A Bayesian Multimodal Framework Integrating Digital Biomarkers and Clinician-Rated Measures for Expressive Negative Symptoms in Schizophrenia

Khan A[1,3]; Lindenmayer JP[2,3]; Insel BJ [1,3]; Kothare H[4]; Neumann M[4]; Ramanarayanan V [4]; Yavorsky C[6]

1: Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY; 2: New York University, New York, NY; 3; Manhattan Psychiatric Center, New York, NY; 4: Modality.AI, San Francisco, CA; 5: University of California San Francisco, San Francisco, CA; 6: Valis Biosciences Inc, Great Neck, NY

METHODOLOGICAL QUESTION

Negative symptoms remain among the most disabling features of schizophrenia and are typically assessed using clinician-rated scales that require extensive training and demonstrate variable inter-rater reliability. Digital biomarkers derived from speech behavior offer objective measurement opportunities but are generally evaluated using conventional predictive modeling approaches that provide limited support for clinical decision-making. We developed a Bayesian decision analysis framework that integrates multimodal acoustic biomarkers, PANSS Expressive Negative Symptom scores, and BNSS expressive symptom ratings to estimate individualized posterior distributions of expressive negative symptom severity. The approach is intended to supplement rather than replace clinician-administered assessments by combining objective digital measures with established clinical rating scales to reduce uncertainty in symptom severity estimation.

BACKGROUND

Recent studies have demonstrated associations between acoustic speech features and expressive negative symptoms, particularly alogia. However, translation of these biomarkers into clinically actionable tools remains limited. The authors prior work demonstrated that individuals with treatment-resistant schizophrenia (TRS) differed from healthy controls on acoustic features including Cepstral Peak Prominence (CPP), canonical timing alignment, reading duration, and spontaneous speech duration, and that several speech measures were associated with PANSS and BNSS expressive symptom ratings.

This analysis extends those findings by evaluating whether digital biomarkers can be integrated with clinician-rated measures within a Bayesian framework to generate individualized estimates of expressive negative symptom severity while explicitly quantifying uncertainty.

OBJECTIVES

- To evaluate whether integrating clinician-rated measures of expressive negative symptoms with speech and facial-expression digital biomarkers improves the identification and classification of expressive negative symptom severity in individuals with treatment-resistant schizophrenia
- To determine whether a Bayesian multimodal framework can generate individualized probabilities of expressive negative symptom severity, improve discrimination compared with clinician-only or digital-only models, and quantify uncertainty to support clinical-trial assessment and future digital-endpoint development.

PARTICIPANT DEMOGRAPHICS

Characteristic	TRS (n=67)	HC (n=63)	P value
Age, years; mean ± SD	43.0 ± 12.0	33.9 ± 10.2*	<0.0001
Age range, years	21–67	20–63	
Sex	n (%)	n (%)	0.68
Male	52 (77.6)	46 (73.0)	
Female	15 (22.4)	17 (27.0)	
Race	n (%)	n (%)	<0.0001
Asian	4 (6.0)	20 (31.7)	
African American	38 (56.7)	17 (27.0)	
Caucasian	9 (13.4)	19 (30.2)	
Unknown	16 (23.9)	7 (11.1)	
Note	*Age missing for one HC.*		

ANALYTIC FRAMEWORK

Sample: The analysis included 67 participants with treatment-resistant schizophrenia who had complete clinician-rated and digital assessment data. Healthy controls were not included because participant-level control data.
Reference Outcome: Expressive negative symptom (ENS) severity was defined using the PANSS Expressive Negative subfactor:

$$\text{PANSS ENS} = N1 + N3 + N6 + G7$$

where N1 = Blunted Affect, N3 = Poor Rapport, N6 = Lack of Spontaneity and Flow of Conversation, and G7 = Motor Retardation. For this exploratory analysis, higher ENS was defined as a score ≥12, corresponding to the sample median; 38 of 67 participants met this criterion.
Predictors: Digital predictors comprised reading-task CPP, speaking duration, canonical timing alignment, and percentage pause time, together with spontaneous-speech duration and F0 variability. Sixteen facial-behavior variables were standardized and reduced to four principal components within each cross-validation training fold to limit dimensionality and prevent information leakage. Clinical predictors were BNSS Alogia and BNSS Blunted Affect scores.

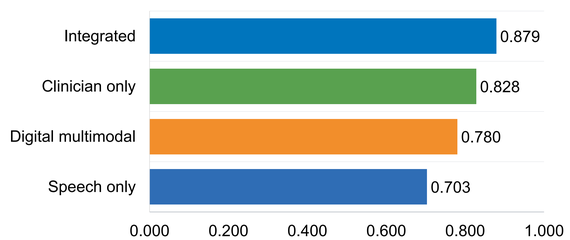
Bayesian Model: A Bayesian probit model estimated each participant's probability of higher ENS:

$$P(Y_i = 1 \mid X_i) = \Phi(\alpha + X_i\boldsymbol{\beta})$$

where $Y_i = 1$indicates PANSS ENS ≥12, $X_i$contains the clinical and digital predictors, and $\Phi$is the standard normal cumulative distribution function. Weakly informative priors were specified for the intercept and standardized regression coefficients. Posterior coefficients were estimated using four Gibbs-sampling chains with 18,000 retained draws; all $\hat{R}$values were ≤1.003, indicating satisfactory chain convergence. Discrimination was evaluated using repeated stratified five-fold cross-validation with 20 repeats. Models were compared using area under the receiver operating characteristic curve, accuracy, balanced accuracy, sensitivity, and specificity. AUC 95% confidence intervals were estimated using 5,000 stratified bootstrap samples.
Multiplicity: Speech-clinical correlations were considered exploratory. Both nominal Pvalues and Benjamini–Hochberg false-discovery-rate-adjusted $q$values were reported; conclusions were not based solely on nominal statistical significance.

Standardized posterior coefficient	Mean	95% CrI	Pr(β>0)
BNSS Blunted Affect	0.914	0.336 to 1.515	99.9%
BNSS Alogia	0.448	−0.142 to 1.057	93.2%
*Facial PC1: Spontaneous S_avg (0.368), spontaneous mouth opening (0.361), reading mouth opening (0.358), reading S_avg (0.358), and reading/spontaneous vLL_abs_avg (0.350/0.331)	−0.408	−0.669 to −0.168	<0.1%
**Facial PC4: Reading and spontaneous S_ratio_avg (0.496/0.483), spontaneous and reading vJC_abs_avg (0.367/0.328), and spontaneous vLL_abs_avg (0.250)	−0.454	−0.845 to −0.083	0.9%
Reading duration	0.330	−0.356 to 1.033	82.4%
***Reading CPP: The strength of the dominant cepstral peak associated with periodic vocal-fold vibration, relative to the estimated cepstral background or regression line, during the standardized reading task	0.159	−0.425 to 0.743	70.9%

CROSS-VALIDATED MODEL PERFORMANCE



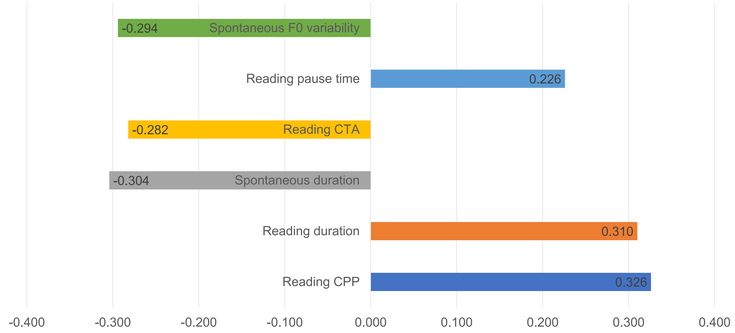
Model	AUC (95% CI)	Accuracy
Speech only	0.703 (0.571–0.827)	64.2%
Digital multimodal	0.780 (0.655–0.887)	73.1%
Clinician only	0.828 (0.716–0.926)	82.1%
Integrated	**0.879 (0.788–0.957)**	**83.6%**

Integrated operating characteristics	Sensitivity	Specificity	Posterior-risk groups
Accuracy 83.6%; balanced accuracy 83.1%	86.8%	79.3%	Low 25 | Intermediate 8 | High 34

Outcome threshold is sample-derived and exploratory; AUCs are based only on the 67 TRS records.
Facial PC1: A broad **facial activation and mouth-movement component. It explained 37.9% of the variance among the facial variables.*
***Facial PC4: A more specific facial-configuration or movement-ratio component. It explained 9.2% of the facial-variable variance.*
****Indicates the prominence and regularity of periodic structure in the voice. It may reflect voice quality, harmonic organization, vocal-fold vibration, and possible dysphonia or subharmonic activity.*

SPEECH–CLINICAL ASSOCIATIONS

Correlations with PANSS Expressive Negative severity (n=67). Positive values indicate higher feature values with greater severity; negative values indicate the reverse.



Key findings: Reading CPP (ρ=0.326, P=0.007), reading duration (ρ=0.310, P=0.011), spontaneous duration (ρ=−0.304, P=0.013), reading CTA (ρ=−0.282, P=0.021), and spontaneous F0 variability (ρ=−0.294, P=0.016).
Multiplicity: No displayed speech–clinical correlation remained significant after BH correction (minimum q=0.079).

CLINICAL CHARACTERISTICS

Clinical assessment	Mean ± SD
PANSS total	71.73 ± 10.91
Marder Negative factor	21.28 ± 3.98
PANSS Expressive Negative	11.57 ± 2.51
PANSS Positive subscale	15.28 ± 4.16
Marder Positive factor	21.84 ± 4.88
BNSS Alogia: quantity	2.67 ± 1.08
BNSS Alogia: elaboration	2.88 ± 1.09
BNSS Alogia	5.55 ± 2.09
BNSS Blunted Affect	8.46 ± 2.43
BNSS total	37.90 ± 9.03
Simpson–Angus Scale total	1.43 ± 2.44

CONCLUSIONS

- The Bayesian integrated model, combining BNSS expressive-domain ratings with speech and facial biomarkers, demonstrated better discrimination of higher ENS than either the digital-only model (AUC 0.879 vs. 0.780) or clinician-only model (AUC 0.828). The integrated model achieved 83.6% accuracy, 86.8% sensitivity, and 79.3% specificity, suggesting that the modalities provided complementary rather than redundant information.
- The Bayesian probit framework successfully generated individualized posterior probabilities of higher ENS and supported classification into low-, intermediate-, and high-probability groups. Posterior estimation demonstrated satisfactory convergence across four sampling chains ($\hat{R} \leq 1.003$), supporting the technical feasibility of integrating clinician-rated and digitally derived measures within a unified probabilistic model. This approach also offers an interpretable method for expressing uncertainty rather than relying solely on a single symptom score or deterministic classification.
- BNSS Blunted Affect demonstrated the clearest independent association with higher ENS after adjustment for the other predictors (posterior mean $\beta = 0.914$, 95% CrI 0.336–1.515). Two facial principal components also retained credible intervals excluding zero, indicating that facial-behavior data contributed information beyond the clinician ratings and speech measures. Because principal-component direction is mathematically determined, these facial findings require clinical interpretation and replication before individual facial behaviors can be identified as biomarkers.
- Speech markers demonstrated modest but clinically coherent associations with ENS. Greater severity was associated with higher reading CPP and longer reading duration, whereas shorter spontaneous-speech duration, lower canonical timing alignment, and reduced F0 variability were associated with greater ENS. However, none of the displayed speech–clinical correlations remained statistically significant following Benjamini–Hochberg correction; these findings should therefore be considered hypothesis-generating.
- The results support the feasibility of using Bayesian multimodal integration to supplement clinician-administered assessments. Combining BNSS ratings with objective speech and facial measures may improve identification of higher expressive symptom burden while providing participant-level estimates of uncertainty that could be useful for clinical-trial enrichment, monitoring, and future digital-endpoint development.
- These findings require prospective external validation using a larger and more clinically diverse sample, a prespecified and clinically meaningful ENS threshold, and independently confirmed digital-feature definitions.

NYU School of Medicine NYU LANGONE MEDICAL CENTER · modality.ai · The Nathan S. Kline Institute for Psychiatric Research · VALIS BIOSCIENCE