Language and Communication in Psychosis: Digital Tools as Novel Opportunities for Biomarker and Intervention

Sunny X. Tang, M.D., Sunghye Cho, Ph.D., Reno Kriz, B.A., Olivia H. Franco, B.S., Jenna Harowitz, B.A., Suh Jung Park, Raquel E. Gur, M.D., Ph.D., Lyle H. Ungar, Ph.D., Mahendra T. Bhati, M.D., Christian G. Kohler, M.D., Monica E. Calkins, Ph.D., Daniel H. Wolf, M.D., João Sedoc, Ph.D., Mark Liberman, Ph.D.

Table 1: Experiment 1
Sample, Access and Use of Technology and Social Media

Variable	HC	CR	PD	p
n	12	22	21	
Age (mean ± SD, yrs)	23.3 ± 4.0	22.0 ± 2.9	24.2 ± 3.4	0.10
Sex				0.03
Female (n, %)	8 (67%)	11 (50%)	5 (24%)	
Male (n, %)	4 (33%)	11 (50%)	16 (76%)	
Access to Technology				
Mobile phone	12 (100%)	22 (100%)	21 (100%)	1.00
Smartphone	12 (100%)	21 (95%)	20 (95%)	1.00
Computer	10 (83%)	20 (91%)	20 (95%)	0.53
Internet	12 (100%)	22 (100%)	20 (95%)	0.60
Social Media Use				
≥ Weekly access	7 (58%)	16 (73%)	13 (62%)	0.62
≥ Weekly posting	3 (25%)	9 (41%)	1 (5%)	0.02
Facebook (ever)	11 (92%)	16 (73%)	15 (71%)	0.40
Twitter (ever)	3 (25%)	5 (23%)	2 (10%)	0.48

Note: HC – Healthy control; CR – Clinical risk for psychosis; PD – psychotic disorder, including schizophrenia, schizoaffective disorder, bipolar I disorder, and unspecified psychotic disorder

Table 1: Experiment 2
Sample and Language Features

	HC	SSD	p value	Cohen's d
Sample				
n	11	20		
Cohort			0.10	
Cohort 1	5	15		
Cohort 2	6	5		
Age (mean years ± SD)	35.6 ± 5.8	36.5 ± 7.2	0.75	0.12
Sex (n, %)				
Female	7 (64%)	9 (45%)	0.32	
Male	4 (36%)	11 (55%)		
Race (n, %)			0.12	
African American	3 (30%)	13 (65%)		
Asian	0 (0%)	1 (5%)		
Caucasian	7 (70%)	6 (30%)		
Education Level	15.8 ± 2.2	13.4 ± 2.5	0.01	-1.00
Recording Characteristics				
Recording Duration (min)	11.6 ± 2.2	12.7 ± 4.5	0.48	0.29
Mean Sentence Length	17.5 ± 3.1	14.4 ± 4.3	0.04	0.81
Word Count	1748.8 ± 448.0	1782.3 ± 908.2	0.92	0.04
Language Measures				
TLC Global Score	0.0 ± 0.0	0.5 ± 1.0	0.13	0.56
TLC Total Score	0.9 ± 1.7	4.4 ± 9.2	0.10	0.46
Next-Sentence Predictability	0.96 ± 0.03	0.94 + 0.04	0.20	-0.44

Note: HC – healthy control participants; SSD – participants with schizophrenia spectrum disorder; TLC – Scale for the Assessment of Thought Language and Communication (Andreasen, 1986). TLC global score is an overall impression of speech and language disturbance based on standard anchors. TLC total score is summed using the published formula, Total = 2*(Sum of items 1-11) + (Sum of items 12-18).

Table 3: Parts-of-Speech Frequencies in SSD and HC

	HC (N=11)	SSD (N=20)	p-value	Cohen's d
Adverb	10.65 (0.95)	8.11 (1.76)	< 0.001	1.66
Determiner	7.50 (0.96)	6.53 (1.25)	0.02	0.83
Pronoun	11.77 (1.47)	13.41 (2.67)	0.04	-0.71
Speech errors	0.05 (0.07)	0.14 (0.16)	0.04	-0.66
Adjective	7.10 (1.57)	6.19 (0.78)	0.06	0.82
Preposition	8.84 (1.37)	7.97 (1.41)	0.11	0.62
Particle	2.65 (0.52)	2.35 (0.50)	0.14	0.59
Conjunction	5.33 (1.35)	4.61 (1.40)	0.17	0.53
Noun	13.16 (0.93)	13.67 (2.16)	0.74	-0.28
Interjection	6.07 (1.66)	6.35 (2.45)	0.75	-0.12
Verb	19.34 (1.74)	19.55 (2.60)	0.79	-0.09

Background / Objectives

- Psychotic disorders (PD), including schizophrenia spectrum disorders (SSD) produce impairments in interpersonal processing, including language
- Interpersonal processing impairment has major negative impacts on functioning and outcomes.
- Funded by the 2018 ASCP Early Career Research Award, this study investigates the opportunity for digital tools to serve as platforms for novel interventions and biomarkers.
- Experiment 1 (Ex1) evaluates access and use of technology and social media in young adults with PD, clinical risk for psychosis (CR) and healthy control (HC) individuals without psychosis symptoms.
- Experiment 2 (Ex2) compares automated natural language processing (NLP) methods for detecting linguistic changes in PD with traditional clinical ratings.

Experiment 1:

- **Participants:** Participants were screened through PERC (Psychosis Evaluation and Recovery Center), an early psychosis intervention program, and LiBI (Lifetime Brain Institute) at the University of Pennsylvania. N=55, Age 18-32 years
- **Instruments:** Participants were surveyed regarding their access to technology and use of social media, specifically Facebook and Twitter, as a part of a larger effort to investigate social media language and usage in individuals with psychosis.
- **Statistical Analyses:** Statistical analyses were conducted in R. Categorical variables were compared among groups Fisher's exact test. Continuous variables were compared using one-way ANOVA. Significance was two-tailed with α=0.05.
- **Results:** There were no significant differences among groups in access to mobile phones, smartphones, computers, or the internet. Social media access rates were similar for all 3 groups. Individuals with psychotic disorders, but not clinical risk, were less likely to actively post at a weekly or higher frequency compared to psychosis free individuals. Decreased active social media posting was unique to psychotic disorders and did not occur with other psychiatric diagnoses or demographic variables. Variation in age, sex, and Caucasian vs. non-Caucasian race did not affect posting frequency

Experiment 2:

- **Participants:** Open-ended interviews were transcribed for two cohorts (Cohort1 = 15 SSD + 5 HC; Cohort2 = 5SSD + 6HC). SSD was not enriched for presence of thought disorder.
- **Clinical Language Evaluation:** Participant speech was rated by a blinded expert clinician on published anchors from the Scale for the Assessment of Thought Language and Communication (Andreasen, 1986).
- **Natural Language Processing:** Sentence-parsing and part-of-speech tagging were completed with spaCy. Sentence embeddings and probabilities of participant-spoken sentences following both participant- and interviewer-spoken sentences were calculated using Bidirectional Encoder Representations from Transformers (BERT).
- **Statistical Analyses:** Statistical analyses were conducted in R. Language measures departed from normality and were compared with the Wilcoxon rank sum test. Other continuous measures were compared between groups with Student's t-test. Categorical variables were compared with Chi-squared test. Leave-one-out cross validation was used in the Naïve Bayes models predicting group categorization.
- **Results:** In this pilot sample un-enriched for thought disorder, NLP methods were significantly better than a standardized clinical rating scale at detecting subtle language differences in individuals with SSD. NLP features reflect difference parts-of-speech frequencies, word choices, and increased sentence embedding distance in SSD interviewer-participant exchanges.

Conclusions

- The results encourage further development of internet and social media-based interventions and treatment monitoring for young people with psychosis.
- Lower active engagement may reflect impairments in social cognition and functioning.
- NLP tools show promise for sensitive discernment of a linguistic biomarker in psychosis.
- Linguistic biomarkers can potentially be automatically and objectively extracted from both digital and non-digital sources to aid in diagnosis and monitoring treatment effect.

Disclosure: Dr. Tang is a consultant for Winterlight Labs and holds an equity position at North Shore Therapeutics. The other authors have no financial disclosures.
Funding Early Career Research Award, American Society of Clinical Psychopharmacology; T32 MH019112 (SXT);
PERC supported by the Substance Abuse and Mental Health Services Administration (SAMSHA); Microsoft Research Dissertation Grant (JS);
DARPA grant HR0011-15-C-0115 (the LORELEI program), MH-62103, and M01RR0040.
Acknowledgements: We thank the participants, as well as Salvatore Giorgi, M.A., Thomas Hohing, Zeeshan Huque, Suhjung Park, Eehwa Ung, Lyndsay Schmidt, LaTonya McCurry, Victoria Pietruszka and Kosha Ruparel from the University of Pennsylvania.
Corresponding Author: Sunny X. Tang, M.D. - stang3@northwell.edu

Figure 1: Group Effects on Clinical Language Ratings and BERT Next-Sentence Probability

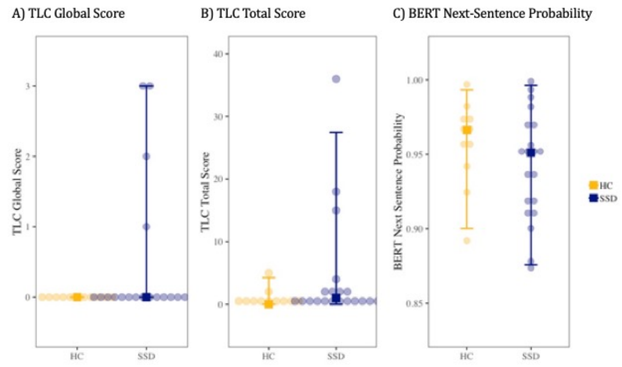

Figure 1. Group Effects on Clinical Language Ratings and BERT Next-Sentence Probability: Median and interquartile range are displayed for SSD (blue) and HC (yellow) participants. There were no significant group differences for: A) Global summary score from the Scale for the Assessment of Thought Language and Communication (p=0.13, Cohen's d=0.56; Andreasen 1986). B) Total TLC summed per the published formula (p=0.10, Cohen's d=0.48). C) Next-sentence probability score calculated using Bi-directional Encoder Representations from Transformers (BERT; p=0.20, Cohen's d=0.44). All sentence pairs where the second sentence was spoken by the participant were included. Per Shapiro-Wilk tests, the distributions were significantly non-normal. Comparisons were made between groups with Wilcoxon rank sum tests.

Figure 2: Sentence Embedding Distance by Interviewer-Participant Exchanges

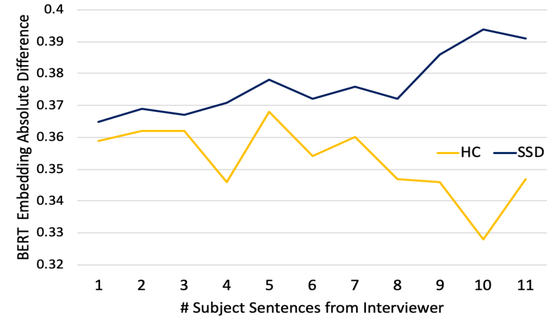

Figure 2 – Sentence Embedding Distance by Interviewer-Participant Exchange:. Sentence embeddings were calculated for each sentence in each interviewer-participant exchange (when the interviewer speaks, and then the participant responds). SSD response embeddings deviated significantly farther from initial interviewer prompts than HC responses.

Figure 3: Using Language Features to Predict SSD Group Status

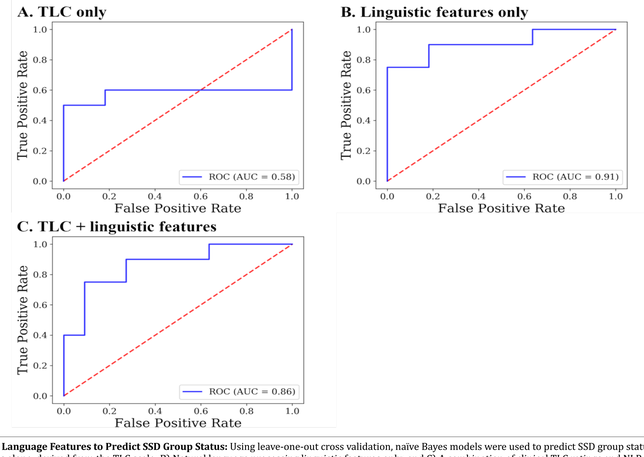

Figure 4 -Using Language Features to Predict SSD Group Status: Using leave-one-out cross validation, naïve Bayes models were used to predict SSD group status using A) Clinical linguistic features alone, derived from the TLC scale; B) Natural language processing linguistic features only; and C) A combination of clinical TLC ratings and NLP features.