Brazilian Journal of Psychiatry

bjp

Revista Brasileira de Psiquiatria

REVIEW ARTICLE

Lowering costs for large-scale screening in psychosis: a systematic review and meta-analysis of performance and value of information for speech-based psychiatric evaluation

Felipe **Argolo**,[1,2]* Guilherme **Magnavita**,[3]* Natalia Bezerra **Mota**,[4,5] Carolina **Ziebold**,[1] Dirceu **Mabunda**,[6] Pedro M. **Pan**,[1] André **Zugman**,[7] Ary **Gadelha**,[1] Cheryl **Corcoran**,[8,9] Rodrigo A. **Bressan**[1,2]

[1]Universidade Federal de São Paulo, São Paulo, SP, Brazil. [2]King's College London, London, UK. [3]Universidade de São Paulo (USP), São Paulo, SP, Brazil. [4]Brain Institute, Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil. [5]Departamento de Física, Universidade Federal de Pernambuco (UFPE), Recife, PE, Brazil. [6]Faculdade de Medicina, Universidade Eduardo Mondlane, Maputo, Mozambique. [7]National Institute of Mental Health (NIMH), Bethesda, MD, USA. [8]Icahn School of Medicine at Mount Sinai, New York, NY, USA. [9]Mental Illness Research, Education and Clinical Center (MIRECC VISN2), New York, NY, USA. * These authors have contributed equally to this manuscript.

Objective: Obstacles for computational tools in psychiatry include gathering robust evidence and keeping implementation costs reasonable. We report a systematic review of automated speech evaluation for the psychosis spectrum and analyze the value of information for a screening program in a healthcare system with a limited number of psychiatrists (Maputo, Mozambique).
Methods: Original studies on speech analysis for forecasting of conversion in individuals at clinical high risk (CHR) for psychosis, diagnosis of manifested psychotic disorder, and first-episode psychosis (FEP) were included in this review. Studies addressing non-verbal components of speech (e.g., pitch, tone) were excluded.
Results: Of 168 works identified, 28 original studies were included. Valuable speech features included direct measures (e.g., relative word counting) and mathematical embeddings (e.g.: word-to-vector, graphs). Accuracy estimates reported for schizophrenia diagnosis and CHR conversion ranged from 71 to 100% across studies. Studies used structured interviews, directed tasks, or prompted free speech. Directed-task protocols were faster while seemingly maintaining performance. The expected value of perfect information is USD 9.34 million. Imperfect tests would nevertheless yield high value.
Conclusion: Accuracy for screening and diagnosis was high. Larger studies are needed to enhance precision of classificatory estimates. Automated analysis presents itself as a feasible, low-cost method which should be especially useful for regions in which the physician pool is insufficient to meet demand.

Keywords: Psychosis; automated speech analysis; speech analysis; verbal behavior; schizophrenia

Introduction

As the research community closes gaps in translational neuroscience, psychiatry revisits classical topics, such as verbal behavior in mental disorders.[1] Evidence from the early 20th century suggested that speech contains valuable information for psychopathology (reviewed in Cohen & Elvevåg[2]). The first descriptions of psychosis already distinguished language disturbance as a core feature in the disease.[3-5] Recent efforts combine computational processing of human natural language (natural language processing, NLP) with clinical expertise in order to obtain predictive models, with prominent results in the psychosis

spectrum. We systematically reviewed the literature to assess the current knowledge on automated speech evaluation. Our study compiles 1) common theoretical grounds, 2) pooled accuracy estimates for predictive models, and 3) the value of perfect and imperfect information for developing a screening program for clinical high risk (CHR) of psychosis and subsequent intervention to prevent psychosis.

Biology of speech production in humans

Verbal behavior mechanisms have been extensively studied in humans using numerous heuristics, such as

formal complexity of language,[6,7] errors in spontaneous production,[8] clinical presentation of aphasia,[9] and comparative grammar.[10]

The investigation of specific speech features to assess underlying brain structure is connected to the very earliest days of psychometrics, when pioneer Francis Galton designed the word association test, later modified by Jung.[11] Other instruments[12,13] stimulated spontaneous speech production from context with total or partial freedom for the subject, such as the Thematic Apperception Test (TAT).

Symbolic structure, phonological variations, and semantic connections were considered to communicate subtle characteristics intrinsic to one's brain phenotype. A preponderance of evidence supports – and most theories agree with – the existence of hierarchical, sequentially distinct stages in language production, from conception of words to articulation of speech.[14] These differences are topologically related to an evolutionary hierarchy of cortical development,[15] with cortical expansion in temporoparietal and frontal hub regions.[16]

Verbal impairments in psychosis

Early psychopathologists carefully described communication impairments in psychosis which were generally regarded to reflect deep disturbances of thought.[17] These features included word salad, speech disorganization, clanging, derailment, tangentiality, and assonance.[3,18] Although observations were consistent, limitations (e.g., real-time measurement and subjective scoring) precluded practical applications of speech assessment.[19]

Abnormal neuroimaging findings in multiple cortical and subcortical regions are observed in schizophrenia.[20-24] Within the speech domain, semantic processing seems to be altered in associative areas, namely the dorsolateral prefrontal cortex (DLPC) and inferior parietal cortex (IPC).[25]

Recent approaches have employed automated analysis of speech, in which mathematical models provide insight about observed behavioral patterns, with implications for risk assessment,[26] diagnostic support, and prognostic monitoring.[27] This avenue of investigation has led to a steady accrual of evidence that supports the potential of automated speech analysis to transform prediction and diagnosis of psychosis, although an overarching framework is still lacking.

Value of information for speech screening in distributed health systems

Randomized clinical trials (RCTs) allow estimation of the causal effect of interventions – or "decisions" – in nearly counterfactual scenarios with minimal assumptions. However, RCTs are not available for all interventions, in all outcomes, for all timeframes, in every population. In such cases, decision science models allow us to expand the conclusions of empirical studies given a set of mathematical assumptions. As such, we may integrate data from different sources in the medical literature to simulate a population under two – or more – counterfactual scenarios, or decisions, as if they were parallel arms in an RCT.

We can incorporate population heterogeneity, adapt findings from studies in one population to another by changing the distribution of effect modifiers, and incorporate cost and life expectancy information to extend the horizon of analysis of the original RCT. All this surplus information comes with the price that the answer is only correct if the mathematical assumptions embedded in the model are correct.

As there is no empirical RCT to evaluate the impact of a population-based screening program for identification of psychosis risk, preventive intervention, and the effect thereof on reducing incidence of schizophrenia (improving quality of life or life expectancy), we chose to simulate the Mozambique population under the counterfactual scenarios of "screening and treating" vs. "no screening." We model the expected benefit of a distributed screening program as the increase in quality-adjusted life years (QALYs) in a scenario with few mental health specialists. Effectiveness and related costs are simulated for the city of Maputo, Mozambique, which serves a country of nearly 30 million people, yet has only 30 primary health care facilities, four general hospitals, 13 psychiatrists, and one psychiatric hospital.

The value of information of such a screening tool for a potential funding agency is estimated as the expected impact given the metrics of sensitivity and specificity achieved by the instrument. The expected value of perfect information (EVPI) represents an upper bound for investment in a new test. We conducted our analysis to calculate both the EVPI (e.g., for a perfect test) and the value of information at lower sensitivity and specificity (e.g., an imperfect test), to calculate the upper limits of costs for developing software to detect individuals with psychosis risk at that accuracy level.

Methods

Systematic review

We proceeded according to the Cochrane protocol and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines for reporting. We report: 1) main common theoretical grounds; 2) accuracy findings for risk assessment in CHR and diagnosis in schizophrenia; and 3) the value of information for implementing decentralized screening.

Eligibility criteria

We included all original studies assessing psychosis and psychosis risk through automated analysis of speech. The reference lists of articles found through database searches were hand-searched to identify additional relevant publications. Studies were included regardless of population characteristics, types of outcomes, or design. We limited the scope of this review to verbal behavior, excluding studies of nonverbal aspects of expression, such as intonation and pitch (e.g., prosody and pauses). Although they carry information, audio wave analyses require different assumptions, methods, and techniques. We consider analysis of transcriptions to be more feasible in clinical settings and

less susceptible to noise from technological issues and the environment. They also rely on the quality of the recorded audio; hence, they are not robust to input noise. Studies published through August 2018 in English, Spanish, and Portuguese were included.

Literature sources and study selection

We used the PubMed search engine (granting access to the MEDLINE database and additional references from the National Library of Medicine), including all material published through August 2018.

Keywords

We designed a search query with a combination of semantically similar keywords to capture (a) "automated analysis" of (b) "speech" in (c) "psychosis." Consonant with the objective of investigating conditions with underlying related phenotypes (e.g., endophenotypes), we directly searched for disorders with consistently reported coheritability, in addition to general terms.[28-30]

Search query:

(a) ("automated" OR "computerized" OR "computational") AND ("analysis" OR "assessment" OR "evaluation") AND
(b) ("speech" OR "semantics" OR "language" OR "prosody" OR "pauses" OR "acoustics" OR "paralinguistic" OR "fundamental frequency" OR "nonverbal expression") AND
(c) ("mental disorder" OR "psychiatric" OR "psychotic" OR "schizophrenia")

Data extraction

For each manuscript, we assessed the following key data: 1) study design; 2) demographics; 3) sample characteristics, including psychopathological measures and/or diagnostic criteria; 4) protocol for elicitation of verbal response; 5) speech metrics used; 6) validity analysis/classification techniques; 7) main findings and classification accuracy (if available); and 8) software toolkit (when available) and reference language corpus used for analyses.

Accuracy estimates

We used pooled discriminatory estimates obtained from the review to assess accuracy performance. Errors and confidence intervals (CIs) for classification were obtained, assuming an underlying binomial distribution.

Value of information model methods

Model overview

We analyzed the value of perfect and imperfect information for implementing a screening program based on automated speech analysis to assess psychosis and psychosis risk in Maputo, the capital of Mozambique,

a country with less than 15 psychiatrists[31] for a population of 28,571,310 (source: http://www.ine.gov.mz/). We developed a Markov cohort to assess the value of information of a software for screening once for CHR of psychosis over a lifetime horizon of cost per QALY. The model includes two different strategies: screening once + preventive treatment vs. no screening. For the no-screening strategy, four health states are possible: regular risk, CHR, schizophrenia, and death. For the screen-and-treat strategy, six health states are possible: untreated regular risk, treated regular risk, untreated high risk, treated high risk, schizophrenia, and death (absorbing state). Possible state transitions are shown in Figure 1.

Model outcomes were life expectancy, expected QALY, and expected costs of the intervention, but not of the software. Transition probabilities between health states were modeled on the basis of multiple sources from the literature; probabilities and costs which could not be drawn from published literature or official documents were estimated by our board of experts. Details are provided in Table S1, available as online-only supplementary material. Our model calculated the probability of being in each health state at any given time by multiplying the previous state with the corresponding transition matrix. Costing was considered from the screening and treatment program funding-agency perspective. We present a more detailed description of how we modeled clinical progression of disease in the supplementary material.

Intervention

The intervention we designed consisted of a single screening assessment lasting 10-15 minutes followed by 6 months of treatment with cognitive therapy, following the work of Morrison et al.[32] For the imperfect information analysis, the screening tool had a sensitivity of 70% and specificity of 70%. As reported by Bearden et al.,[33] anual methods already possessed the ability to detect transition to schizophrenia within CHR patients in 2011. Therefore, detecting prodromal states and attenuated symptoms with 70% accuracy is not only feasible, but likely an underestimation of current models. For the perfect information scenario, sensitivity and specificity for the CHR state are set at 100% at the time of screening. Treatment is assumed to be maintained for 6 months for all positively screened patients, except for those who dropped out, developed threshold psychosis (e.g., schizophrenia), or died in the first year. After treatment termination, costs related to treatment and follow-up were set at 0, and – as mentioned above – treatment benefit was only maintained until the end of the second year. In our base case, we assumed a 10% risk of loss to follow-up, with a constant rate of schizophrenia incidence during the 6 months of intervention and dropouts immediately losing all later – but not previous – protective effects from treatment.

Utilities

Regular-risk patients were assumed to have perfect utility. CHR patients were assumed to have a utility of 0.95 due to the burden of prodromal symptoms, as defined by our
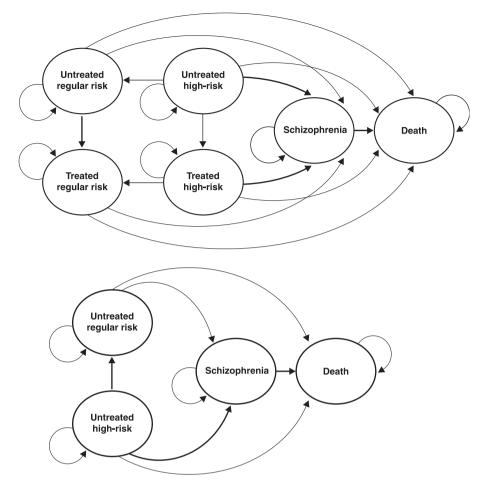
Figure 1 Bubble diagram for the screen + treat strategy (top) and no-screening strategy (bottom)

board of experts based on utility values for schizophrenia patients. Treatment itself was associated with a 0.02 decrease in utility for all treated false-positive patients due to the burden of attending appointments. The schizophrenia state had a utility of 0.75; this refers to the perceived utility of moderate to severe schizophrenia among healthy people according to the work of Lenert et al.[34] One of the authors (DM), a psychiatrist at the Mozambique Ministry of Health, felt that this disease level best reflected the average experience of a patient with schizophrenia in Mozambique, even after treatment, due to diagnostic delay.

Setting

The program is meant to be deployed at the 30 primary health care facilities, four general hospitals, and single psychiatric hospital of the capital, Maputo. Patients would be screened by a program-trained psychologist at one of these centers, and those who screened positive would then receive therapy at one of the primary facilities on an outpatient basis. As the capital has a population of 2,500,000, assuming the most conservative estimate from van Os et al.[35] of 1.9% for 1-year prevalence of psychotic episodes, we would still have a pool of 47,500 patients in need of our screening program in 1 year, far

above the maximum load our psychologists could take in. If we screened 9,000 patients a year (i.e., 0.7 screening sessions per center per day), our ability to treat both true positives and false positives would not be overextended. We would therefore be able to screen 45,000 patients during the 5-year period we used to calculate costs related to program deployment and fixed costs. We chose to calculate costs aiming to screen 30,000 patients, to yield a more conservative estimate.

Costs

We adopted the costing perspective of the program's funding agency: therefore, we included costs to train personnel, buy the screening hardware (smartphones), pay for internet access and power (to charge phones), and compensate psychologists for screening and providing therapy. Costs were calculated from a bottom-up approach – i.e., individual costs from the program workforce and infrastructure were added together. As the program would not fund regular schizophrenia treatment for Mozambique, patients who developed threshold psychosis (e.g., schizophrenia) would be accepted for treatment by the government health system and drop out of the program. Also, as all costs are intervention-related, the no-screening strategy had a total cost of 0. Fixed costs

related to program set-up were divided by the estimated number of patients we expected the program to cover. All data related to costs and the expected number of patients to be screened were provided by the Mozambique Ministry of Health. Fixed costs and installation costs were calculated for a 5-year period for a total of 30,000 people; as noted above, this yields a conservative estimate if compared to a longer horizon, since some of our costs – e.g., personnel training – would not accrue immediately after the 5-year window. All costs are given in 2018 dollars (Dirceu Mabunda, personal communication, 2018 Oct 08).

Value of information analysis

Willingness-to-pay (WTP) for a QALY was set at three times the current value of Mozambique's gross domestic product (GDP) per capita, or USD 1,247.154. This is a commonly used (albeit conservative) QALY estimate for developing countries. Similar thresholds for the U.S. would range between USD 50,000 to USD 150,000 per QALY. Differences in expected QALY between strategies were calculated and then multiplied by the WTP value. The program maintenance cost was subtracted from that, and the remaining value would be our value of information for that setting. We calculated both the EVPI and the value for an imperfect test with sensitivity and specificity of 70%.

Sensitivity and scenario analyses

We conducted a deterministic sensitivity analysis on the impact of the initial prevalence of the CHR state in the population, as there is little knowledge of this input in the literature and the prevalence in Mozambique may be different from the one reported in Conrad et al.[36] We display our EVPI, incremental life year (LY), and incremental QALYs as functions of a range of possible prevalence values, set between 5 and 50%, based on the bounds of psychosis-like symptoms shown in the work of van Os et al.[35] and in the screening phase of Morrison et al.[32]

Validation and calibration

We validated our modeling of the intervention – a risk ratio of 0.263 for schizophrenia conversion according to the CHR conversion risk meta-analysis, followed by a sudden loss of effect compared to controls after the first year of intervention – by predicting the expected risk ratio for conversion to schizophrenia at 3 years and comparing it to the risk ratio reported by Morrison et al. for the 3-year follow-up of their original RCT cohort.[37,38] We also compare the performance of our model against some of the calibration benchmarks used as references to construct it. Our cumulative incidence of psychosis among the CHR state at 12, 24, and 48 months is compared to the work of Schmidt et al.,[39] while our remission risk among CHR individuals at 6 months is compared to the work of Polari et al.[40]

Ethical aspects

For automated speech analysis to be helpful in the detection and screening of mental illness, technological barriers are not the only issue. Predictions are based on indirect information about cognitive functioning, as high-risk phenotypes are inferred from language patterns.

Even if high concordance between such algorithms and personalized (face-to-face) diagnostics is achieved, one must keep in mind that psychiatric diagnostics are structured around crucial criteria. We emphasize the need for later in-person confirmation of initial findings. This is incorporated in the screening system outlined, where NLP is used a screening tool. Individuals showing seemingly altered patterns are referred to centers for further evaluation.

Results

One hundred and sixty-eight papers were initially screened, of which 28 were included in the final review, with 24 including psychosis patients and four including CHR individuals. Of the 24 studies including psychotic individuals, three also included first-episode psychosis (FEP) groups. The list of studies included in this systematic review is shown in Table 1.

Included studies: design, sample size, protocol time

Studies were fairly heterogeneous. Protocol time ranged from 1-minute structured tasks[60] to 60 minutes of open interviews,[54] while sample sizes ranged from 34[27] to 665 subjects.[60] Nevertheless, 75% of the studies (n=21) included between 40 and 110 patients. One study used a small collection of five speech samples: four schizophrenia patients and one control.[61]

There were no large studies providing estimate accuracies for computerized classification. Generally, open interviews were conducted in scenarios close to clinical settings,[45] while task-based assessment took place in artificial environments.[41]

Protocols

Texts were obtained from recordings of either 1) tasks, such as describing a picture for 1 minute,[41] 2) prompted free speech,[50] 3) structured interviews and scales, such as the Indiana Psychiatric Illness Interview (IPII),[43] and 4) open-ended clinical interviews.[27]

As seen in Table 2, protocols based on tasks and prompted free speech were generally faster, reporting the same pattern of findings as longer protocols did. Open protocols and unstructured interviews resulted in higher heterogeneity due to variability within interviewer-generated stimuli, as well as distinct individual responses.

Theoretical grounds

Different ways of measuring similar features yielded valuable information from prediction. Semantic coherence, phrase length, and the frequency of a specific class of words (determiners) were features used to achieve good classification performance for psychosis onset in patients at risk.[26,45] The underlying constructs being

Table 1 Full summary of studies included for systematic review of automated speech analysis to assess psychosis

Reference	Journal	N (controls)	Central age	Sample description	Protocol	Protocol time (min)	Metrics	Main findings
Palaniyappan[41]	Progress in Neuropsycho-pharmacology & Biological Psychiatry	56 (0)	34.6 (SD 10.4); 32.9 (SD 8.9)	Schizophrenia (34) and bipolar disorder (22).	Three 1-minute instances of freely generated speech on the presentation of the pictures for the TLI.	3	Connectedness (graph analysis)	Moderate association between speech connectedness and brain markers, global functioning, and thought disorder.
Pauselli[42]	Psychiatry Research	105 (47)	33.2 (SD 9.9)	58 patients (two groups: derailment and no derailment) with DSM-IV diagnosis of schizophrenia or first-episode non-affective psychosis (schizophreniform disorder and psychotic disorder, not otherwise specified).	The subject is asked to say as many words belonging to a semantic category (e.g., animals, vegetables) as possible in a certain amount of time, usually 60 seconds.	1	Mean similarity, coherence, coherence 5, coherence 10	Significant differences among groups, number of words, coherence 5, coherence 10.
Minor[43]	Psychological Medicine	81 (0)	49.7 (SD 10.71)	81 outpatients from a Midwestern VA Medical Center. All participants had a DSM-IV diagnosis of schizophrenia (n=56) or schizoaffective disorder (n=25) confirmed via the structured clinical interview for DSM-IV-TR Disorders - Patient Edition.	Automated analysis was conducted on speech generated in response to the IPII, a semi-structured interview that assesses perceptions of one's life and illness. The open-ended nature of the IPII was a key reason for its selection; its format differs from many structural interviews and speech tasks in that subjects control how long they speak with little input or affective prompting from examiners. IPII interviews were typically 30-60 min in length, allowing subjects to generate substantial samples for analysis (total words: mean 2,786, SD 2,117).	30-60	Deep cohesion, referential cohesion, word concreteness, syntactic simplicity, syllables per word, type-token ratio	Correlation with neurocognition, social cognition, and metacognition.
Gupta[44]	Schizophrenia Research	84 (43)	Ultra high-risk 19.33 (SD 1.44); control 18.76 (SD 2.63)	84 individuals (ultra high-risk = 41, control = 43).	Participants were instructed to write a brief story about an image depicting a woman washing dishes while two children take cookies from a jar. Participants were given up to 10 min to produce their narratives.	10	Three measures of referential overlap in themes (words vs. words; sentence; words; sentence vs. sentence)	Group differences in referential cohesion.
Corcoran[45]	World Psychiatry	59 (40)	Several groups: UCLA 17.36 (SD 3.7), 16.46 (SD 3.0), 18.06 (SD 2.8), 15.86 (SD 1.7)/NYC 22.26 (SD 3.4), 21.26 (SD 3.6)	Ultra high-risk and FEP: two centers located at NYC and UCLA.	UCLA: Caplan's Story Game, in which participants retell and then answer questions about a story they hear ("what do you like about it?"; "is it true?"), and then construct and tell a new story; NYC: open-ended narrative interviews of about one hour were obtained by interviewers trained by an expert in qualitative research methods.	N/A; 60	Maximum semantic coherence, variance in semantic coherence, minimum semantic coherence, frequency of use of possessive pronouns.	Psychosis in the UCLA cohort accuracy 83% using the logistic regression classifier. CHR with respect to psychosis onset (p < 0.05 upon label randomization), with a true negative ratio of 0.82 (24/29) and a true positive ratio of 0.60 (3/5), that is, an overall accuracy of 0.79. AUC of 0.87 in the receiver operating characteristic. NYC: true negative ratio of 0.82 (24/29) and a true positive ratio of 0.60 (3/5), i.e., an overall accuracy of 0.79. AUC of 0.72 for FEP.
Hong[46]	Psychiatry Research	39 (16)	33.21	23 schizophrenia and 16 controls.	Narrative of five emotions.	N/A	Generic (e.g., average word length/word repetition), word identity, dictionary (LIWC), language models (LMs, PP)	Classification accuracy of 65% using one story and 74% using all stories.

Table 1 (*continued*)

Reference	Journal	N (controls)	Central age	Sample description	Protocol time (min)	Protocol	Metrics	Main findings
Buck & Penn[47]	Journal of Nervous and Mental Disease	90 (48)	25-60	DSM-IV criteria for either schizophrenia or schizoaffective.		NET.	Words per sentence; pronoun use.	The AUC in predicting group membership was 0.823 (p < 0.001) for words per sentence and 0.790 (p < 0.001) for pronoun use, indicating acceptable to good sensitivity and specificity in identifying individuals with schizophrenia and non-clinical controls using these lexical characteristics.
Bonfils[48]	Psychiatry Research	45 (N/A)	48.5	Schizophrenia (17) and schizoaffective (28)Participants were eligible for the study if they were receiving mental health services at either a VA Medical Center or a local community mental health center, were older than 18 years of age, had a diagnosis of schizophrenia.	30-60	The IPII is a semi-structured interview that asks participants to tell the story of their lives in as much detail as possible. Participants were interviewed by trained research assistants. Interviews were typically less than 1 hour (n=38, 84%).	LIWC (64 word categories); means and SDs for overall word count, lexical categories, and hope variables. Considering the large proportion of the sample diagnosed with schizoaffective disorder (62%), the authors ran a series of independent t tests to assess for any impact of diagnosis.	Speech features correlated with "hope."
Moe[49]	Schizophrenia Research	47 (15)	41.6	DSM-IV-TR + SADS.	N/A	IPII.	Idea density	
Fineberg[50]	Psychological Medicine	46 (23)	35.2	Psychosis (schizophrenia, schizoaffective, and bipolar); 23 subjects with psychosis from the inpatient psychiatric hospital and outpatient clinic.	10	Recorded and transcribed speech. Prompt: We like to begin by hearing about you. Would you tell us a little about yourself?	Word category frequencies (LWIC)	Lexical markers previously identified as specific language changes in depression and psychosis are probably markers of illness in general.
Mota[51]	Front in Psychology	73 (28)	34 schizophrenia (39 bipolar)	25 schizophrenia, 20 bipolar; DSM-IV + PANSS and BPRS.	N/A	Reports from the most recent memorable dream, followed by questions about regular dreaming (translated from Portuguese): Do your dreams usually resemble your daily life? Do your dreams usually resemble your psychotic symptoms? Do your dreams change following changes in medication? Also about lucid dreaming: Can you be aware of dreaming during sleep? Can you control your dream when this happens? How frequently does this happen? How do you feel when you wake up from these dreams?	Graph analysis	There was no clinical advantage for lucid dreamers among psychotic patients, even for the diagnostic question specifically related to lack of judgment and insight.
Bedi[26]	NPJ Schizophrenia	34 (29)	22	At risk for psychosis.	60	Participants were asked to describe changes they had experienced and the impact of these changes, what had been helpful or unhelpful for them, and their expectations for the future.	LSA; maximum phrase length, use of determiners (e.g., which)	100% accuracy to predict psychosis after 2.5 years.
Elvevåg[52]	Journal of Neurolinguistics	83 (30)	N/A	Individuals were selected as part of two large cohort studies, one a family study of individuals from families with a high density of schizophrenia and another a longitudinal study of first episode patients with schizophrenia over time.	N/A	Participants were asked to talk about whatever came to mind, perhaps what they did yesterday or what they would like to be doing.	LSA, LDA	0.77 accuracy.

Table 1 (*continued*)

Reference	Journal	N (controls)	Central age	Sample description	Protocol time (min)	Protocol	Metrics	Main findings
Rosenstein[53]	*Cortex*	122 (76)	N/A	Schizophrenia (n=28), unaffected siblings (n=18), and unrelated healthy control participants (n=76).	N/A	N/A	Word recall, LSA, character n-gram	Accuracy ~ 70%. LSA was the most accurate predictor. Taken individually, the semantic feature is most predictive, while a model combining the features improves accuracy of group membership prediction slightly above the semantic feature alone as well as over the human rating approach.
Elvevåg[27]	*Schizophrenia Research*	51 (25)	34	SM-IV criteria for schizophrenia or schizoaffective disorder.	45	Speech generated in a 45 minute semi-structured clinical interview with open-ended questions (including questions regarding symptoms, current events, as well as why some people believe in God and what free-will is).	LSA	82.4% accuracy (78.4% cross-valid).
Tagamets[20]	*Cortex*	22 (11)	40	11 schizophrenia and 11 controls.	N/A	Free speech about religion.	CSA, fMRI, behavioral data	In persons with schizophrenia, coherence was mainly related to auditory and visual regions, depending on the modality of monitoring, but superior/middle temporal cortex related to coherence regardless of task; these findings are consistent with existing evidence for a role of the superior temporal cortex in thought disorder, and demonstrate that computational measures of semantic content capture objective measures of coherence in speech that can be usefully related to underlying neurophysiological processes.
Mota[54]	*PLoS One*	24 (8)	N/A	DSM-IV - Schizophrenia and mania.	20-60	Subjects were requested to report exclusively on their most recently experienced dream, which served as an anchor topic. Recordings proceeded without interference from the interviewer.	Graphs: nodes, edges, average total degree, largest connected component, largest strongly connected component, parallel edges, loops with one, two, and three nodes (L1, L2, L3), waking nodes, waking edges, density, diameter, average shortest path	High AUROC values vs. controls.
Thomas[55]	*British Journal of Psychiatry*	65 (16)	N/A	First psychotic episode; schizophrenia and mania.	N/A	Interview with five sections.	% well-formed major sentences, mean maximum depth of embedding, % deviant sentences, % syntactically deviant sentences, % errors of commission	Results indicated that, even at the earliest stages of illness, schizophrenic speech is distinct from that of other groups.
García-Laredo[56]	*Medicine (Baltimore)*	62 (15)	39.6	Paranoid schizophrenia; bipolar with and without psychotic symptoms; controls; DSM-IV.	1	PVF-FAS: participants were requested to produce the maximum possible number of words beginning with a specific letter (F, A, and S). SVF: participants were requested to produce the maximum possible number of words that belong to the animal (high fluency) and tools (low fluency) categories.	Phonemic verbal fluency, semantic verbal fluency.	Performance was lower in psychotic groups (schizophrenia and psychotic bipolar).
Buck et al.[57]	*Journal of Clinical Psychology*	41 (0)	49.2	23 schizophrenia and 18 schizoaffective disorder.	30-60	IPII, TEPS, PANSS, LIWC.	Relative frequencies of various content and word type categories	Results revealed that relatively higher levels of both anticipatory and consummatory anhedonia were linked with fewer past-related words and lesser use of first-person plural pronouns.

Table 1 (*continued*)

Reference	Journal	N (controls)	Central age	Sample description	Protocol	Protocol time (min)	Metrics	Main findings
Buck et al.[58]	*Comprehensive Psychiatry*	58 (0)	49.9	34 schizophrenia and 24 schizoaffective disorder.	IPII (30 ~ 60 minutes), MAS (informant-rated: self-reflectivity, understanding the mind of others, mastery, and decentration), LIWC.	N/A	LIWC	Lexical characteristics indicative of cognitive complexity were significantly related to level of metacognitive capacity, while social cognition was related to second-person pronoun use, articles, and prepositions, and pronoun use overall.
Holshausen[59]	*Cortex*	165 (0)	66	Geriatric schizophrenia patients.	Semantic fluency task participants were required to find as many different animals as possible in the span of 90 seconds.	1.5	LSA (average vector length, cosine)	Mostly negative findings; small impacts (delta R^2 in logistic regression blocks ~0.03). Dependent: severity of deficits in community affairs on the CDRS; unusualness was significantly associated with semantic fluency and phonological fluency, disconnectedness in speech, and impaired functioning, even after considering the contribution of premorbid cognition, positive and negative symptoms, and demographic variables.
Nicodemus[60]	*Cortex*	665 (307 controls; 164 siblings)	N/A	Siblings study - DSM-IV criteria for schizophrenia or schizoaffective disorder, depressed type, all siblings were free from schizophrenia spectrum disorders.	The authors used a category fluency task where a participant generated words in response to the cue animal for one minute.	1	LSA: coherence, total number of valid words, N words sequence coherence, vector length, cosine to "animal," average cosine between all terms.	Some genes (*DISC1*; *KIAA0319*; *ZNF804A*) associated with LSA derived variables.
Bearden[33]	*Journal of the American Academy of Child and Adolescent Psychiatry*	105 (51)	17	CHR participants met criteria for of one of three prodromal syndrome categories, as assessed by the SIPS 27: 1) attenuated (subthreshold) psychotic symptoms; 2) transient, recent-onset psychotic symptoms; or 3) a substantial drop in social/role functioning in conjunction with schizotypal personality disorder diagnosis or a first-degree relative with a psychotic disorder.	A three part audiotape was played. In the first and the last part, the subject listens to a brief audiotaped story and is asked to retell it, as well as to answer a set of open-ended questions about the narrative. Examples of questions: What did (or didn't) you like about that story? Do you think this is a true story? Why (or why not)? In the middle part, the participant is asked to select one of four topics and asked to construct their own story.	20-25	FTD	Illogical thinking was uniquely predictive of subsequent conversion to psychosis, whereas poverty of content and referential cohesion were significant predictors of social and role functioning, respectively.
Cabana[61]	*Schizophrenia Research*	5 (1)	N/A	4 samples of schizophrenia patients and 1 control.	Exploratory analysis with focus on measures.	N/A	Graphs; entropy	Mean tropic entropy was higher in patients.
Shakeel[62]	*Cortex*	48 (24)	N/A	24 schizophrenia patients and 24 controls.	N/A	N/A	Word repetition, semantic category word generation	The performance of patients with schizophrenia was significantly inferior to that of healthy control subjects in both the "pre-scan" and "intra-scan" sessions of the computerized task.

AUC = area under the curve; AUROC = area under the receiver operating characteristic curve; BPRS = Brief Psychiatric Rating Scale; CHR = clinical high-risk; FEP = first-episode psychosis; fMRI = functional magnetic resonance imaging; FTD = formal thought disorder; IPII = India Psychiatric Illness Interview; LDA = linear discriminant analysis; LIWC = Linguistic Inquiry Word Count Software; LMS = learning management system; LSA = latent semantic analysis; MAS = Metacognitive Assessment Scale; N/A = not available; NET = Narrative of Emotions Task; NYC = New York City; PANSS = Positive and Negative Syndrome Scale; PP = Perplexity; PVF-FAS = Phonemic Verbal Fluency-FAS; SADS = seasonal affective disorder; SD =standard deviation; SIPS = Structured Interview for Prodromal Syndromes; TEPS = Temporal Enjoyment of Pleasures Scale; TLI = Thought and Language Index; UCLA = University of California, Los Angeles.

Table 2 Source data for accuracy estimates used in value of information model of automated speech analysis to assess psychosis

Reference	Accuracy	True negative	True positive	Time to event (years)	AUC	Cross-sectional
Corcoran[45] - Sample A	0.83	-	-	2	0.87	No
Corcoran[45] - Sample B (cross-validation)	0.79	0.82 (24/29)	0.60 (3/5)	2.5	0.72	No
Bedi[26] (different classifier for sample in Corcoran[45]; Bedi[26] reports cross-validation)	1	1 (29/29)	1 (5/5)	2.5	1	No
Buck & Penn[47]	-	-	-	-	0.823/0.790 (two features)	Yes
Hong[46]	0.744	-	-	-	-	Yes
Rosenstein[53]	0.708	0.93 (259/273)	0.55 (44/80)	-	-	Yes
Elvevåg[52]	0.771	0.60 (18/30)	0.868 (46/53)	-	-	Yes
Elvevåg[27]	0.824 (cross-valid: 0.784)	-	-	-	-	Yes
Bearden[33]*	0.705	0.710	0.690	2	-	No

AUC = area under the curve.
* Non-automated, manually obtained metrics.

assessed reflect classical clinical characteristics, such as disorganized speech and alogia.

Graph analysis

Graph-based embeddings normally rely on the shortest path between nodes and/or on weighted edges as distance measures. Secondary characteristics include values associated with network structure, such as centrality, modularity, and connectedness.[54]

Palaniyappan et al.[41] reported that connectedness in speech-derived graphs correlated moderately with degree of centrality in the brain through resting-state functional magnetic resonance imaging (fMRI). Other potential biomarkers include candidate genes (*DISC1*, *KIAA0319*, *ZNF804A*) that are associated with speech features.[60]

Derailment and semantic coherence

Carl Schneider[63] described as derailment (*entgleisen*) the unexpected association of ideas among psychosis patients. Several authors have deployed similar strategies to assess this feature. Sentences are mapped to a set of ordered objects from which coherence can be measured.

Words are first lemmatized into their roots and/or tagged according to syntactical function.

(1) *Purify → Pure*
  Then, they are mapped to new structures according to a given morphism (*f*) from the set of words to real valued n-dimensional vectors. Their similarity is evaluated adopting an arbitrary metric space and a distance measure. For the common cosine distance:
(2) f : word $\to$ u
(3) f : (*Pure*) = v = [0.11,0.12,0.43,0.75, …]
(4) f : (*Love*) = v′ = [0.11,0.44,0.99,0.33, …]
(5) Dist. (*Pure, Love*) = cos (v,v′)

where u, v, and v′ are vector mappings of words and *d* is a real value bounded by the cosine function in the interval [0, 1]. The distance is taken as a measure of semantic coherence. For each sample, semantic coherence was

calculated from phrases relative to previous ones.[26] These measures are used as input features for predictive models (e.g., logistic regression, support vector machines [SVMs]).

Accuracy

Six studies reported classification accuracy using automated speech processing.

Classification of psychotic disorder

Elvevåg et al.[27] achieved 82.4% accuracy in distinguishing language in schizophrenia from that of healthy individuals (78.4% using cross-validation). Speech samples were gathered from a language task ("tell me the story of Cinderella"), processed using latent semantic analysis (LSA) (TK Landauer), and classified using a linear discriminant classifier. A similar experiment by this group[52] used transcripts from a variety of prompt questions. Features analyzed included surface features, statistical language features, and semantic features derived from LSA. Fisher discriminant analysis correctly classified 86.8% (46 out of 53) patients and 60% (18 out of 30) healthy individuals (healthy family members and unrelated healthy controls), giving an overall cross-validated classification accuracy of 77.1%.

Buck & Penn[47] used transcripts from the Narrative of Emotions Task (NET) processed by the Linguistic Inquiry Word Count (LIWC) software, which calculates the relative frequency of 83 word categories in a given text. The area under the receiver operating characteristic (AUROC) curve was 0.823 for "words per sentence" and 0.790 for pronoun use.

Hong[46] elicited autobiographical experiences to produce four classes of input: 1) generic features (e.g., words per sentence); 2) word identity features (e.g., frequency of specific words); 3) dictionary features (categories deriving from dictionaries); and 4) language model features (bigram probabilistic model). A SVM was used to evaluate

Table 3 Incremental QALYs and LYs

| | Imperfect screening | | Perfect screening |
	No screening	Screen + treat	Screen + treat
LYs	42.26	42.39	42.45
QALYs	41.83	42.00	42.09
Discounted LYs	22.50	22.55	22.57
Discounted QALYs	22.24	22.32	22.35
Costs (USD)	-	25.11	11.29
Discounted costs (USD)	-	24.89	11.21
Incremental LYs		0.13	0.19
Incremental QALYs		0.17	0.26
Incremental discounted LYs		0.04	0.07
Incremental discounted QALYs		0.07	0.11
Expected value of information (million USD)		5.78	9.34

LYs = life years; QALYs = quality-adjusted life years.

performance and select valuable features. The accuracy reported for the final model was 0.744.

Semi-structured interviews[53] processed with LSA and analyzed with proportional-odds logistic regression yielded accuracy similar to that obtained with manual analysis (approximately 70%).[33] There was no added information when both were combined.

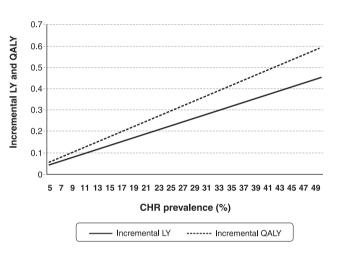*Prediction of conversion of clinical high risk (CHR) to psychosis*

Initial investigations[33] suggest that speech-based analysis provided good classification performance (70.5%) when predicting CHR conversion to psychosis within 2 years, even when carried out with manually obtained features (true negative: 71%; true positive: 69%).

More recent studies evaluating conversion of CHR state to psychosis within a 2.5-year range suggest high accuracy ratings, including error free results.[26] Reproduction in a larger sample suggested more parsimonious estimates (83%; cross-validation: 79%). This work included a slightly shorter follow-up of 2 years, using LSA and part-of-speech tagging as features. Values were processed with singular value decomposition and classified with a logistic regression model.[45] This was also the only study available reporting classification accuracy for FEP, achieving a 72% mark.

Model metrics matched our inputs, and the assumption of no effect after 1 year matched the 3-year follow up estimate by Morrison et al. in 2007[37] (Figure S1, available as online-only supplementary material).

Results for the base-case and perfect-screening analyses are presented in Table 3. The intervention increased both life expectancy (0.13 years) and QALY (0.17) in the imperfect screening analysis. Under perfect test characteristics, there is a modest improvement in efficacy, with a life expectancy increase of 0.19 years and of 0.26 in QALYs. However, costs are reduced by decreasing therapy personnel costs with false positives.

In Figure 2, increasing CHR prevalence values with imperfect information (sensitivity and specificity of 70%) also increased increments of both QALY and life years in a linear fashion between the likely prevalence range of 5 to 50%, with increments between 0.1 QALY and 0.6 QALY.

Figure 2 Increments in life years (LYs) and quality-adjusted life years (QALYs) as a function of clinical high-risk (CHR) prevalence (%)

Discussion

The role of language in psychopathology has been studied for more than a century; however, recent technological and conceptual advances now allow for innovative approaches. Fast mathematical approaches and low-cost electronic devices provide feasible ways of recording, processing, and obtaining information from patients' verbal expression.[64,65]

The current literature comprises studies with improved ecological validity, such as direct analyses of free speech during clinical assessment. Multiple independent groups from several nations are aiming at robust candidates for clinical practice.

We detected two broad experimental setting styles: some include recording situations close to normal conversations taking place in a healthcare service, while others study brief, well-defined tasks. Most studies seek to identify key features of psychosis, and have achieved good classificatory performance using features as input for machine learning models.

Cross-sectional studies achieved fairly good accuracy for diagnosis of schizophrenia (above 70% in all studies). Using data from longitudinal studies, researchers were

capable of predicting conversion to psychosis within 2.0-2.5 years of assessment with high accuracy (70-100% across studies).

Only Mota et al.[54] included non-English recordings, with speech samples in Brazilian Portuguese. As no studies included large samples, the CIs for performance estimates remain wide. Future meta-analyses including studies with larger sample sizes may settle the diagnostic characteristics. Further studies addressing classification performance will refine and confirm the reliability of accuracy estimates and clinical effectiveness.

Models have shown improved performance when incorporating NLP variables.[26] Most studies have used LSA, which relies on single value decomposition for embeddings. These models may yet improve even further, since other methods (e.g., skip-n-gram; long-short term memory neural networks) have achieved higher marks in tasks from other domains.[66]

Aside from small samples, study-level bias may arise from heterogeneity in settings. Results have been fairly consistent across studies; however, elicitation procedures and analysis strategies vary. We noticed that recent efforts center around predictive models using ecologically obtainable speech samples.

For our specific case of Mozambique (and likely for other countries with limited numbers of psychiatrists), developing software for CHR detection would be valuable for any likely prevalence range we could expect, even if the screening tool had relatively poor performance (sensitivity 70% and specificity 70%) and we assumed a punitive WTP of three times the GDP per capita. It is likely that current algorithms would perform better at detecting prodromal symptoms of psychosis, since current software already fares much better in detecting subsets of patients in the high-risk state who are more likely to convert to schizophrenia. Therefore, any effort to develop software to detect CHR in Mozambique would be expected to yield a large health benefit, probably overshadowing its cost by an ample margin.

One limitation of the value-of-information model is that follow-up of CHR patients who underwent therapy ended at 3 years after randomization. Nevertheless, this may not be problematic, as rates of conversion to psychosis steadily fall off in the years following initial ascertainment.[67] Similarly, we included some of the costs to set up the screening program, but assumed it would continue for a limited time period of 5 years. Were it held for longer, the cost-benefit balance might be even more favorable towards screening, since fixed set-up costs would be shared by a wider population.

In sum, this systematic review reports the first meta-analysis of performance and value of information estimates for speech-based psychiatric evaluation using electronic devices.

A decentralized system coupled with support for focused treatment would be highly efficient. The estimates obtained suggest language analysis would be a valuable screening asset, even for imperfect screening metrics which would be easy to obtain, as compared to similar tools.

Construct validity

Mental disorders compromise higher-order functions of the brain. In psychosis, speech may be a valuable window for assessment, manifesting thought disorganization through several disturbances. The notions of early psychopathologists regarding semantic incoherence are now formalized with NLP to provide reliable signals.

Current tools available for language assessment, such as structured interviews, rely on operator characteristics and adequate training. Nevertheless, it is virtually impossible to remove bias from interviewer cues. Conversely, automated analysis of speech provides clinical indicators which can be interviewer-neutral.

Automated analysis of speech is coherently articulated with biological findings,[41] neuroimaging,[20] measurable behavior,[2] abstractions,[27] and clinical applications.[26] Although behavioral and technical advances are supported by the clinical results gathered, the identification of putative biomarkers is still under development. Initial cross-validation of computational variables with brain activity (fMRI) has been achieved in in the psychosis spectrum,[20,68] but only once[41] with NLP. The relation between speech impairment[60] and previously implied genes (e.g., DISC1) supports discrete language patterns as potentially valuable endophenotypes for psychosis.

Further exploration of natural language structure and its counterparts in the brain will enhance the adequacy of mathematical models. Language carries a hierarchical structure which is efficiently embedded in hyperbolic spaces.[69] Different embedding strategies (e.g., LSTM networks, Poincaré embeddings) and distance measures yield similar or better results in non-clinical tasks.[66]

We restricted our findings to processing of symbolic outputs. Nonverbal cues, such as intonation and pauses, may provide different information about neurocognitive characteristics in psychosis.[19] These perspectives are complementary, and will benefit from combined analysis in order to enhance efficiency.

Acknowledgements

Disclosure

The authors report no conflicts of interest.

References

1 Huys QJ, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci. 2016;19:404-13.
2 Cohen AS, Elvevåg B. Automated computerized analysis of speech in psychiatric disorders. Curr Opin Psychiatry. 2014;27:203-9.

3 Kuperberg GR. Language in schizophrenia part 1: an introduction. Lang Linguist Compass. 2010;4:576-89.

4 Kraepelin E. [100 years of psychiatry]. Vertex. 2010;21:317-20.

5 Bleuler E. [Dementia praecox or the group of schizophrenias]. Vertex. 2010;21:394-400.

6 Chomsky N. Syntactic structures. Berlin: Mouton; 1957.

7 Jager G, Rogers J. Formal language theory: refining the Chomsky hierarchy. Philos Trans R Soc Lond B Biol Sci. 2012;367:1956-70.

8 Fromkin VA. Grammatical aspects of speech errors. In: Newmeyer FJ. Linguistic theory: extensions and implications. Cambridge: Cambridge; 1988. Linguistics: The Cambridge Survey, v. 2, p. 117-38.

9 Tremblay P, Dick AS. Broca and Wernicke are dead, or moving past the classic model of language neurobiology. Brain Lang. 2016;162.60-71.

10 Bresnan J, Kaplan R, Peters S, Zaenen A. Cross-serial dependencies in Dutch. In: Savitch WJ, Bach E, Marsh W, Safran-Naveh G, editors. The formal complexity of natural language. Dordrecht: Springer; 1982. Studies in Linguistics and Philosophy (formerly Synthese Language Library), v. 33, p. 286-319..

11 Jung CG. The structure and dynamics of the psyche [Internet]. Vol. 8. 1969 [cited 2019 Feb 10]. www.jungiananalysts.org.uk/wp-content/uploads/2018/07/C.-G.-Jung-Collected-Works-Volume-8_-The-Structure-and-Dynamics-of-the-Psyche.pdf

12 Rorschach H. Psychodiagnostics. Oxford: Grune & Stratton Psychodiagnostics; 1942.

13 Freed H, Ecker WF. The thematic apperception test. Dis Nerv Syst. 1946;7:146-51.

14 Baese-Berk M, Goldrick M. Mechanisms of interaction in speech production. Lang Cogn Process. 2009;24:527-54.

15 Aboitiz F. A Brain for speech. Evolutionary continuity in primate and human auditory-vocal processing. Front Neurosci. 2018;12:174.

16 Hill J, Inder T, Neil J, Dierker D, Harwell J, Van Essen D. Similar patterns of cortical expansion during human development and evolution. Proc Natl Acad Sci U S A. 2010;107:1315-40.

17 Andreasen NC, Grove WM. Thought, language, and communication in schizophrenia: diagnosis and prognosis. Schizophr Bull. 1986;12:348-59.

18 Engels H, Heynick F, van der Staak C. Emil Kraepelin's dream speech: a psychoanalytic interpretation. Int J Psychoanal. 2003;84:1281-94.

19 Cohen AS, Mitchell KR, Docherty NM, Horan WP. Vocal expression in schizophrenia: less than meets the ear. J Abnorm Psychol. 2016;125:299-309.

20 Tagamets MA, Cortes CR, Griego JA, Elvevåg B. Neural correlates of the relationship between discourse coherence and sensory monitoring in schizophrenia. Cortex. 2014;55:77-87.

21 Hagoort P. Nodes and networks in the neural architecture for language: Broca's region and beyond. Curr Opin Neurobiol. 2014;28:136-41.

22 Xiang HD, Fonteijn HM, Norris DG, Hagoort P. Topographical functional connectivity pattern in the perisylvian language networks. Cereb Cortex. 2010;20:549-60.

23 Sabb FW, van Erp TG, Hardt ME, Dapretto M, Caplan R, Cannon TD, et al. Language network dysfunction as a predictor of outcome in youth at clinical high risk for psychosis. Schizophr Res. 2010;116:173-83.

24 Collin G, Seidman LJ, Keshavan MS, Stone WS, Qi Z, Zhang T, et al. Functional connectome organization predicts conversion to psychosis in clinical high-risk youth from the SHARP program. Mol Psychiatry. 2018 Nov 8. doi: 10.1038/s41380-018-0288-x. [Epub ahead of print]

25 Kuperberg GR, West WC, Lakshmanan BM, Goff D. Functional magnetic resonance imaging reveals neuroanatomical dissociations during semantic integration in schizophrenia. Biol Psychiatry. 2008;64:407-18.

26 Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, et al. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015;1:15030.

27 Elvevåg B, Foltz PW, Weinberger DR, Goldberg TE. Quantifying incoherence in speech: an automated methodology and novel application to schizophrenia. Schizophr Res. 2007;93:304-16.

28 Cross-Disorder Group of the Psychiatric Genomics Consortium, Lee SH, Ripke S, Neale BM, Faraone SV, Purcell SM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. Nat Genet. 2013;45:984-94.

29 Cross-Disorder Group of the Psychiatric Genomics Consortium. Identification of risk loci with shared effects on five major psychiatric disorders: a genome-wide analysis. Lancet. 2013;381:1371-9.

30 Geschwind DH, Flint J. Genetics and genomics of psychiatric disease. Science. 2015;349:1489-94.

31 Wagenaar BH, Cumbe V, Raunig-Berhó M, Rao D, Kohrt BA, Stergachis A, et al. Outpatient mental health services in Mozambique: use and treatments [Internet]. 2016 Feb 1 [cited 2020 Mar 12]. ps.psychiatryonline.org/doi/full/10.1176/appi.ps.201500508

32 Morrison AP, French P, Walford L, Lewis SW, Kilcommons A, Green J, et al. Cognitive therapy for the prevention of psychosis in people at ultra-high risk: randomised controlled trial. Br J Psychiatry. 2004;185:291-7.

33 Bearden CE, Wu KN, Caplan R, Cannon TD. Thought disorder and communication deviance as predictors of outcome in youth at clinical high risk for psychosis. J Am Acad Child Adolesc Psychiatry. 2011;50:669-80.

34 Lenert LA, Sturley AP, Rapaport MH, Chavez S, Mohr PE, Rupnow M. Public preferences for health states with schizophrenia and a mapping function to estimate utilities from positive and negative symptom scale scores. Schizophr Res. 2004;71:155-65.

35 van Os J, Linscott RJ, Myin-Germeys I, Delespaul P, Krabbendam L. A systematic review and meta-analysis of the psychosis continuum: evidence for a psychosis proneness-persistence-impairment model of psychotic disorder. Psychol Med. 2009;39:179-95.

36 Conrad AM, Lewin TJ, Sly KA, Schall U, Halpin SA, Hunter M, et al. Utility of risk-status for predicting psychosis and related outcomes: evaluation of a 10-year cohort of presenters to a specialised early psychosis community mental health service. Psychiatry Res. 2017;247:336-44.

37 Preti A, Cella M. Randomized-controlled trials in people at ultra high risk of psychosis: a review of treatment effectiveness. Schizophr Res. 2010;123:30-6.

38 Morrison AP, French P, Parker S, Roberts M, Stevens H, Bentall RP, et al. Three-year follow-up of a randomized controlled trial of cognitive therapy for the prevention of psychosis in people at ultrahigh risk. Schizophr Bull. 2007;33:682-7.

39 Schmidt SJ, Schultze-Lutter F, Schimmelmann BG, Maric NP, Salokangas RK, Riecher-Rössler A, et al. EPA guidance on the early intervention in clinical high risk states of psychoses. Eur Psychiatry. 2015;30:388-404.

40 Polari A, Lavoie S, Yuen HP, Amminger P, Berger G, Chen E, et al. Clinical trajectories in the ultra-high risk for psychosis population. Schizophr Res. 2018;197:550-6.

41 Palaniyappan L, Mota NB, Oowise S, Balain V, Copelli M, Ribeiro S, et al. Speech structure links the neural and socio-behavioural correlates of psychotic disorders. Prog Neuropsychopharmacol Biol Psychiatry. 2019;88:112-20.

42 Pauselli L, Halpern B, Cleary SD, Ku B, Covington MA, Compton MT. Computational linguistic analysis applied to a semantic fluency task to measure derailment and tangentiality in schizophrenia. Psychiatry Res. 2018;263:74-9.

43 Minor KS, Willits JA, Marggraf MP, Jones MN, Lysaker PH. Measuring disorganized speech in schizophrenia: automated analysis explains variance in cognitive deficits beyond clinician-rated scales. Psychol Med. 2019;49:440-8.

44 Gupta T, Hespos SJ, Horton WS, Mittal VA. Automated analysis of written narratives reveals abnormalities in referential cohesion in youth at ultra high risk for psychosis. Schizophr Res. 2018;192:82-8.

45 Corcoran CM, Carrillo F, Fernández-Slezak D, Bedi G, Klim C, Javitt DC, et al. Prediction of psychosis across protocols and risk cohorts using automated language analysis. World Psychiatry. 2018;17:67-75.

46 Hong K, Nenkova A, March ME, Parker AP, Verma R, Kohler CG. Lexical use in emotional autobiographical narratives of persons with schizophrenia and healthy controls. Psychiatry Res. 2015;225:40-9.

47 Buck B, Penn DL. Lexical characteristics of emotional narratives in schizophrenia: relationships with symptoms, functioning, and social cognition. J Nerv Ment Dis. 2015;203:702-8.

48 Bonfils KA, Luther L, Firmin RL, Lysaker PH, Minor KS, Salyers MP. Language and hope in schizophrenia-spectrum disorders. Psychiatry Res. 2016;245:8-14.

49 Moe AM, Breitborde NJ, Shakeel MK, Gallagher CJ, Docherty NM. Idea density in the life-stories of people with schizophrenia:

associations with narrative qualities and psychiatric symptoms. Schizophr Res. 2016;172:201-5.

50 Fineberg SK, Leavitt J, Deutsch-Link S, Dealy S, Landry CD, Pirruccio K, et al. Self-reference in psychosis and depression: a language marker of illness. Psychol Med. 2016;46:2605-15.

51 Mota NB, Resende A, Mota-Rolim SA, Copelli M, Ribeiro S. Psychosis and the control of lucid dreaming. Front Psychol. 2016;7:294.

52 Elvevåg B, Foltz PW, Rosenstein M, Delisi LE. An automated method to analyze language use in patients with schizophrenia and their first-degree relatives. J Neurolinguistics. 2010;23:270-84.

53 Rosenstein M, Diaz-Asper C, Foltz PW, Elvevåg B. A computational language approach to modeling prose recall in schizophrenia. Cortex. 2014;55:148-66.

54 Mota NB, Vasconcelos N, Lemos N, Pieretti AC, Kinouchi O, Cecchi GA, et al. Speech graphs provide a quantitative measure of thought disorder in psychosis. PLoS One. 2012;7:e34928.

55 Thomas P, Romme M, Hamelijnck J. Psychiatry and the politics of the underclass. Br J Psychiatry. 1996;169:401-4.

56 García-Laredo E, Maestú F, Castellanos MÁ, Molina JD, Peréz-Moreno E. The relationship between educational years and Phonemic Verbal Fluency (PVF) and Semantic Verbal Fluency (SVF) tasks in Spanish patients diagnosed with schizophrenia, bipolar disorder, and psychotic bipolar disorder. Medicine (Baltimore). 2015;94: e1596.

57 Buck B, Minor KS, Lysaker PH. Lexical characteristics of anticipatory and consummatory anhedonia in schizophrenia: a study of language in spontaneous life narratives. J Clin Psychol. 2015;71:696-706.

58 Buck B, Minor KS, Lysaker PH. Differential lexical correlates of social cognition and metacognition in schizophrenia; a study of spontaneously-generated life narratives. Compr Psychiatry. 2015;58:138-45.

59 Holshausen K, Harvey PD, Elvevåg B, Foltz PW, Bowie CR. Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. Cortex. 2014;55:88-96.

60 Nicodemus KK, Elvevåg B, Foltz PW, Rosenstein M, Diaz-Asper C, Weinberger DR. Category fluency, latent semantic analysis and schizophrenia: a candidate gene approach. Cortex. 2014;55:182-91.

61 Cabana Á, Valle-Lisboa JC, Elvevåg B, Mizraji E. Detecting order-disorder transitions in discourse: iImplications for schizophrenia. Schizophr Res. 2011;131:157-64.

62 Shakeel MK, Halahalli HN, Kumar K, Jain S, John JP. Utility of a computerized, paced semantic verbal fluency paradigm in differentiating schizophrenia and healthy subjects. Asian J Psychiatr. 2014;7:22-7.

63 Schneider C. Die Psychologie der Schizophrenen: und ihre Bedeutung für die Klinik der Schizophrenie. Stuttgart: G. Thieme; 1930.

64 Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. Trends Cogn Sci. 2012;16:72-80.

65 Iter D, Yoon J, Jurafsky D. Automatic detection of incoherent speech for diagnosing schizophreniaIn:Loveys K, Niederhoffer K, Prud'hommeaux E, Resnik R, Resnik F, editors. Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: from Keyboard to Clinic; 2018; New Orleans, USA. Stroudsburg: Association for Computational Linguistics; 2018. p. 136-46.

66 Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space [Internet]. [cited 2020 Feb 12]. arxiv.org/pdf/1301.3781.pdf

67 Nelson B, Yuen HP, Wood SJ, Lin A, Spiliotacopoulos D, Bruxner A, et al. Long-term follow-up of a group at ultra high risk ("Prodromal") for psychosis: the PACE 400 study. JAMA Psychiatry. 2013;70:793-802.

68 Powers AR, Mathys C, Corlett PR. Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. Science. 2017;357:596-600.

69 Nickel M, Kiela D. Poincaré embeddings for learning hierarchical representations. Adv Neural Inf Process Syst. 2017:6338-47.