ISCTM
International Society for CNS Clinical Trials and Methodology

Defining and Standardizing Vocal Biomarker Outcomes Working Group

Co-chairs: Alex S Cohen, Jan A. Sedway

ISCTM 22nd Annual Meeting, Washington, D.C.,
February 18, 2026

Conflicts & Declarations

Alex Cohen
Professor at Louisiana State University in the Department of Psychology and the Center for Computation & Technology
Part owner of Quantic Innovations, which provides digital phenotyping solutions for clinical research/care.

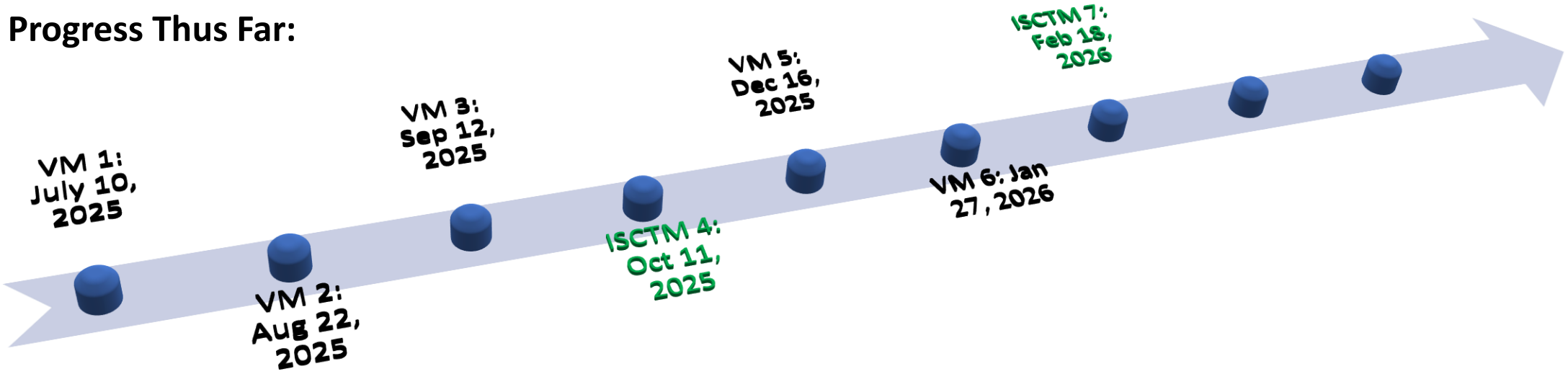
Jan Sedway
Share holder and employee for NetraMark Holdings, Inc; contractor for Clario.

ISCTM Working Group Summary

Working Group Objectives:

1. Discuss the various methods of collecting speech latency, analyze the evidence for each approach, come to a consensus on FDA program approach (i.e., COA Qualification, MIDD, CPIM, ISTAND, etc)
2. Write a manuscript detailing the above process and decisions
3. Determine and document the additional steps needed for FDA approach

Progress Thus Far:



Subgroups / Manuscript Section Leads

Introduction:	Jan Sedway & Alex Cohen
Background/Literature Review:	Anhye Kim & Chris Brady
Potential Mechanisms: Cognition:	Hardik Kothare & Bill Simpson
Definitions & Measurement:	Leif Simmatis & Anzar Abbas & Sunny Tang
Regulatory Considerations & Validation Strategies:	Ni Kim & Marc Cantillon
Other COUs:	Michael Spilka & Ryan Berry
Next Steps (data sharing):	Derek Buhl & Bill Simpson

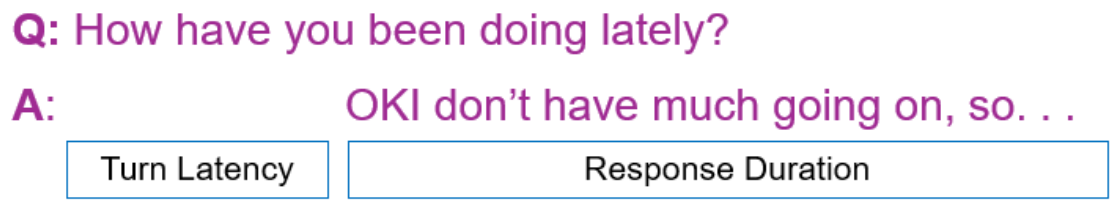
Ground Rules

- Each group will briefly present [~5 minutes]
- Hold questions until after presentations

Speech Latencies & Turn Latencies

Speech Latencies: General term for speech "pauses" in between meaningful events (e.g., stimulus, response).

Turn Latencies: Initial pause separating different speakers.



Construct (Signal): Reflects an integration of cognitive, social cognitive and motivational systems.

Why Turn Latencies?

- Interpretable as a reaction time, in milliseconds
- Pauses (more generally) reflect one of the most replicable speech features tied to psychopathology (i.e., Alzheimer's dementia, Schizophrenia negative symptoms, Depression).
- Can experimentally manipulate cognitive load to see changes in pauses.
- Interviews contain many turn latencies, leading to reliable (internally consistent) estimates.
- Can be obtained from audio recordings of clinical interviews
- Potentially generalizable across people, languages, socio-demographic factors (at least, compared to language)

Consensus Context of Use

Consensus Primary COU:

- Turn Latencies derived from a clinical interview.
- Solely applied to enrichment: with the purpose of minimizing placebo response & identifying patients with appropriate disease state
- Solely used for depression v Bipolar-depression

Section 3: Background/Literature Review

Chris Brady, PsyD & Anhye Kim, MD, PhD

Existing evidence for using (turn) latencies to understand depression

- **Consistently observed symptoms in depressive state across time and culture**
  - Delayed verbal responses have long been observed in depressive states across time and cultures, from early clinical descriptions to modern quantitative speech analyses.(Ref: Kraepelin, 1921; Hamilton, 1960; Parker & Hadzi-Pavlovic (CORE), 1996; Sobin et al.(MARS), 1998; Yamamoto et al., 2020; Cohen et al., 2024)
- **Diagnostic marker: Depression vs Healthy or other psychiatric ds**
  - Diagnostic utility: Depression vs Healthy or Non-depressed (Ref: Xu et al., 2019; Yamamoto et al., 2020; Cohen et al., 2024)
  - Differential diagnosis: Depression vs Bipolar disorder (Ref: Yamamoto et al., 2020) or Schizophrenia (Ref: Xu et al., 2019)
- **Monitoring marker: severity, detect changes**
  - Cross-sectional: Correlation with severity (HAMD-17) (Ref: Yamamoto et al., 2020; Cohen et al., 2024)
  - Longitudinal: Track changes in depression severity (Ref: Yamamoto et al., 2020; Cohen et al., 2024)
- **PD & Response marker**
  - limited and requires further investigation

Discussion point

- **Main message of bullet #1 (Background)**

- Focused on the positive findings in the significant correlations btw turn latency & depression

- Should pause time be included within the scope of this section?

- **Speech latency (average or total pause time) <= Turn latency (response time) + Word latency**
- **Title**: Defining and Standardizing Speech Latencies for Clinical Trials
- **Purpose:** To develop an objective, reliable and valid measure of patient characteristics for identifying subgroups related to trial appropriateness based on "turn latencies"

- Should negative or non-significant findings also be addressed for balance?

- **Even if not fully incorporated into this section, it may be worth to understand contextual limitations across studies with negative or null findings (Study design, data sources, interview settings, measurement methods, demographic factors etc.)**
  --> relevant to "Definition and measurement section"

- Are additional latency-related constructs relevant to include?

Psychometrics of studies using turn latencies for enrichment

Construct Definition & Content Validity

- Literature establishes a measurable relationship between turn latency and depression along with discriminate validity, however, clinical trials using this available resource are limited to 5 studies

Validity

- Stable when symptoms are stable and shifts with symptom change (Cohen et al., 2024)
- Aligned with psychomotor retardation and cognitive slowing with depression
- Convergent validity based on correlations between latency and traditional measures MADRS (Mundt et al., 2007; Mundt et al., 2012; Cohen et al., 2024; Seigel et al., 2024); QIDS-C, HAM-D, QIDS-SR (Mundt et al., 2012); PANSS negative symptoms (Cohen et al., 2025).
- Generalizes across languages with calibration (Cohen et al., 2024)

Reliability

- High internal consistency with coefficients from 0.94 to 0.97 (Cohen et al., 2024) and 0.98 (Seigel et al., 2024)
- Moderate test–retest reliability with ICC values of 0.67 to 0.70 (Cohen et al., 2024)
- Consistent reliability findings across all five studies

Sensitivity and Specificity for Enrichment

- Excluding "normal-latency" participants leads to stronger treatment effects with smaller sample (Cohen et al., 2024; Seigel et al., 2024, and Cohen et al., 2025)
- Turn latency predicts depression severity with strong performance: sensitivity 0.77 and specificity 0.94 (Cohen et al2024)
- Cross disorder psychometric robustness based on data from studies of individuals with active depression (bipolar) and schizophrenia (Cohen et al., 2024; Seigel et al., 2024, and Cohen et al., 2025)

Discussion Points

Psychometric Considerations

- Is validation compared to Structured Interviews scores enough?

Overall Considerations

- Standardization of language model and analysis?
- Standardized elicitation tasks and clear latency definitions are essential for reproducibility.
- Device type, noise, sampling rate, and overlap handling influence measurement precision.
- Ethical considerations include privacy, consent, and transparent automated processing.
- Key gaps include global normative datasets, preregistered validation, and language-aware calibration.

Section 4: Possible Mechanisms, Cognition

Hardik Kothare and Bill Simpson

Coverage for today's session

- Review the areas of focus for the section
- Are there any concepts we're missing?

Why cognition as a primary mechanism?

•Conversational turn-taking requires rapid cognitive processing, response selection,and initiation. Slowed processing or impaired cognitive-motor integration plausiblyincreases turn latencies.

•*Harrison, J. E. (2019). Screening and measurement of cognitive impairment in psychiatry. CNS Spectrums, 24(1), 144–153*

•**Mechanism hypothesis:** Turn latency serves as a behavioural manifestation of slowedcognitive processing and/or delayed response initiation.

Empirical studies show a pattern of cognitive deficits associated with speech and language production in schizophrenia and major depressive disorder (e.g., Schneider et al., 2023).

Both schizophrenia and depression show slowed processing speed, delayed response initiation, and impaired cognitive-motor integration. Possible neurobiological substrates: reduced cortical inhibition (Lefebvre et al., 2024) and reduced striatal dopamine transmission (Leong et al., 2025).

Example findings for depression

- **Psychomotor Retardation**

•Psychomotor retardation (slowed speech, movement and cognitive processing) is a definingfeature of major depressive disorder. Slowed speech and increased response latency trackdepressive symptom severity.

•*Buyukdura et al., 2010; Sobin et al., 1997*

- **Speech Timing and Clinical Correlates**

•Objective measures (speech pause time, speech rate) show longer pauses and slower speech in depressed individuals.

•Improvements in symptom severity correlate with normalization of speech timing, supporting a cognitive-motor basis.

•*Mundt et al., 2007; Wiseman et al., 2025*

Example findings for schizophrenia

•**Psychomotor and cognitive slowing:** Psychomotor slowing is a clinically observable and enduring feature ofschizophrenia, affecting motor initiation, planning, and execution. Slowed cognitive processing, planningdeficits, and impaired response selection contribute to prolonged behavioral output.

•**Conversational turn-taking:** Empirical studies show longer gaps and pauses between speaker turns in freeconversation and semi-structured interviews. Pauses and turn-timing abnormalities correlate with negativesymptoms, blunted affect, and psychomotor retardation.

- *Morrens et al., 2007; Walther & Mittal, 2016; Fauviaux et al., 2025; Lucarini et al., 2022, 2024; Cohen et al., 2024*

Discussion

•Our section will summarize data examining cognitive-motor integration (psychomotor retardation) and processing speed as key cognitive concepts

- Are there any concepts we're missing?
- Should we include a discussion around "normal" and how specific these cognitive differences are to depression/schizophrenia?

Section 5: Definitions and Measurement

Leif Simmatis & Anzar Abbas & Sunny Tang

Summarization/analysis

- Speech turn latency represents the duration of a specific cognitive sequence, and thus serves as a proxy for brain network functioning
  - Auditory reception, information processing, response generation, and vocal initiation
  - Psychiatric illnesses often disrupt relevant brain networks
  - Variance also arises from cultural and demographic factors
  - Time of day effects, including age modulation (Veneman et al., 2013; Iskandar et al., 2016)
- Turn-taking can be quantified as the gap between the end of one speaker's bout and the start of another speaker's, sometimes also called floor transfer offset (FTO) (Levinson and Torreia, 2015)
- Measures include gaps, overlaps, and pauses (total number relative to the total number of possible turn transitions, expressed in %); temporal properties (the median duration expressed in ms, and the total duration relative to the total speaking duration, expressed in %) (Fauviaux et al., 2025)
  - Requires transcription and diarization

Recording/interview methodology

- Speech convergence (i.e., alignment of speaking characteristics) between interviewer and interviewee might motivate careful consideration of interview style
- Explicit question-answer pairs can also be valuable for promoting reliable measurement, with measurement of speaker onset being a critical feature
- “Sequence organization” and complexity of questions are also related to subsequent turn-taking patterns and timing (Roberts et al., 2015)

Recording conditions

- Ideally multiple microphones; directional condenser microphones mounted close to the mouth (e.g., Lucarini et al., 2024) to ensure even loudness and to avoid audio contamination
  - Improves the reliability of speaker identification (Lucarini et al., 2024)
- Automated analysis include voice activity detection and speaker change detection to assist in manual annotation of records (Riad et al., 2022)
- Acoustic isolation to aid speaker identification and diarization
- Teleconferencing software may lead to exaggerated turn exchange times, potentially due to interruption of normal conversational cadence via transmission delays (Boland et al., 2022)
- High-quality audio data (44kHz, uncompressed) are preferred (e.g., Fauviaux et al., 2025)

Software/measurement approaches

- Diarization is used to extract the identities of individuals in a conversation
  - Aided by voice activity detection, speaker identification technologies using machine learning
- Manual annotation can be used as an additional/ alternative approach
  - Specialized software like Praat, Audacity, etc. can be used to insert timestamps and/or correct errors made using automated methods

Section 7 – Other COUs

Michael Spilka and Ryan Berry

Summary

This section will describe other potential contexts and use cases for speech latencies in clinical trials, including:

1. Relevant indications other than depression
2. Other modalities to elicit speech and other speech latency variables
3. Other potential uses of turn latency besides enrichment
4. Brief discussion of the speech latency construct stability/variability across COUs

1) Relevant indications other than depression

- **Schizophrenia**
  - Relevance of vocal biomarkers in SZ, speech latency and negative symptoms
- **Dementia and/or transdiagnostic marker of cognitive impairment**
  - Speech research in dementia
  - Potential use beyond to dementia to transdiagnostic marker of cognitive impairment in CNS disorders?
- **Other psychiatric indications and/or comorbidities**
  - Mania in bipolar disorder: opposite end of mania-depression dimension for turn latency?
  - Anxiety: summary of relevant research and discussion of implications for comorbid depression and anxiety?

2) Other modalities to elicit speech and other speech latency variables

- **Other modalities**
  - Clinical interviews conducted outside the clinic (e.g., telehealth)
  - Other types of interviews/conversations (e.g., digital avatars)
  - Does turn latency generalize across different sources of speech?
- **Speech latency variables**
  - Are there suitable speech latency alternatives to turn latency (e.g., speech rate) that can be used for non-conversational speech tasks (e.g., a picture description task)

3) Use cases other than enrichment

- **Patient stratification/precision medicine**
  - E.g., depression subtypes, apathy/negative symptoms
- **Speech latencies as surrogate endpoints**
- **Speech latencies for quality monitoring of clinical ratings?**
  - E.g., discrepancy between speech latencies and clinical ratings to detect rater issues
  - Can a similar approach be applied to detect placebo response?

4) Brief discussion of the speech latency construct stability/variability across COUs

- **Does turn latency reflect the same underlying construct across indications, speech sources, uses cases?**
  - Should we expect it to? Does it need to?
  - Recommendation for focus on ensuring vocal biomarker is fit-for-purpose, regardless of generalizability across contexts

Section 6: Regulatory

Ni Khin and Marc Cantillon

FDA Regulatory Pathways for Vocal Biomarkers in MDD

Pathway	Full Name	Purpose	Timeline
CPIM	Critical Path Innovation Meeting	Early discussion forum; non-binding scientific exchange	1 meeting
MIDD	Model-Informed Drug Development	Exposure-response modeling; trial design optimization	2 meetings
BQP	Biomarker Qualification Program	Full qualification for multi-program use	3 stages (LOI, QP, FQP)
ISTAND	Innovative Science and Technology Approaches for New Drugs	Digital health tech not fitting biomarker/COA categories	3 stages (pilot)

Enrichment Strategy for MDD Trials (All Arms)

Enrichment Type	Application	Advantage
Prognostic	Enroll patients with vocal biomarkers predicting higher likelihood of response	Increases effect size
All-Arms Design	Apply same criteria to drug AND placebo groups	Balanced comparison

Context of Use (COU): Vocal acoustic features (speech pause time, prosody) as prognostic biomarkers for Major Depressive Disorder treatment response

Key Considerations & Pathway Selection

Critical Questions

- Which pathway fits development stage?
- Is analytical validation sufficient?
- How does enrichment affect label breadth?
- What is the proposed COU?

Pathway Recommendations

- **Early stage:** CPIM for FDA feedback
- **Modeling phase:** MIDD for trial optimization
- **Multi-program:** BQP for full qualification
- **Single-program:** IND pathway

EMA Qualification Pathway for Vocal Biomarkers

EMA Pathway	Full Name	Purpose	Outcome
QoNM	Qualification of Novel Methodologies	Biomarkers, digital health tech, imaging, COAs	QO or QA/LOS
QO	Qualification Opinion	Full qualification with public consultation	Published by EMA
QA/LOS	Qualification Advice / Letter of Support	Methods not ready for full qualification	Encourages development

Key Difference: EMA pathway managed by CHMP (Committee for Medicinal Products for Human Use). 8-9 biomarkers fully qualified since 2008. Must comply with GDPR for digital biomarkers.