Standardizing Audio and Video Capture to assess biomarkers in Clinical Trials – Avoiding Garbage in/Garbage out

Submitter Daniel DeBonis

Affiliation EMA Wellness

SUBMISSION DETAILS

I agree to provide poster pdf for attendee download. Yes

I have used the poster abstract template to develop my abstract. Yes

Methodological Issue Being Addressed Developing a rigorous, standardized protocol for collecting audio and video data across clinical sites to enable high-quality LLMs scoring models, quantitative biomarker analyses, including voice, speech, and non-verbal cues.

Introduction At the ISCTM Biomarker Working Group meeting in Amsterdam (Autumn 2025), participants raised a critical question: are there established standards for capturing audio and video during clinical interviews used in CNS research? Consensus among attendees was that no unified, field-accepted standards currently exist.

The absence of standardized recording practices introduces substantial variability across study sites. Differences in interview rooms, microphone characteristics, camera placement, lighting, and acoustics contribute to inconsistent audio-video quality. These inconsistencies directly affect rater quality control, eligibility review on scales such as MADRS and PANSS, and the reliability of downstream biomarker analyses. Poor recording quality is also a frequent source of transcription failure and degraded performance for large language models (LLMs) increasingly used for interview review and scoring.

Methods This work integrated acoustic engineering principles, prior validation studies, and practical constraints of CNS trial environments to derive minimum, implementable standards for audio and video capture. In addition, a small pilot dataset was generated to assess whether recordings obtained under these standards measurably improve automated transcription quality relative to typical site recordings.

The proposed standards were derived by mapping common sources of recording failure—such as spectral distortion, inconsistent framing, and variable signal-to-noise ratios—to established requirements for reliable speech analysis, transcription, and video-based feature extraction. Prior work from the Bridge2AI-Voice Consortium demonstrated that voice recordings meeting basic acoustic criteria are sufficient for biomarker analysis; however, recommended headset-based configurations are impractical for structured clinical interviews. The present framework adapts these validated acoustic requirements to CNS interview settings and extends them to video capture.

Results Audio Standards:

Microphones must exhibit a flat frequency response within ±3 dB across the 50–5000 Hz speech bandwidth to preserve clinically relevant vocal features such as speech rate, prosody, and pause structure. Multiple commercially available tabletop microphones under $100 meet these specifications and are compatible with rater-administered interviews without interfering with rapport or scale administration.

Video Standards:
Video capture must meet a minimum of 720p HD resolution, with stable device positioning, consistent lighting, and fixed participant distance. Because tablets and laptops used for eCOA entry cannot reliably maintain static framing, a dedicated, mounted webcam tethered to the eCOA device is recommended. Numerous commercially available webcams under $50 exceed these minimum requirements.

Testing and Quality Control Procedures:
A brief pre-interview verification protocol is proposed. Audio is validated using a microphone frequency response check via device-appropriate audio analysis software. Video consistency is verified using auto-framing or face-centering tools to ensure stable participant positioning. These checks are integrated into routine visit workflows and supported by rater training materials and visual setup guides.

Pilot Comparison Testing:
To evaluate the practical impact of these standards, ten test interviews with healthy volunteers (3–5 minutes each) were recorded using the specified microphone, webcam, and setup procedures. Automated transcription was performed an Azure speech-to-text model. Transcription quality was assessed using the model's native transcription confidence metric, reported on a 0–1.0 scale.

Across the ten test interviews, transcription confidence scores ranged from 0.95 to 1.0 for both models. As a benchmark, the same transcription model was applied to two professionally recorded, studio-quality interviews used in a rater training program; confidence scores for these recordings were uniformly 1.0. By comparison, a randomly selected sample of 20 site-recorded interviews used for an analysis presented at ISCTM (Fall 2025) demonstrated substantially lower transcription confidence, with a mean score of approximately 0.75.

Conclusion There is currently no unified standard for audio and video capture in CNS clinical trials, despite their growing importance for rater quality control, transcription-dependent LLM scoring, and biomarker research. Pilot data suggest that recordings obtained using the proposed standards approach studio-quality transcription confidence and substantially outperform typical site recordings. As the field continues to adopt LLM scoring methodologies, study sponsors will need high confidence in their recordings and transcriptions. The same principles apply for human reviews of audio and video recorded interviews.

While audio biomarkers are already supported by readily available analytic models, video-based biomarker analysis remains in earlier stages of development. The adoption of recognized, minimum video capture standards may help accelerate progress in this area by enabling more consistent

data collection and cross-study comparability. We propose submission of these recommendations to the ISCTM working group for refinement and potential adoption at the 2026 annual meeting

Co-Authors

Suzanna Newton[1], Pat Harrington[1], **Daniel DeBonis[1]**

[1] EMA Wellness

Keywords

Keywords
Biomarkers
Audio Video Capture
Standards

Guidelines I have read and understand the Poster Guidelines

Disclosures All are full time employees of EMA Wellness. Dan DeBonis is an equity holder in EMA Wellness and Signant Health.