for early psychosis (STEP).[7–12] However the referral process to STEPs is far from being optimal. Although most STEP referrals are from primary care,14/01/2023 14:13:00 many general practitioners lack awareness of high-risk symptoms or are not confident with recognising the psychotic state, both of which could lead to young people not receiving the care they need.[13,14] On the other hand, although increasing the number of referrals has been shown to be beneficial at the population level, it also leads to increases in STEPs workload thus contributing to the pressure on the care system. The extra work is caused by higher number of assessments requested as well as a need for an increased engagement (including dedicated liaison practitioners) with primary care providers to identify and refer people experiencing, or at risk of, psychotic illness.[9,15]

Here, we investigate if digital behavioural tests can be used as an effective tool that allows differentiation between people referred to STEPs and the general population, and if they show potential to facilitate and standardise the referral process. Specifically, we use data from a digital version of the trail making task (TMT), a standard method for assessment of cognitive function,[16,17] and the mirror game (MG), a novel way of assessing socio-motor functioning (motor coordination and interpersonal synchronisation).[18,19] Our choice of the tasks is based on a significant body of research showing that assessment of movement, behaviour and cognitive function allows to accurately differentiate between people with schizophrenia and general population.[18,20–31] In particular, motor and executive functions[18,30] as well as eye movements[24,27] were shown to hold promising diagnostic potential. In addition, deficits in motor coordination were recently shown to be markers of long-term clinical outcomes,[31] while performance in TMT was shown to differ in people at clinical high-risk for psychosis who transitioned from those who did not transition to psychosis.[32] Finally, we test feasibility of using cheap off-the-shelf components (laptop computer with a plug-in sensor) for simplifying neuropsychiatric assessment and introducing standardised digital tests to clinical practice.[20]

METHODS

Study design and participants

The study was designed as a prospective, cross-sectional feasibility study in a group of service users accepted for an assessment for psychosis including people with first episode psychosis or assessed as being at risk of developing psychosis. Control cohort was recruited independently at the University of Exeter (UoE). Demographic and clinical characteristics of participants can be found in Table 1.

Service users (SU) were identified and recruited by Devon Partnership NHS Trust (DPT) and Avon

and Wiltshire Mental Health Partnership NHS Trust (AWP). In total we recruited 32 participants, all of which were included in the analysis (we do not have data about number of screen participants, none of the participants dropped-out). The inclusion criteria were being accepted for an assessment for psychosis or risk of developing psychosis by a consultant psychiatrist or a trained specialist with experience in at-risk mental states. The exclusion criteria were: (1) Lacking capacity to provide informed consent for inclusion. The clinical team had an opportunity to assess the mental capacity at the CAARMS (comprehensive assessment of at-risk mental state (ARMS)[33] appointment before the potential participant was approached to request consent to contact. (2) Insufficient understanding of English to follow the test instructions. (3) Any suspected organic cause of psychosis (i.e., head injury, epilepsy or dementia). (4) Taking antipsychotic medication for longer than 4 months before the start of the study. Each participant was offered £10 for participating and if necessary, reimbursement for reasonable travel expenses (after producing a receipt). SU were recruited between 19/07/2018 – 23/05/2019. The study was reviewed by Research Ethics Committee (REC) and received approval from Health Research Authority (HRA) and Health and Care Research Wales (HCRW); project ID 236262, protocol number 1718/26, REC reference 18/SW/0065.

Control cohort (CC) was identified at UoE. In total we recruited 86 participants, of which 43 played the same version of MG as SU and were used to identify the n=32 CC matching by age and gender the SU as close as possible. Participants were volunteers recruited by personally approaching potential participants, putting posters around UoE campus and at Exeter's community centres, social media adverts, and snowball sampling. The exclusion criteria were: (1) Moderate, or more severe, symptoms of depression, assessed by means of Patient Health Questionnaire-9 (PHQ-9).[34] For ethical reasons we excluded the question concerning thoughts of suicide and self-harm. All participants scoring above 9, indicating at least moderate levels of depression, were signposted to several sources of support. (2) A diagnosis of depression, an anxiety disorder or schizophrenia. (3) Taking any psychopharmacological medication. Participants who indicated that they were having difficulties with mental health were directed to the UoE wellbeing centre. (4) Suffering from seizures. (5) English not being one of their first languages. This criterion was introduced to try to minimise the chances of misinterpretations due to the extensive use of questionnaires in the study. CC was additionally screened using Community Assessment of Psychic Experiences-42 (CAPE-42);[35] question 14, which asks about suicidal ideation and loads onto the depressive subscale, was excluded for ethical reasons. Each participant was offered £5 or one course credit for participating. CC was recruited by AW between 25/05/2018 – 26/11/2018 as part of his Master's degree project. The recruitment of the CC was approved by University of Exeter, College of Life and Environmental Sciences (CLES), Psychology Ethics Committee, eCLESPsy000568 v2.1.

Participant flow chart is presented in Fig. 1. All participants gave written informed consent prior to the study.

Table 1 Demographic and clinical characteristics of participants.

	Service users (N=32)		**Controls cohort (N=32)**		**Statistics**
	Mean	Min-Max	Mean	Min-Max	
Age (years)	28.8	18-58	24.2	18-54	U=394.5, p=0.11
Sex (male/female)	16/16		12/20		Chi^2=1.01, p=0.31 (Pearson)
CAARMS (score, number of participants)		0, n=16; 2, n=4; 4, n=12			
CAPE-42			1.41	0.85-2.36	
Anti-psychotic medication	15 participants, < 4 months				

Note: U – Mann-Whitney statistic, Chi^2 – Chi-squared statistic, CAARMS – score of the comprehensive assessment of at risk mental state, CAPE-42 – Community Assessment of Psychic Experiences-42.

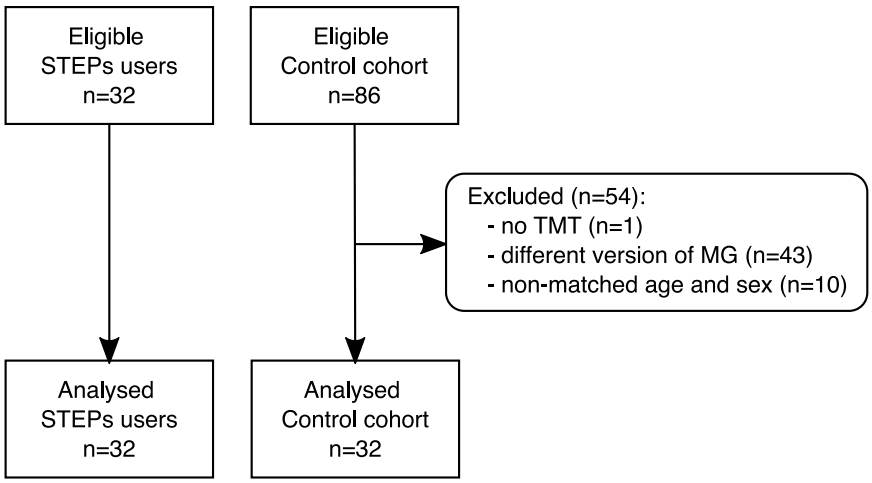

Figure 1 Participant flow chart.

Mirror Game

The mirror game (MG) used in this study was based on the algorithm described by Zhai and colleagues[36] and followed closely our earlier work on establishing socio-motor markers of schizophrenia.[18] MG is a movement task that can be used to assess socio-motor functioning (motor coordination and interpersonal synchronisation).[18,19] We used two MG tasks. The first task was a Solo game, where participants were asked to move their hand freely in a horizontal direction. Participants were given the following instruction: "Please move your hand left and right, create an interesting motion and enjoy playing." The second task was a Leader-Follower game. In the second task an animated image of a robot appeared on the screen. The animation showed the robot controlling its own dot. The dot moved horizontally according to a pre-generated movement pattern. Participants were tasked with following the dot's movement as closely as possible whilst it was on screen. Participants were given the following instruction: "Please try to follow the movement of an animated robot as accurately as you can." During the Leader-Follower game the robot was also presenting a parametric positive social feedback (smiling) as described by Cohen and colleagues.[37]

The two tasks were grouped into one session. The session consisted of the Solo game, three repetitions of the Leader-Follower game and another Solo game. Each game lasted for one minute. The session was repeated three times. Participants were free to take breaks between the games and sessions. Each Leader-Follower game used a different pre-generated movement pattern. Patterns were the same for each participant. Only the last two sessions were included in the analysis. We excluded the 1st session to allow participants to get familiar with the task. The SU participants were sitting in front of a 17'' diagonal laptop computer (1100x680 pixels screen resolution), the CC was using a 23'' diagonal computer monitor; the image displayed on the computer monitor was scaled down to use the central 17'' diagonal part of the monitor and have the same resolution as the laptop display. Movement of the hand was recorded using a leap motion sensor[38] and displayed as a dot on the screen. Participants used their dominant hand to control the horizontal position of a dot on the screen. The computer set-ups were different in the two groups due to the need of collecting the data simultaneously in multiple locations and additional research goals for the experiments with the CC that are not a part of the presented analysis.

For analysis we used the recorded position of the participant hand (Solo and Leader-Follower) and the trajectory of the movement generated by the computer (Leader-Follower). Recorded position data is in arbitrary units in the range [-0.5, 0.5]; variable sampling rate, 90-140 Hz (S) and 40-70Hz (Leader-Follower). Pre-processing included:

- resampling to 100Hz with linear interpolation,
- low pass filtering with 5 Hz cut-off done using phase preserving Butterworth filter of degree 2,
- omitting the first and last 5s of the recording,
- estimation of movement velocity, using a fourth-order finite difference scheme.

Trail-making task

The trail-making task (TMT)[16,39] is a valid, public domain test of visual attention and working memory, as well as executive control.[40] It has two parts, which were alternated. In each part, participants must click on 25 dots in a specified order as quickly and accurately as possible. The visual attention part (TMT A) had participants click numbers in ascending order, 1-25. The executive control part (TMT B) had participants alternate between clicking on numbers and letters, both in ascending order (1-A-2-B-3-C etc.). Participants completed each part three times, alternating between TMT A and B, starting with TMT A. Only the last two repetitions were included in the analysis. We excluded the 1st repetition to allow participants to get familiar with the task. We used a digital version of the task implemented in PEBL: The Psychology Experiment Building Language.[41] In the original study protocol the TMT was used as a non-diagnostic attention measuring task. It was retrospectively included in the analysis after literature review[37,42–44] and data analysis indicated that including participants' performance in this task could be beneficial for differentiating between the SU and CC.

For analysis we used the times between each individual mouse click made by the participant, we also used times between mouse clicks made on the correct targets.

Testing procedure

The stages of the research session are presented in Table 2. Both tasks and examples of collected data are shown in Fig. 2.

Table 2 Stages of the research session

Stage	Description
1.	Time for questions and signing informed consent.
1a.	CC additionally answered a questionnaire (state/ moment PANAS)[45] and the finger-tapping test,[46] not a part of the presented analysis.
2.	1st MG session (excluded from analysis): Solo x 1 – 1 minute of participant's own movement; Leader-Follower x 3 – three repetitions of the participant following the avatar (1 minute each); Solo x 1.
3.	Break (at least 1 minute).
4.	2nd MG session: Solo x 1; Leader-Follower x 3; Solo x 1.
5.	Break (at least 1 minute).
6.	3rd MG session: Solo x 1; Leader-Follower x 3; Solo x 1.
7.	Break (at least 1 minute).
7a.	CC additionally answered a questionnaire (state/ moment PANAS), not a part of the presented analysis.
8.	TMT: 3 x TMT A and B. (1st repetition excluded from analysis.)
8a.	SU and CC were asked to answer a MG acceptability questionnaire, not a part of the presented analysis.
8b.	CC continued with remaining part of the research session.

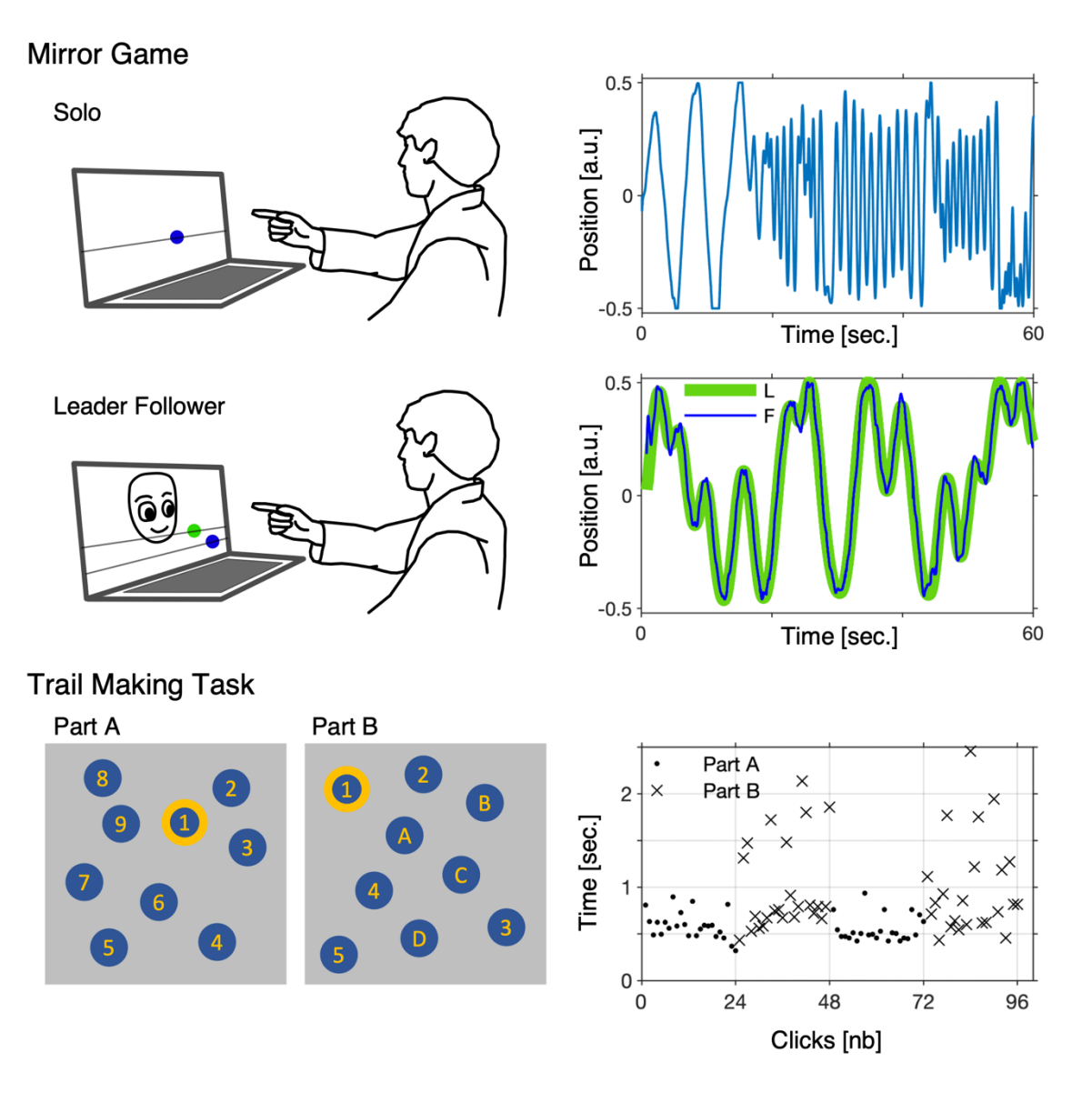

Figure 2 Illustration of the Mirror Game (MG) and the Trail Making Task (TMT) together with examples of collected data. In the MG participant sat in front of a computer with a connected leap motion sensor. In the Solo game *(first row) the participant was instructed: "Please move your hand left and right, create an interesting motion and enjoy playing." We recorded the horizontal hand movement (blue). In the* Leader-Follower game *(second row) the participant was instructed: "Please try to follow the movement of an animated robot as accurately as you can." We recorded movement generated by the computer (leader (L), green) and movement of the participant (follower (F), blue). In the TMT participant sat in front of the same laptop computer but was using a computer mouse to complete the task. In the TMT (third row) the participant was asked to connect a set of 25 dots as quickly and accurately as possible (in order given by numbers (Part A) and alternating numbers and letters (Part B)). We recorded the time between each click a person made on the screen and analysed both parts together (Part A, dots; Part B, crosses). For the sake of clarity, we show simplified illustration with 9 dots instead of 25.*

Sample size

For the feasibility study we recruited as many eligible SU as possible for the duration of the study. We approached all eligible participants that had been identified as appropriate for assessment for risk of psychosis by the Early Intervention for Psychosis Team. Convenience sampling allowed us to proceed with the study as quickly as possible and assess what are feasible sample sizes for future research. Sample size of the CC was driven by the research objectives of AW's Master's degree project.

Features extracted from data

Selection of features for classification was informed by our earlier work,[18] and modified to better fit machine learning methodology employed in the current study. Instead of using distributions (histograms) as in the previous work, here we use a set of their properties (e.g., mean, standard deviation or median). As previously, the data was concatenated or averaged across the repetitions. Before averaging or concatenating the movement data, we remove parts where participants' moves reached the edges of the sensor range (-0.5 or 0.5 value). The complete list of the features, and their description, is presented in Table 3.

Table 3 List of point measures and data features used to estimate them.

Point measure label	Description of data features
TMT ICT mean	Inter-click times (ICT) – times in msec. between each mouse click made by the participant while completing parts A and B of TMT. As features we used mean, standard deviation (std) and median. ICTs are a natural extension of the completion times used typically as TMT measure.
TMT ICT std	
TMT ICT median	
TMT IoTCT mean	Inter-on-target-click times, (IoTCT) – times in msec. between on target (correct) mouse clicks made by the participant while completing parts A and B of TMT. As features we used mean, standard deviation and median. IoTCTs are a natural extension of the completion times used typically as TMT measure.
TMT IoTCT std	
TMT IoTCT median	
S GWS p pf	Global wavelet spectrum (GWS) based on position (p) or velocity (v) time-series from the solo (S) task. GWS is a normalised time-average of a wavelet power spectrum based on continuous wavelet transform.[57] We only consider frequencies in the 0.25 and 5Hz band. As features we used the frequency of the highest peak (pf) of the GWS, and mean GWS frequency estimated as: $\mu_{GWS} = \int_{0.25}^{5} \omega\, \mathrm{GWS}(\omega) d\omega.$ We used GWS, rather than Fourier analysis, because it is better suited to characterise non-stationary time-series from the solo task.[19,47]
S GWS p mean	
S GWS v pf	
S GWS v mean	
LF GWS pos. mean	GWS based on position or velocity time-series of the human participant from the Leader-Follower (LF) task. As features we used the mean GWS frequency and GWS value (power) at 5Hz. We used GWS, rather than Fourier analysis, because it is better suited to characterise non-stationary time-series from the Leader-Follower task.[19,47]
LF GWS pos. 5HZ	
LF GWS vel. mean	
LF GWS vel. 5HZ	
LF RP frequency	Distribution of relative phase across wavelet frequency bands (limited to 1/15 – 2 Hz band), estimated as a circular mean over time of the phase of the wavelet cross-spectrum (WCS)[18,19,47] computed between the leader and follower positions time-series in the Leader-Follower task. As a feature we used the frequency at which the phase value drops below -pi/4. WCS allows to quantify relative phase (lag) between the leader and follower movements.
LF RP time mean	Relative phase estimated as a circular mean over frequencies (limited to 1/15 – 2 Hz band) of the phase of the WCS computed between the leader and follower positions time-series in the Leader-Follower task. As features we used mean, standard deviation and median.
LF RP time std	
LF RP time median	

Role of founding source

The founders had no involvement in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

RESULTS

Classes

We used the group (CC or SU), as the primary classification outcome (predicted variable). We used results of the CAARMS (i.e. ARMS, psychotic or neither) as an additional predicted variable. CAARMS was completed by trained personnel of STEPs before completion of the MG and TMT tasks.

Classification methods

For classification we employed Naïve Bayes (NB) classifier[48] as implemented in Matlab 2022b function fitcnb[49] with default settings. To avoid overfitting, we are using only three out of the 18 features (Table 3), namely one from the 6 measures estimated from the TMT data, one from the 4 measures estimated from the MG S task and one from the 8 measures estimated from MG Leader-Follower task. To select the features, we used the value of Cliff's delta,[50] a non-parametric measure of effect size. The set of classification features is chosen based on the training data, meaning that it is selected separately in each training-testing split (fold).

Additionally, we replicated the classification results using a k-nearest neighbours (kNN) classifier.[48] Specifically, we used a cosine distance to measure distances between the points in the 18-dimensional space defined by the 18 z-scored point measures (Table 3) and find the k-nearest neighbours. Cosine distance is defined as $1-cos(\theta)$, where θ is the angle between vectors defined by the point measures coordinates and the origin of the coordinate system.

Training and testing

To evaluate the performance of the classifiers, we used two training-testing splits. A 25% hold-out (HO) training-testing split and a leave-one-out (L1O) training-testing split (corresponding to 2% hold-out in our case). Parameters and hyperparameters of the classifier are identified using only the training set. Hold-out data is used only for testing and is unseen by the classifier during the training.

In the 25% HO split we select at random 25% of the data (8 out of 32 participants in each cohort). We train the classifier using the remaining 75% of the data (24 CC and 24 SU datasets). We use the 16 participants (8 CC and 8 SU datasets) unseen by the classifier during training to construct confusion matrix, and compute performance metrics. To estimate 95% confidence intervals of the classifier performance we repeat the 25% HO split 1000 times (1000 folds).

The leave-one-out (L1O) training-testing split uses 63 participants (98% of data) to train the classifier and 1 participant to test the model. L1O training-testing split allows to test the methodology 64 times. The L1O split simulates situation where we classify a new participant using classifier based on all the data available prior to the arrival of the new patient. To construct confusion matrix and compute performance metrics we compare the original classes with combined predictions of each of the 64 splits.

Classification results

The proposed methodology allows classification of the CC and SU participants with an outstanding (0.9-1) or excellent (0.8-0.9)[51] accuracy, sensitivity, specificity and precision. The only exception being the acceptable (0.7-0.8) sensitivity in case of the HO and kNN classifier. However, the methods fail to differentiate between SU with and without at-risk-mental-state (CAARMS score of 0 and CAARMS score > 0). Since there are only 16 participants with CAARMS = 0 and 16 participants with CAARMS > 0 we only use the L1O training-testing split. See Table 4 for details.

Table 4 Results of using two classification methods naïve Bayes and k-nearest neighbours (kNN) and two types of training-testing split (HO and L1O) for classifying SU and CC and results of L1O training-testing split classifying SU with and without at-risk-mental-state (CAARMS = 0 vs. CAARMS > 0).

		Groups	TN	FP	TP	FN	AUC	Accuracy	Sensitivity	Specificity	Precision
Naïve Bayes	HO	CC, n=8 SU, n=8	8 (6-8)	0 (0-2)	7 (5-8)	1 (0-3)	0.92 (0.75-1)	0.88 (0.75-1)	0.88 (0.62-1)	1 (0.75-1)	1 (0.73-1)
	L1O	CC, n=32 SU, n=32	30	2	27	5	0.91	0.89	0.84	0.94	0.93
	L1O	CAARMS=0, n=16 CAARMS>0, n=16	4	12	13	3	0.54	0.53	0.81	0.25	0.52
kNN	HO	CC, n=8 SU, n=8	8 (6-8)	0 (0-2)	6 (5-8)	2 (0-3)	0.91 (0.75-1)	0.88 (0.69-1)	0.75 (0.62-1)	1 (0.75-1)	1 (0.71-1)
	L1O	CC, n=32 SU, n=32	31	1	27	5	0.93	0.91	0.84	0.97	0.96
	L1O	CAARMS=0, n=16 CAARMS>0, n=16	2	14	8	8	0.22	0.31	0.5	0.13	0.36

Note: TN- true negative, FP – false positive, TP – true positive, FN – false negative. AUC – area under receiver operating characteristic (ROC) curve. For the HO we show median and (2.5 - 97.5 centiles) based on the 1000 folds.

To better understand the difference in performance of the proposed methodology in the two cases (CC vs. SU and SU CAARMS = 0 vs SU CAARMS > 0), we compared distributions of the point measures in the 3 groups. We did not find any statistically significant differences between the SU CAARMS = 0 vs SU CAARMS > 0 groups. However, we observed that the values of many point measures in the SU CAARMS > 0 cohort differ more from the CC than the SU CAARMS = 0 cohort;

see Figure 3. Specifically, we observed that the effect size (Cliff's delta) is overall higher for the point measures from the MG task while remaining relatively unchanged for TMT. The difference observed in the pattern of change between the SU CAARMS = 0 vs SU CAARMS > 0 groups is statistically significant $p < 0.0016$; bootstrap test with 10000 random splits of the SU cohort of observing simultaneous low change in effect sizes for TMT (smaller than the median change of the 6 TMT point measures), medium change in effect sizes for Solo MG task (larger than the median change of the 4 Solo MG point measures) and large change in effect sizes for Leader-Follower MG task (larger than the median change of the 8 Leader-Follower MG point measures).

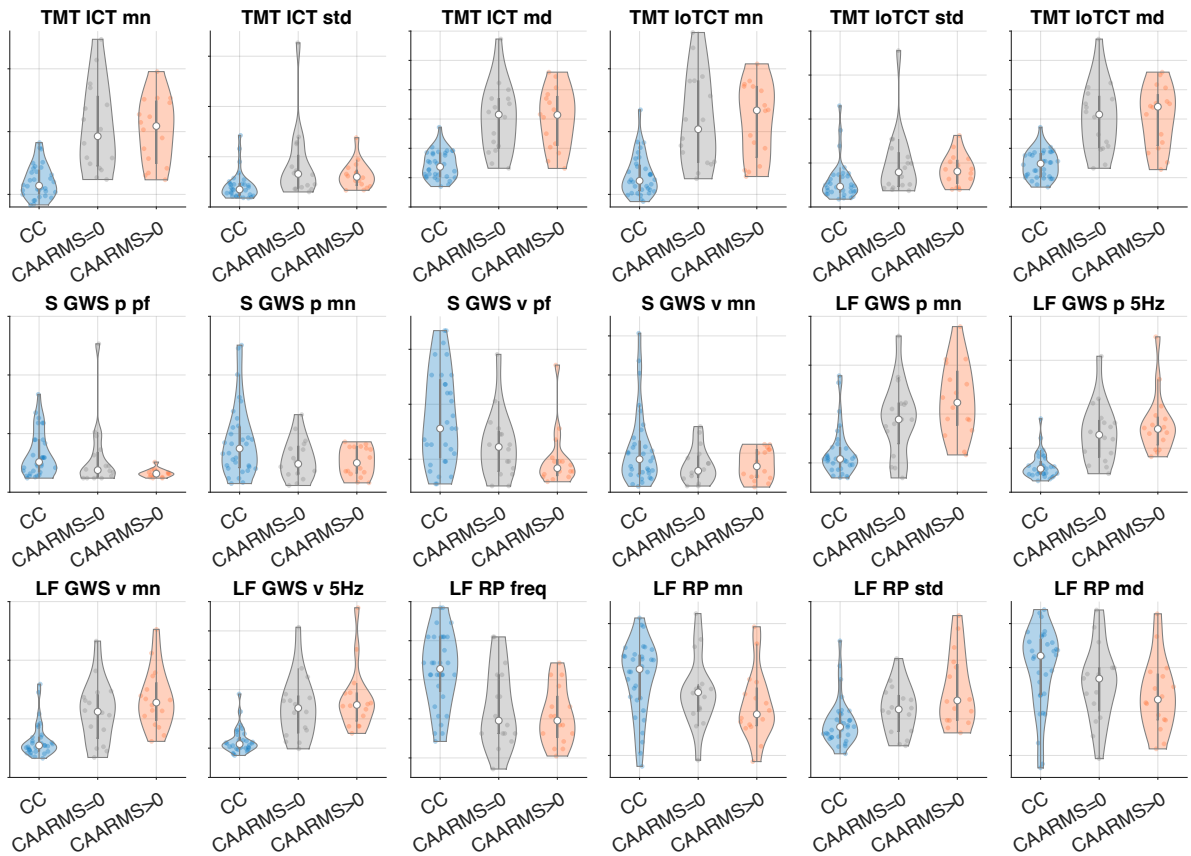

Figure 3 Distributions of point measure values in CC (blue) and two groups of SU, CAARMS=0 (gray) and CAARMS>0 (orange). Violin plots illustrate distributions of the values, white dot shows median, gray vertical bar shows IQR (middle 50% of values), scatterplots in each violin plot show all individual values. In all plots y-axis shows z-scored values in arbitrary units.

Table 5 Table of difference between point measures in CC and SU (combined), CC and SU with CAARMS=0 and CC and SU with CAARMS>0.

		CC vs. SU	CC vs. SU CAARMS = 0		CC vs. CAARMS > 0		
Point measure	**Effect size**	**p-value**	**Effect size**	**p-value**	**Effect size**	**p-value**	$\lvert es_{>0} - es_{=0} \rvert$
TMT ICT mean	0.81	<0.001	0.8	<0.001	0.82	<0.001	0.016
TMT ICT std	0.54	0.0028	0.53	0.023	0.56	0.0058	0.027
TMT ICT median	0.84	<0.001	0.84	<0.001	0.85	<0.001	0.01
TMT IoTCT mean	0.79	<0.001	0.78	<0.001	0.79	<0.001	0.016
TMT IoTCT std	0.46	0.016	0.43	0.075	0.49	0.023	0.063
TMT IoTCT median	0.83	<0.001	0.84	<0.001	0.83	<0.001	0.014
S GWS p pf	0.42	0.0028	0.24	0.46	0.61	<0.001	0.37
S GWS p mean	0.34	0.016	0.32	0.2	0.37	0.023	0.047
S GWS v pf	0.44	0.016	0.34	0.2	0.54	0.0012	0.2
S GWS v mean	0.23	0.58	0.25	0.31	0.21	0.31	0.043
LF GWS pos. mean	0.68	<0.001	0.54	<0.001	0.82	<0.001	0.28
LF GWS pos. 5HZ	0.82	<0.001	0.73	<0.001	0.91	<0.001	0.18
LF GWS vel. mean	0.68	<0.001	0.53	0.0012	0.83	<0.001	0.31
LF GWS vel. 5HZ	0.81	<0.001	0.71	<0.001	0.91	<0.001	0.2

LF RP frequency	0.59	0.001	0.51	0.023	0.66	0.0027	0.15
LF RP time mean	0.35	0.001	0.26	0.023	0.44	0.0058	0.18
LF RP time std	0.47	0.0028	0.34	0.042	0.61	0.0058	0.28
LF RP time median	0.36	0.0028	0.27	0.042	0.45	0.012	0.17

Note: Effect size (Cliff's delta) and p-value of Kolmogorov-Smirnov test. See also Table 3 and Fig. 3. $|es_{>0} - es_{=0}|$ *is the difference between effect sizes of CC vs. SU CAARMS > 0 and CC vs. SU CAARMS = 0.*

DISCUSSION

We presented results of a feasibility study in which we investigated the potential for employing digital behavioural tests in healthcare practice for stratification of specialist teams for early psychosis (STEP) users and healthy cohorts. Presented analysis of the data demonstrated that the two investigated behavioural tests (MG and TMT) can be used to differentiate between STEPs users and healthy cohorts with outstanding accuracy AUC>0.9 using two different classifiers naïve Bayes and k-nearest neighbours and two different training-testing splits, 25% hold-out and leave-one-out. We showed that cheap off-the-shelf equipment (laptop computer and a leap motion sensor) can be used to record clinically relevant behavioural data and that digital behavioural tests hold the prospect to aid clinical practice.

We also identified areas that require further research and development. We observed that the behavioural data from the MG and TMT collected in the current study cannot be used to differentiate between service users without (CAARMS=0) and SU with at-risk-mental-state (CAARMS=2) or first episode psychosis (CAARMS=4). This result might partially reflect limited specificity of the CAARMS assessment, meaning that only 15 - 22% of individuals with at-risk-mental-state develop a full psychotic disorder within 12 months.[51–53] Another possible limitation is the small number of participants available in our SU cohort. Nonetheless, the fact the SU can be so accurately differentiated from the CC confirms that the so-called 'non-cases' among STEPs referral have a range of characteristic behavioural marker and constitute an important clinical cohort that differs from control cohort.[54–57] Moreover, the Specificity = 1 achieved by our proposed methodology means that it most accurately identifies control participants. This is important as misclassification in terms of mental health state in young individuals could have equally serious consequences due to stigma associated with mental health diagnosis.[58]

Furthermore, we observe that SU with CAARMS>0 differ more from the CC than SU with CAARMS=0. This indicates presence of differences between these two cohorts which could be uncovered by means of including additional tasks and additional data modalities. For example, recordings of hand movements during the TMT or recordings of eye-movements during both tasks. Inclusion of eye-movement data could be particularly beneficial since it is demonstrated to have diagnostic potential.[24,27] Additionally, using mechanistic (differential equations) models[59] to combine eye-movements, reaction time and movement data could help to identify people's cognitive strategies e.g., employed to complete neuropsychological tasks.[60,61] Identification of the cognitive strategies and understanding their causal mechanisms would elucidate the role of pathophysiology in perturbed information processing and allow the development of new methodologies for risk and treatment stratification.

Finally, longitudinal study using digital behavioural tests would be instrumental for understanding how motor coordination and other neurological signs change (decline or improvement) in the course of psychosis and why, as shown by Ferruccio and collegues,[31] they allow to predict its long-term severity.

Study limitations

There are two main limitations of the study. First, we did not control the level of education in the two groups and it is know that the performance in TMT is affected by years of education.[17] However, even using only the point measures from the MG allow to classify the CC and SU with AUC=89 (0.73-1.00) and sensitivity=0.75 (95%CI 0.5-1) and specificity=1 (95%CI 0.75-1); naïve Bayes, 25% hold-out and 1000 folds. The second, potential source of bias is short exposure to antipsychotic medication; less than 4 months. We allowed 4 months of antipsychotic medication in order to facilitate recruitment of participants while minimising potential presence of motor side-effects that associated with anti-psychotic drugs.[62] We believe the effect of medication status to be minimal and additionally confound

with CAARMS score. Our earlier study showed results of classification to be independent from anti-psychotic medication status.[18] Furthermore, a recent study showed that neurological signs (e.g., tests of coordination and balance) and their change over 10 years is likely unrelated to exposure to anti-psychotic drugs.[31]

Implications for clinical and research practice
Our findings reinforce the benefits of digital behavioural test and quantitative analysis of their results and their potential for being used as a mobile assessment platform; assessable in home settings as well.[63] Cheap, portable off-the-shelf equipment allows the assessment to take place in a range of indoor locations, while automatic data collection greatly simplifies necessary training.

Digital behavioural tests would benefit research on prognostic instruments for psychosis. Recent review[51] identified heterogeneity in recruitment strategies for high-risk services as one of the factors limiting development of prognostic instruments for psychosis. Digital behavioural test could alleviate this limitation by stratifying and homogenising clinical high-risk populations.

Finally, with further development standardised digital test batteries could supplement and augment neuropsychiatric/ neurological tests making them quicker and easier to apply in routine clinical practice. This has the potential to be extended beyond STEPs users' stratification[64] and would facilitate rapid assessments for all people referred to mental health early intervention services,[65,66] improving their quality of life and wellbeing.

OTHER INFORMATION

Data sharing
The fully anonymised research data supporting this publication are openly available from: https://osf.io/rnzys/.

Ethics approval and full study protocol:
IRAS (Integrated Research Application System) ID: 23626.
REC (Research Ethics Committee) reference: 18/SW/0065, protocol number 1718/26.
University of Exeter, College of Life and Environmental Sciences (CLES), Psychology Ethics Committee: eCLESPsy000568 v2.1.

Further information about the study can be found at https://www.hra.nhs.uk/planning-and-improving-research/application-summaries/research-summaries/movement-and-perspective-taking-as-a-diagnostic-aid-for-psychosis/.

Full study protocol can be accessed at http://hdl.handle.net/10871/132205.

Contributions
Manuscript writing: P.S., K.T.A. Study design and methods development: P.S., S.S., T.E., Ph.S., A.K., K.T.A. Data acquisition and coding: P.S., A.W., S.L., J.W. Data analysis: P.S., K.T.A. Discussion of results and manuscript editing: All authors. Project coordination: P.S., K.T.A.

Competing interests
The authors declare no competing interests.

Sources of funding
The study was funded by: EPSRC Impact Acceleration Account, Impact & Knowledge Exchange Award, Jean Golding Institute seed corn award, Avon & Wiltshire Mental Health Partnership NHS Trust Research Capability Funding.
PS was generously supported by the Wellcome Trust Institutional Strategic Support Award 204909/Z/16/Z. KTA gratefully acknowledges the financial support of the EPSRC via grant EP/T017856/1.

The founders were not involved in study design; in the collection, analysis, and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Acknowledgments: The research team would like to thank all the STEPs users and control participants who generously shared their time and took part in the project. This study would not be possible without them.

BIBLIOGRAPHY

1 Yung AR. Chapter 3 - At-risk mental states. In: Thompson AD, Broome MR, eds. Risk Factors for Psychosis. Academic Press, 2020: 47–57.

2 de Mooij LD, Kikkert M, Theunissen J, *et al.* Dying Too Soon: Excess Mortality in Severe Mental Illness. *Frontiers in Psychiatry* 2019; **10**. https://www.frontiersin.org/articles/10.3389/fpsyt.2019.00855 (accessed Sept 1, 2022).

3 Ride J, Kasteridis P, Gutacker N, Aragon Aragon MJ, Jacobs R. Healthcare Costs for People with Serious Mental Illness in England: An Analysis of Costs Across Primary Care, Hospital Care, and Specialist Mental Healthcare. *Appl Health Econ Health Policy* 2020; **18**: 177–88.

4 Andrew, Alison, Knapp, Martin, McCrone, Paul, Parsonage, Michael, Trachtenberg, Marija. Effective interventions in schizophrenia: the economic case. 2012; published online Nov. http://www.schizophreniacommission.org.uk/the-report/ (accessed Sept 1, 2022).

5 Howes OD, Whitehurst T, Shatalina E, *et al.* The clinical significance of duration of untreated psychosis: an umbrella review and random-effects meta-analysis. *World Psychiatry* 2021; **20**: 75–95.

6 Polari A, Lavoie S, Yuen H-P, *et al.* Clinical trajectories in the ultra-high risk for psychosis population. *Schizophrenia Research* 2018; **197**: 550–6.

7 Mortimer A, Brown T. Early intervention in psychosis: another triumph of hope over experience? *Progress in Neurology and Psychiatry* 2015; **19**: 10–4.

8 O'Connell N, O'Connor K, McGrath D, *et al.* Early Intervention in Psychosis services: A systematic review and narrative synthesis of the barriers and facilitators to implementation. *European Psychiatry* 2022; **65**.

9 England NHS. Implementing the Early Intervention in Psychosis Access and Waiting Time Standard: Guidance. *NHS England, London* 2016. https://www.nice.org.uk/guidance/qs80/resources/implementing-the-early-intervention-in-psychosis-access-and-waiting-time-standard-guidance-2487749725.

10 Singh K, Ghazi F, White R, Sarfo-Adu B, Carter P. Improving access to Early Intervention in Psychosis (EIP): the 2-week wait for cancer comes to psychosis. *BMJ Open Qual* 2018; **7**: e000190.

11 O'Donoghue B, O'Connor K, Thompson A, McGorry P. The need for early intervention for psychosis to persist throughout the COVID-19 pandemic and beyond. *Irish Journal of Psychological Medicine* 2021; **38**: 214–9.

12 Yung AR, Wood SJ, Malla A, Nelson B, McGorry P, Shah J. The reality of at risk mental state services: a response to recent criticisms. *Psychological Medicine* 2021; **51**: 212–8.

13 Strelchuk D, Wiles N, Derrick C, Zammit S, Turner K. Identifying patients at risk of psychosis: a qualitative study of GP views in South West England. *Br J Gen Pract* 2021; **71**: e113–20.

14 Fusar-Poli P, Spencer T, De Micheli A, Curzi V, Nandha S, McGuire P. Outreach and support in South-London (OASIS) 2001—2020: Twenty years of early detection, prognosis and preventive care for young people at risk of psychosis. *European Neuropsychopharmacology* 2020; **39**: 111–22.

15 Perez J, Jin H, Russo DA, *et al.* Clinical effectiveness and cost-effectiveness of tailored intensive liaison between primary and secondary care to identify individuals at risk of a first psychotic illness (the LEGs study): a cluster-randomised controlled trial. *The Lancet Psychiatry* 2015; **2**: 984–93.

16 Fellows RP, Dahmen J, Cook D, Schmitter-Edgecombe M. Multicomponent analysis of a digital Trail Making Test. *The Clinical Neuropsychologist* 2017; **31**: 154–67.

17 Park S-Y, Schott N. The trail-making-test: Comparison between paper-and-pencil and computerized versions in young and healthy older adults. *Applied Neuropsychology: Adult* 2022; **29**: 1208–20.

18 Słowiński P, Alderisio F, Zhai C, *et al.* Unravelling socio-motor biomarkers in schizophrenia. *npj Schizophr* 2017; **3**: 8.

19 Słowiński P, Zhai C, Alderisio F, *et al.* Dynamic similarity promotes interpersonal coordination in joint action. *Journal of The Royal Society Interface* 2016; **13**: 20151093.

20 van Harten PN, Walther S, Kent JS, Sponheim SR, Mittal VA. The clinical and prognostic value of motor abnormalities in psychosis, and the importance of instrumental assessment. *Neuroscience & Biobehavioral Reviews* 2017; **80**: 476–87.

21 Fradkin SI, Erickson MA, Demmin DL, Silverstein SM. Absence of Excess Intra-Individual Variability in Retinal Function in People With Schizophrenia. *Frontiers in Psychiatry* 2020; **11**.

22 Garvey MA, Cuthbert BN. Developing a Motor Systems Domain for the NIMH RDoC Program. *Schizophrenia Bulletin* 2017; **43**: 935–6.

23 Stephenson DD, Shaikh AAE, Shaff NA, *et al.* Differing functional mechanisms underlie cognitive control deficits in psychotic spectrum disorders. *Journal of Psychiatry and Neuroscience* 2020; **45**: 430–40.

24 Morita K, Miura K, Kasai K, Hashimoto R. Eye movement characteristics in schizophrenia: A recent update with clinical implications. *Neuropsychopharmacology Reports* 2020; **40**: 2–9.

25 Vinogradov S, Poole JH, Willis-Shore J, Ober BA, Shenaut GK. Slower and more variable reaction times in schizophrenia: what do they signify? *Schizophrenia Research* 1998; **32**: 183–90.

26 Athanasopoulos F, Saprikis O-V, Margeli M, Klein C, Smyrnis N. Towards Clinically Relevant Oculomotor Biomarkers in Early Schizophrenia. *Frontiers in Behavioral Neuroscience* 2021; **15**.

27 Wolf A, Ueda K, Hirano Y. Recent updates of eye movement abnormalities in patients with schizophrenia: A scoping review. *Psychiatry and Clinical Neurosciences* 2021; **75**: 82–100.

28 Lemvigh CK, Brouwer RM, Pantelis C, *et al.* Heritability of specific cognitive functions and associations with schizophrenia spectrum disorders using CANTAB: a nation-wide twin study. *Psychological Medicine* 2022; **52**: 1101–14.

29 Dean DJ, Scott J, Park S. Interpersonal Coordination in Schizophrenia: A Scoping Review of the Literature. *Schizophrenia Bulletin* 2021; **47**: 1544–56.

30 Bolt LK, Amminger GP, Farhall J, *et al.* Neurocognition as a predictor of transition to psychotic disorder and functional outcomes in ultra-high risk participants: Findings from the NEURAPRO randomized clinical trial. *Schizophrenia Research* 2019; **206**: 67–74.

31 Ferruccio NP, Tosato S, Lappin JM, *et al.* Neurological Signs at the First Psychotic Episode as Correlates of Long-Term Outcome: Results From the AESOP-10 Study. *Schizophrenia Bulletin* 2021; **47**: 118–27.

32 Hedges EP, See C, Si S, McGuire P, Dickson H, Kempton MJ. Meta-analysis of longitudinal neurocognitive performance in people at clinical high-risk for psychosis. *Psychological Medicine* 2022; **52**: 2009–16.

33 Yung AR, Yung AR, Pan Yuen H, *et al.* Mapping the Onset of Psychosis: The Comprehensive Assessment of At-Risk Mental States. *Aust N Z J Psychiatry* 2005; **39**: 964–71.

34 Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. *J GEN INTERN MED* 2001; **16**: 606–13.

35 Stefanis NC, Hanssen M, Smirnis NK, *et al.* Evidence that three dimensions of psychosis have a distribution in the general population. *Psychological Medicine* 2002; **32**: 347–58.

36 Zhai C, Alderisio F, Słowiński P, Tsaneva-Atanasova K, Bernardo M di. Design of a Virtual Player for Joint Improvisation with Humans in the Mirror Game. *PLOS ONE* 2016; **11**: e0154361.

37 Cohen L, Khoramshahi M, Salesse RN, *et al.* Influence of facial feedback during a cooperative human-robot task in schizophrenia. *Sci Rep* 2017; **7**: 15023.

38 Ultraleap. Tracking | Leap Motion Controller | Ultraleap. 2022. https://www.ultraleap.com/product/leap-motion-controller/ (accessed Oct 30, 2022).

39 Brown EC, Casey A, Fisch RI, Neuringer C. Trail Making Test as a screening device for the detection of brain damage. *Journal of Consulting Psychology* 1958; **22**: 469–74.

40 Sánchez-Cubillo I, Periáñez JA, Adrover-Roig D, *et al.* Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society* 2009; **15**: 438–50.

41 Mueller, Shane T. The PEBL Trail-making task. https://pebl.sourceforge.net/.

42 Salesse RN, Casties J-F, Capdevielle D, Raffard S. Socio-Motor Improvisation in Schizophrenia: A Case-Control Study in a Sample of Stable Patients. *Front Hum Neurosci* 2021; **15**: 676242.

43 Goldstein G, Neuringer C. Schizophrenic and Organic Signs on the Trail Making Test. *Percept Mot Skills* 1966; **22**: 347–50.

44 Mahurin RK, Velligan DI, Hazleton B, Mark Davis J, Eckert S, Miller AL. Trail Making Test Errors and Executive Function in Schizophrenia and Depression. *The Clinical Neuropsychologist* 2006; **20**: 271–88.

45 Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* 1988; **54**: 1063–70.

46 Shimoyama I, Ninchoji T, Uemura K. The Finger-Tapping Test: A Quantitative Analysis. *Archives of Neurology* 1990; **47**: 681–4.

47 Grinsted A, Moore JC, Jevrejeva S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics* 2004; **11**: 561–6.

48 Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, NY: Springer, 2009 DOI:10.1007/978-0-387-84858-7.

49 Train multiclass naive Bayes model - MATLAB fitcnb - MathWorks United Kingdom. https://uk.mathworks.com/help/stats/fitcnb.html (accessed Nov 25, 2022).

50 Cliff N. Dominance statistics: Ordinal analyses to answer ordinal questions. *Psychological Bulletin* 1993; **114**: 494–509.

51 Oliver D, Arribas M, Radua J, *et al.* Prognostic accuracy and clinical utility of psychometric instruments for individuals at clinical high-risk of psychosis: a systematic review and meta-analysis. *Mol Psychiatry* 2022; : 1–9.

52 Salazar de Pablo G, Radua J, Pereira J, *et al.* Probability of Transition to Psychosis in Individuals at Clinical High Risk: An Updated Meta-analysis. *JAMA Psychiatry* 2021; **78**: 970–8.

53 Fusar-Poli P, Bonoldi I, Yung AR, *et al.* Predicting Psychosis: Meta-analysis of Transition Outcomes in Individuals at High Clinical Risk. *Archives of General Psychiatry* 2012; **69**: 220–9.

54 O'Donoghue B, Lyne J, Renwick L, *et al.* A descriptive study of 'non-cases' and referral rates to an early intervention for psychosis service. *Early Intervention in Psychiatry* 2012; **6**: 276–82.

55 Jordan G, Kinkaid M, Iyer SN, *et al.* Baby or bathwater? Referrals of "non-cases" in a targeted early identification intervention for psychosis. *Soc Psychiatry Psychiatr Epidemiol* 2018; **53**: 757–61.

56 Lindhardt L, Lindhardt M, Haahr UH, Hastrup LH, Simonsen E, Nordgaard J. Help-Seekers in an Early Detection of Psychosis Service: The Non-cases. *Frontiers in Psychiatry* 2021; **12**.

57 Edwards J, Norman R, Kurdyak P, *et al.* Unmet need for mental health services among people screened but not admitted to an early psychosis intervention program. *Schizophrenia Research* 2019; **204**: 55–7.

58 Rainteau N, Salesse RN, Macgregor A, Macioce V, Raffard S, Capdevielle D. Why you can't be in sync with schizophrenia patients. *Schizophrenia Research* 2020; **216**: 504–6.

59 Słowiński P, Al-Ramadhani S, Tsaneva-Atanasova K. Neurologically Motivated Coupling Functions in Models of Motor Coordination. *SIAM J Appl Dyn Syst* 2020; **19**: 208–32.

60 Danion FR, Flanagan JR. Different gaze strategies during eye versus hand tracking of a moving target. *Sci Rep* 2018; **8**: 10059.

61 Brenner E, de la Malla C, Smeets JBJ. Tapping on a target: dealing with uncertainty about its position and motion. *Exp Brain Res* 2022.

62 Kahn RS, Sommer IE, Murray RM, *et al.* Schizophrenia. *Nat Rev Dis Primers* 2015; **1**: 1–23.

63 Koo BM, Vizer LM. Mobile Technology for Cognitive Assessment of Older Adults: A Scoping Review. *Innovation in Aging* 2019; **3**: igy038.

64 Mallawaarachchi SR, Amminger GP, Farhall J, *et al.* Cognitive functioning in ultra-high risk for psychosis individuals with and without depression: Secondary analysis of findings from the NEURAPRO randomized clinical trial. *Schizophrenia Research* 2020; **218**: 48–54.

65 Shah JL. Sub-threshold mental illness in adolescents: within and beyond DSM's boundaries. *Soc Psychiatry Psychiatr Epidemiol* 2015; **50**: 675–7.

66 Iyer SN, Boksa P, Lal S, *et al.* Transforming youth mental health: a Canadian perspective. *Irish Journal of Psychological Medicine* 2015; **32**: 51–60.