Automated assessment of PANSS interview quality

Submission ID 3000983

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

Methodological Issue Being Addressed Can PANSS interview quality, measured as clinician adherence to the SCI-PANSS script, be quantified accurately using automated analysis of audio recordings of the interview?

Introduction The Positive and Negative Syndrome Scale (PANSS) is often relied on as the primary clinical endpoint in schizophrenia clinical trials. Given its role in the assessment of treatment efficacy, standardization of the interview as administered by the clinician is paramount. For this reason, the PANSS interview is often audio recorded for subsequent review by an independent clinician to ensure high interview quality. However, it is impractical to conduct manual reviews of all PANSS interviews in a clinical trial, particularly in a timely enough fashion to correct for poorly conducted interviews. Here, we use natural language processing (NLP) to analyze audio recordings of PANSS interviews to test an automated assessment of interview quality.

Methods Data collection

Two PANSS interviews were conducted in English on a healthy volunteer by a trained clinical psychologist: In the first interview (‘high quality’), the clinician was careful to conduct the interview as specified in Structured Clinical Interview – Positive and Negative Syndrome Scale (SCI-PANSS) and adhere to scripted prompts; In the second interview (‘low quality’), the clinician, while remaining true to the original intention of the prompts, purposefully adjusted the specific wording used by either shortening, summarizing, or adding additional language to the original prompts (e.g., “How anxious have you been feeling in the past week?” instead of “Have you been feeling nervous or worried in the past week?”). Purposefully, both interviews followed the same prompt logic so that the interviews would be comparable.

Data processing and analysis

The audio for each ‘high quality’ and ‘low quality’ interview was processed. First, the speech was transcribed into text using Amazon Transcribe. As part of this transcription, speaker diarization was used to label each speaker. Second, each speaker was programmatically identified as either the clinician or the participant. For the purposes of this analysis, the participant audio/speech was discarded. Then, individual ‘turns’ of the clinician’s speech were matched to PANSS prompts. This was done using a pre-trained MiniLM model, a sentence-transformer that maps text to a 384 dimensional dense vector space for comparison with other text. Purposefully, only the 58 prompts within the SCI-PANSS not subject to skip logic were matched. Using the same method, we quantified the ‘adherence’ of the clinician’s speech to the prompts as specified in the SCI-PANSS. We then compared prompt adherence in the ‘high’ and ‘low’ quality interviews.

Results The average adherence to each of the 58 prompts in the ‘high’ and ‘low’ quality interview was 0.90 and 0.65 respectively. The difference, as evaluated by a Mann-Whitney U test, was significant, with a p-value <0.001.

Conclusion We demonstrate the ability of NLP-based analysis of PANSS interview audio recordings to quantify prompt adherence and subsequent interview quality. In the future, we hope to (1) replicate this on a larger bank of pre-labeled interviews, (2) expand the metrics to include pre-prompt pause duration, rate of speech, total length of clinician speech, emotional valence, presence of external distractions, etc., and (3) expand these analyses to other structured interviews such as the MADRS and the UPDRS.

Co-Authors

* Presenting Author

First Name	Last Name	Affiliation
Georgios	Efstathiadis	Harvard T.H. Chan School of Public Health, Harvard University; Brooklyn Health
Michelle	Worthington	Department of Psychiatry, Yale School of Medicine; Brooklyn Health
Vijay	Yadav	Brooklyn Health
Anzar *	Abbas *	Brooklyn Health

Keywords

Keywords
digital biomarkers
schizophrenia
PANSS
rater quality

Guidelines I have read and understand the Poster Guidelines

Disclosures if applicable All authors are employees of Brooklyn Health, a privately held company that developed the software evaluated in this abstract.