AI-based assessment of clinical interview quality

Georgios Efstathiadis[1,4], Michelle Worthington[2,4], Vijay Yadav[3,4], Anzar Abbas[4]

[1] Harvard University, Cambridge, MA, [2] Yale University, New Haven, CT, [3] University of New South Wales, Australia, [4] Brooklyn Health, Brooklyn, NY

Background

- PANSS is a primary clinical endpoint in SCZ trials
- Proper PANSS administration is critical for trial success
- To ensure high quality, interviews are often recorded
- Recordings are then manually reviewed for quality
- **However, this is both impractical and ineffective**

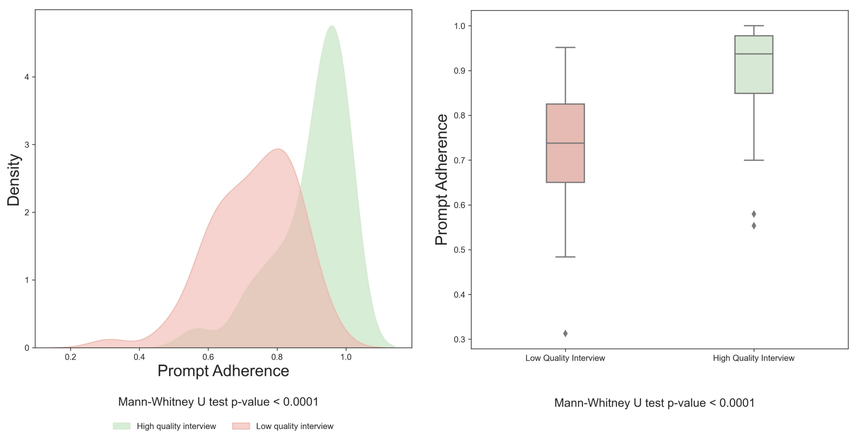
Objective

Use of natural language processing to automatically evaluate assessment quality from recordings of PANSS clinical interviews

Results

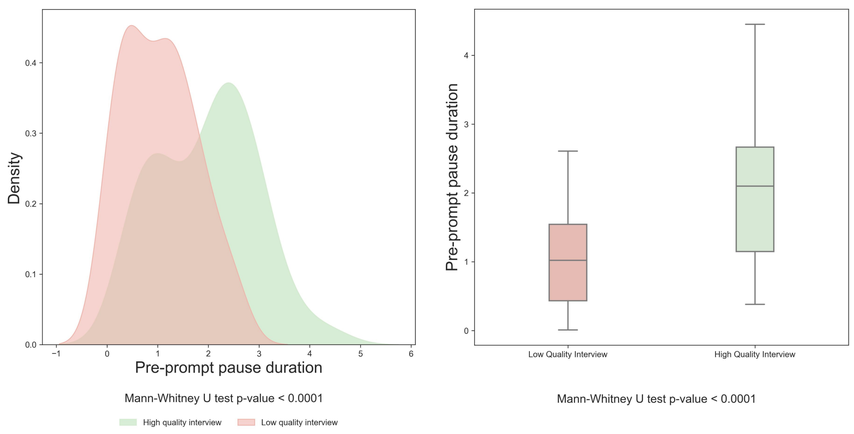
Comparison of high vs. low-quality interviews

Rater adherence to prompts



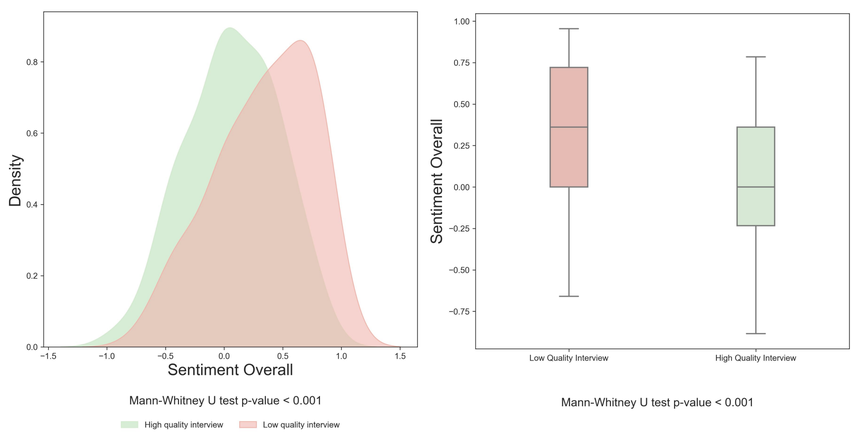
The low-quality interview, with intentional deviations from the script but maintenance of the original intent, showed significantly worse adherence to the 59 prompts not subject to skip logic in the PANSS.

Rater pauses between questions



The low-quality interview, where the rater intentionally rushed through the prompts, showed significantly shorter pauses between questions compared to the high- interview, as well as some interruptions.

Emotional valence of rater speech



The low-quality interview, peppered with encouraging interjections like "good, let's move on", led to a significantly higher emotional valence compared to the high-quality interview, where valence centered at 0.

Methods

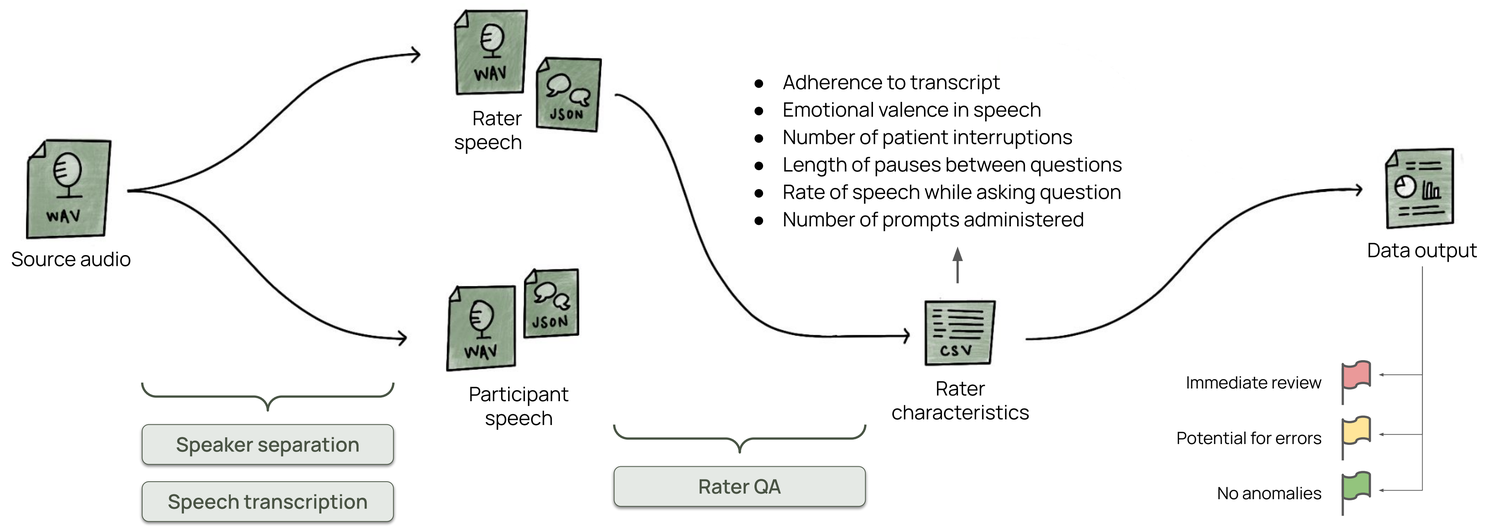
Data collection

- Two clinical interviews conducted by a trained clinical psychologist
  - A 'high quality' interview, where the SCI-PANSS was followed by the book
  - A 'low-quality' interview, where bad practices were intentionally used
- Both interviews followed the same prompt logic so they would be comparable.

Data processing and analysis

1. The recording is split to separate clinician and rater speech
   a. The interview is transcribed and the transcript is separated by speaker
   b. Given knowledge of expected PANSS prompts, each speaker is identified
2. The rater speech transcript is analyzed for a list of rater characteristics features
3. The features are comapred between the low and high quality interview

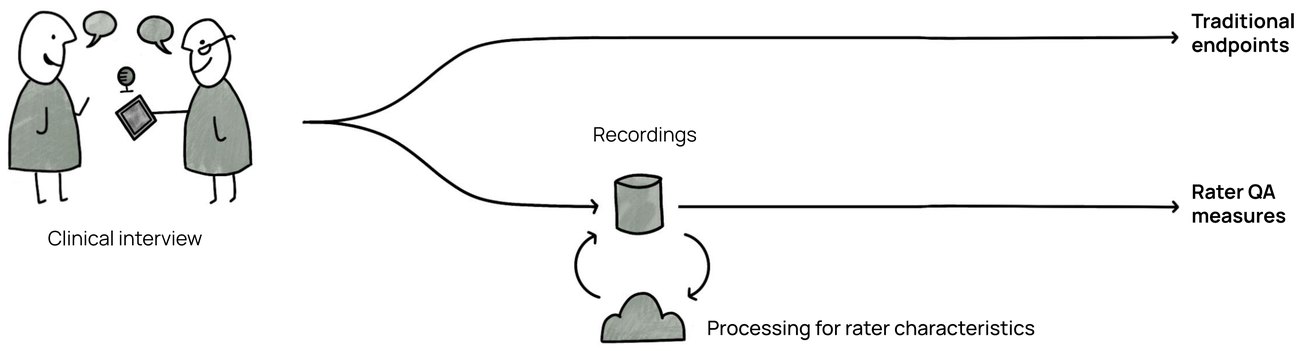
For a full description of the methods, see OpenWillis documentation on www.github.com/bklynhlth/openwillis



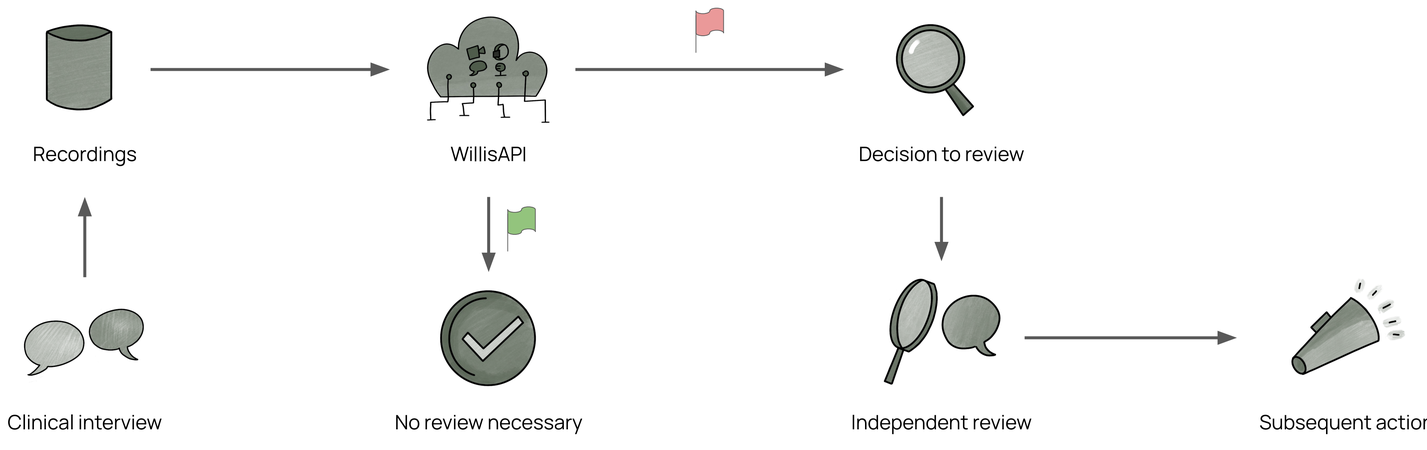
Use case

Automated, real-time interview quality flagging for secondary review

Process interview recordings, typically collected for manual rater QA, for rater behavior



Flag recordings for immediate review; reduce number of interviews that need review



ISCTM 2023 Autumn Conference
includes joint day with ECNP
5-7 October 2023 Barcelona, Spain

Brooklyn Health

Presenting Author
Anzar Abbas, PhD
anzar@brooklyn.health