

The AVOCADOS Study: Evaluating Speech as a Biomarker of Human Spaceflight Adaptation and Operational Status in a Long COVID Analogue

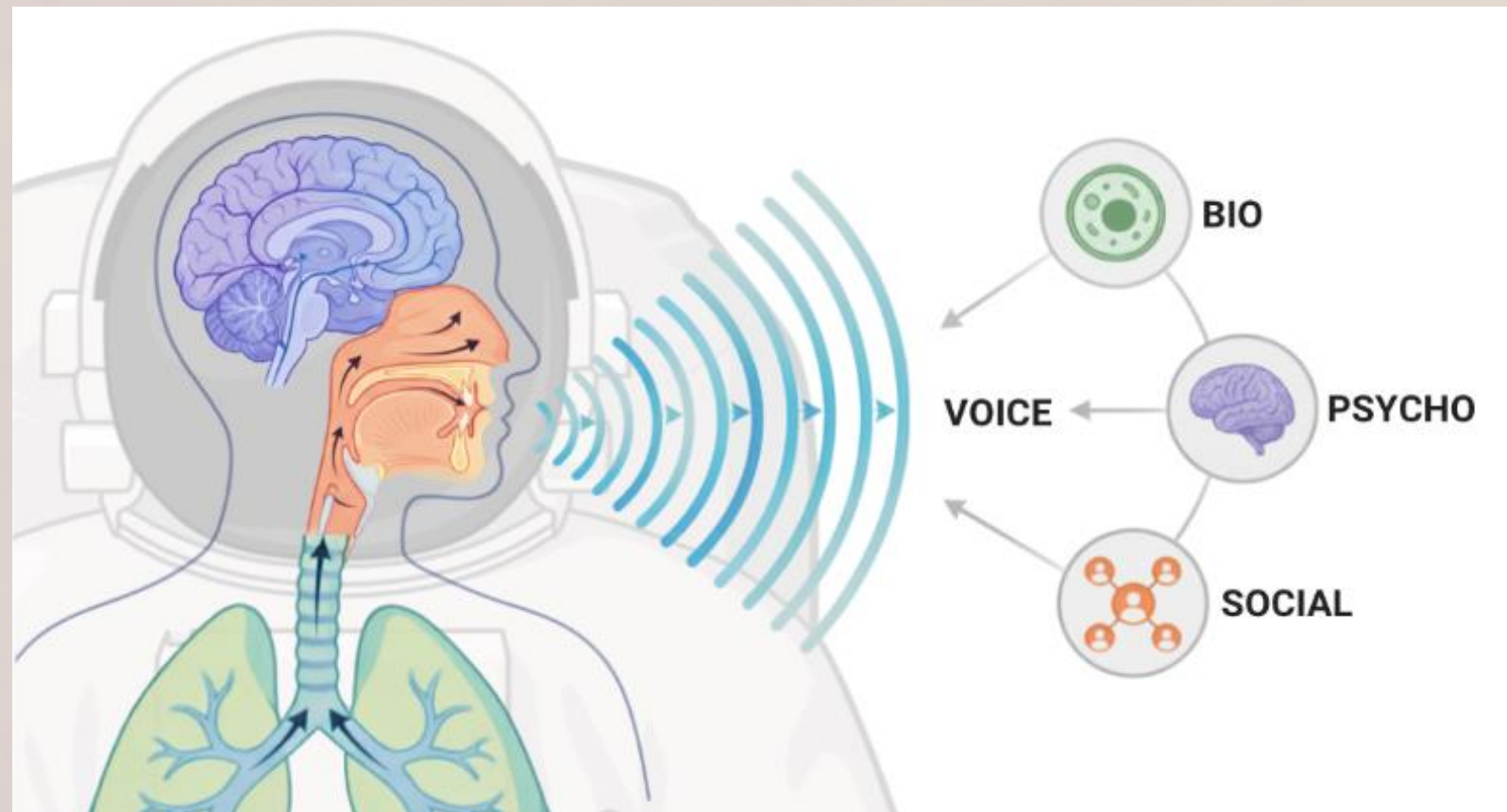
Zoe Swann Baillie, PhD¹

¹Leverhulme Centre for Humanity and Space, University of Leicester, United Kingdom



BACKGROUND

Future long-duration spaceflight missions will require low-burden methods for monitoring astronaut health, adaptation and operational readiness in environments with limited medical support. Speech is an attractive candidate bio-psycho-social biomarker because it can be collected remotely, repeatedly and non-invasively using existing communication systems^{1,4}.



The AVOCADOS (Astronaut Voice Observation to Capture Adaptation, Dysregulation and Operational Status) study investigates whether changes in spaceflight symptoms are reflected in measurable changes in speech production and voice quality². Long COVID was selected as a terrestrial analogue because it is characterised by fluctuating fatigue, cognitive symptoms ("brain fog"), isolation, stress, mitochondrial and endothelial mechanisms, and alterations in daily functioning that may share features with operational challenges encountered during prolonged space missions³.

Aim: To determine whether speech-derived measures are sensitive to day-to-day changes in self-reported health and cognitive status.

METHODS

STUDY DESIGN: Prospective single-participant pilot study conducted over 14 consecutive days. This study was approved by the Research Ethics Committee at the University of Leicester. The participant completed morning and evening symptom assessments and speech recordings using a remote digital survey tool (JotForm).

Self-reported symptom ratings included Fatigue, Brain fog, Low mood, Stress from 0-10, with 5 being their usual state over the last 7 days. Domains were selected per the C19-YRS clinical test.

DEMOGRAPHICS: This pilot participant was an adult female, native British English speaker, with a confirmed diagnosis of Long COVID (symptoms >12 months) and no pre-existing respiratory or voice disorders.

SPEECH TASKS were completed during each recording session:

Rainbow Passage: Clinically validated 190-word reading task providing controlled speech production.

Open Speech: "How are you feeling today?" Naturalistic speech sample reflecting spontaneous communication behaviour.

Verbal Fluency: "How many animals / kitchen utensils / fruits and vegetables can you name in a minute?" The category was randomised to prevent habituation. This is a cognitively demanding speech task requiring lexical retrieval and executive processing.

ACOUSTIC ANALYSIS was conducted using Praat software to extract a total of nineteen speech-derived outcome measures across all three tasks over the duration of the 14 days.

- Speech Production** (Words per minute (WPM), Speech duration, Word count)
- Voice Quality** (Harmonics-to-noise ratio (HNR), Cepstral peak prominence (CPP), Shimmer (local))
- Pitch** (Mean fundamental frequency (F0), F0 standard deviation, F0 coefficient of variation)
- Pauses** (Total pause duration, Mean pause duration, Maximum pause duration, Pause count)
- Additional Features** (Intensity-related measures, Temporal speech characteristics)

Note: this pilot analysis is part of a larger ongoing study of n=55 people, which includes further clinical assessments not yet analysed.

STATISTICAL ANALYSIS

For each speech task and symptom, task-specific regression models were fitted using the general form **Outcome ~ Symptom + Time** where: Outcome = speech-derived feature; Symptom = fatigue, brain fog, low mood or stress; and Time = AM vs PM recording.

Models were fitted using glmmTMB in R. Outcome distributions were selected according to feature characteristics: **Gaussian** (identity link), **Gamma** (log link), **Negative binomial** (log link), or **Tweedie** (log link) depending on the distribution of each acoustic measure (per the DHARMA package in R). A total of over 200 symptom-outcome combinations were screened across tasks. For log-link models, effect sizes were transformed to % change per one-point increase in symptom severity: **% change = [exp(β) - 1] $\times$ 100**

Multiple-comparison correction was performed using the Benjamini-Hochberg false discovery rate (FDR) procedure. Model outputs included: regression coefficient estimates, 95% confidence intervals, raw p-values, and FDR-adjusted p-values.

Candidate findings were prioritised based on: FDR-corrected significance, effect magnitude, biological plausibility, and consistency across speech tasks. As an exploratory single-participant pilot study, analyses were considered hypothesis-generating rather than confirmatory.

RESULTS

Daily symptom ratings fluctuated across the 14-day study period, with fatigue and brain fog showing the greatest variation relative to the participant's self-reported usual state (**FIGURE 1**).

Exploratory modelling identified several associations between symptom severity and speech measures. In Reading, higher fatigue was associated with slower reading speed (-4.4% WPM; FDR = .009) and longer reading duration (+4.7%; FDR = .009). Higher low mood was associated with slower reading (-7.4% words per minute; FDR = .028) and increased reading duration (+8.1%; FDR = .023). In Open Speech, higher stress was associated with increased word production (+13.0%; FDR = .093) and longer speaking duration (+10.6%; FDR = .049). (**FIGURE 2**)

Brain fog was associated with several measures of speech production and voice quality. In Reading, higher brain fog was associated with lower cepstral peak prominence (CPP; β = -0.234, FDR = .093), lower harmonics-to-noise ratio (HNR; β = -0.598, p = .005), slower reading speed (-4.3%, p = .005), and longer reading duration (+4.6%, p = .004). Similar negative associations between brain fog and HNR were observed in Open Speech (β = -0.695, p = .009) and Verbal Fluency (β = -0.818, p = .017). (**FIGURE 3**)

To assess whether HNR findings might reflect recording artefacts, exploratory diagnostics were performed. HNR showed essentially no relationship with mean recording intensity across any task (Rainbow: p = 0.05, p = .82; Open Speech: p = 0.13, p = .54; Verbal Fluency: p = -0.08, p = .71), suggesting that microphone gain issues were unlikely to explain the observed associations. In contrast, HNR was correlated with CPP during Reading (p = 0.75, p < .001) and with CPP during Open Speech (p = 0.51, p = .009), with a weaker, non-significant correlation in Verbal Fluency (p = 0.29, p = .15), indicating that HNR and CPP tracked a common voice-quality dimension in at least two of the three tasks. This suggests genuine variation in voice quality, though noise artifacts cannot be fully dismissed at this stage.

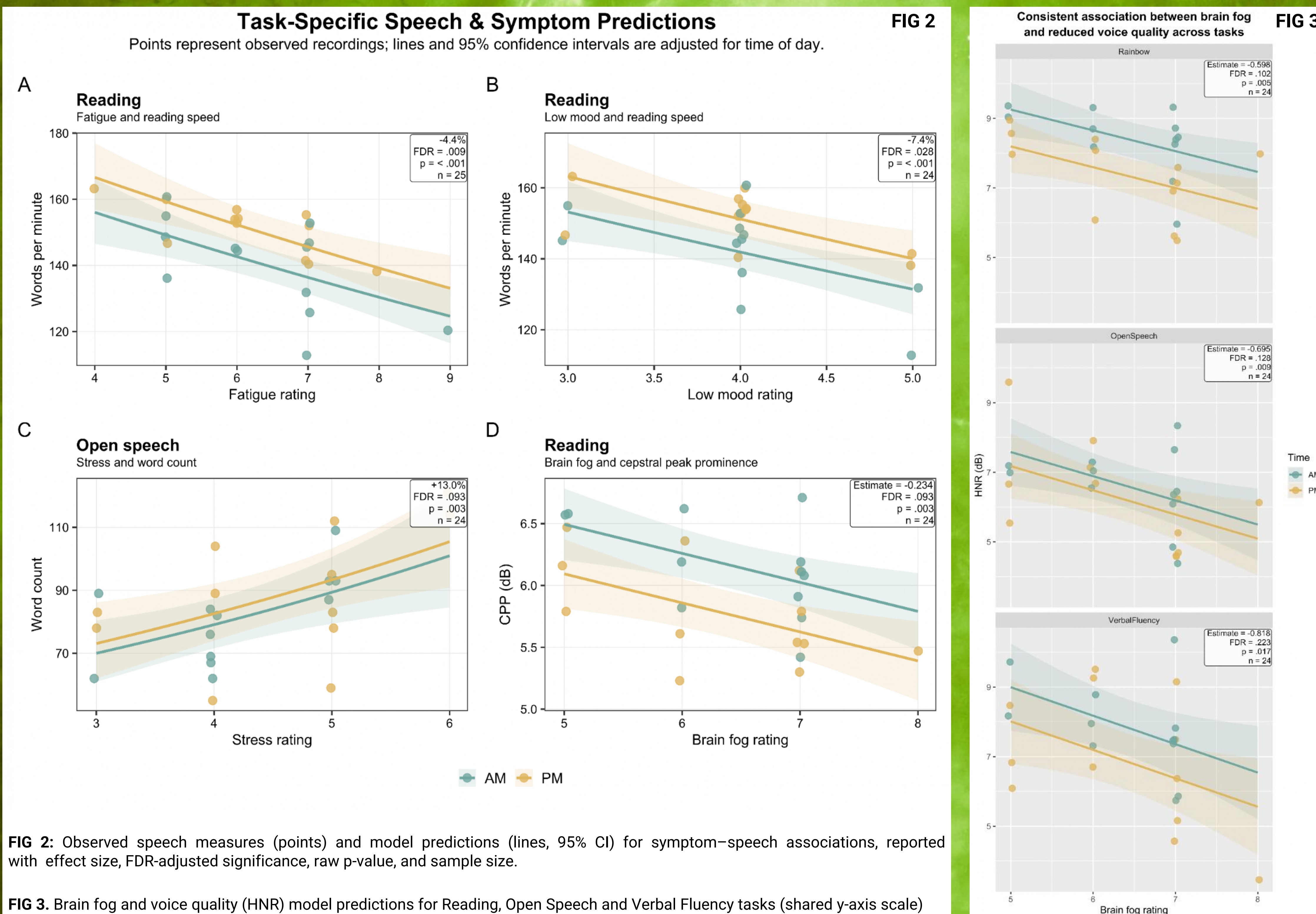


FIG 2: Observed speech measures (points) and model predictions (lines, 95% CI) for symptom-speech associations, reported with effect size, FDR-adjusted significance, raw p-value, and sample size.

FIG 3. Brain fog and voice quality (HNR) model predictions for Reading, Open Speech and Verbal Fluency tasks (shared y-axis scale)

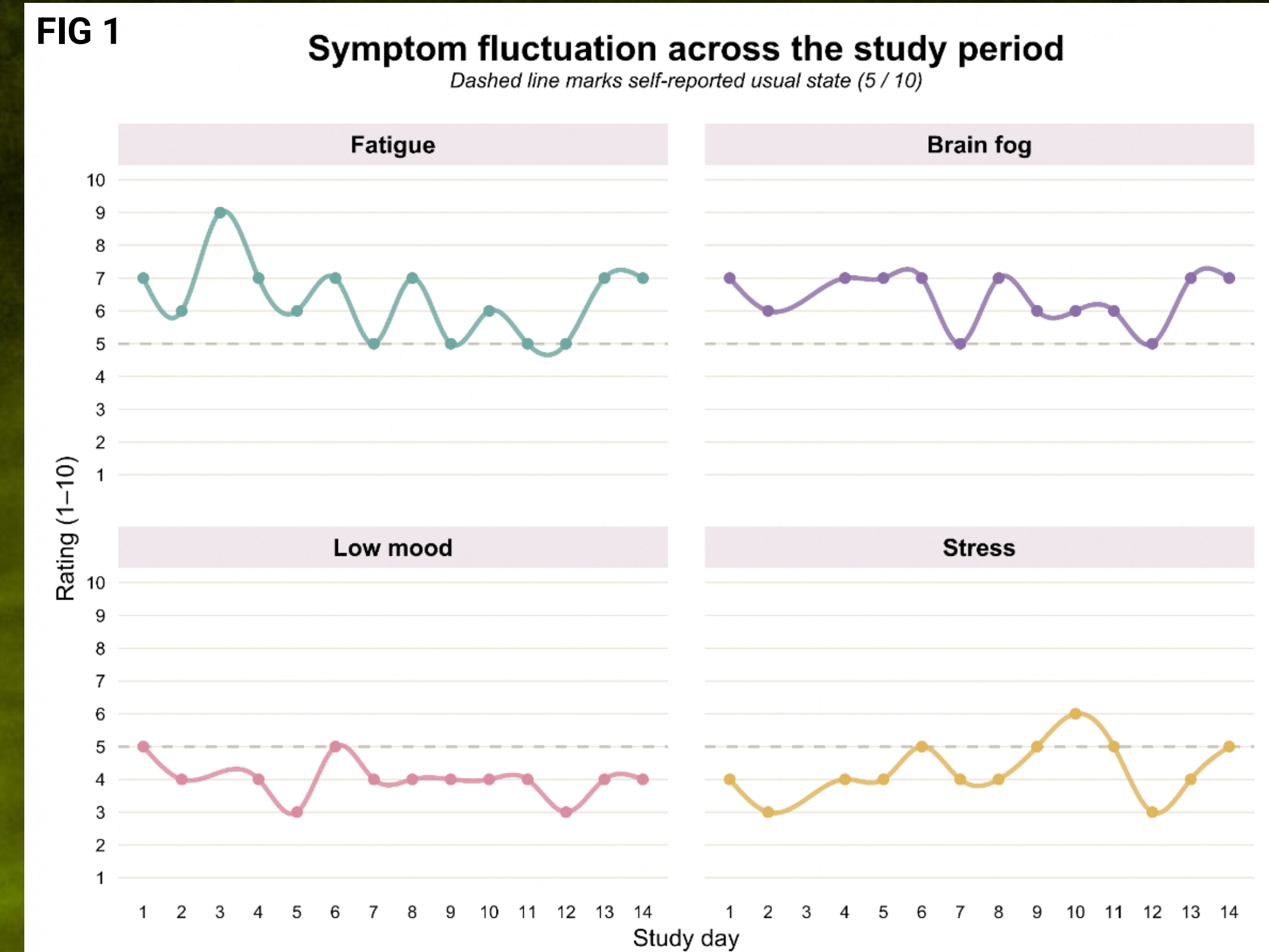


FIG 1: Daily self-reported symptom ratings over 14 days. Dashed line = usual state (5/10).

DISCUSSION

Spaceflight missions need ways to monitor astronaut adaptation and performance despite isolation, communication delay, and limited medical resources. Speech-based monitoring offers key advantages: low crew burden (collected during routine activities), use of existing communication infrastructure, longitudinal tracking, and sensitivity to multiple biopsychosocial systems at once (e.g., respiratory, cognitive, linguistic, motor, social conflict). This makes speech a candidate early indicator of symptoms before more obvious issues appear.

Speech biomarkers identified here (speech rate, duration, word production, and voice quality) tracked changing symptom burden. Reduced voice quality was consistently associated with brain fog across three speech tasks. These measures could support monitoring of cognitive status, fatigue, and operational readiness on long-duration missions.

Limitations: Single-participant exploratory pilot with multiple comparisons; temporal autocorrelation was not modelled; recording-environment effects cannot be fully excluded. Speech monitoring is inherently dependent on communication ability and represents one component of a broader astronaut health-monitoring framework rather than a universal solution. Replication in larger analogue cohorts is needed.

Future Directions: Future work will evaluate replication across participants, speech-based forecasting of symptom flares, linguistic measures, and integration with operational monitoring frameworks relevant to human spaceflight. Because speech biomarkers are inherently individualised, AVOCADOS adopts a precision-monitoring approach that prioritises within-person longitudinal change rather than population-level classification. Future analyses will therefore focus on detecting meaningful deviations from an individual's baseline adaptation and operational status.

Ethical Considerations: Astronauts often mask symptoms to protect mission status. Any use of speech monitoring must be built on trust and support, not surveillance, i.e., voluntary, transparent, and never used punitively, with astronauts given access to and control over their own data to make decisions about pacing, countermeasures, etc.

REFERENCES

- Shamei, A. et al., Sci Rep 13, 8231 (2023).
- Kappen M, et al., Sci Rep. (2022)
- Choukér, A., Stahn, A.C., npj Microgravity 6, 32 (2020).
- Baykaner, K, et al., Front. Bioeng. Biotechnol. 3, (2015). (More available on request.)

ACKNOWLEDGEMENTS

The author would like to thank the participant for their time and engagement throughout this study; colleagues at the University of Leicester and Leverhulme Centre for Humanity and Space (LCHS) for their peer review and support; and the Leverhulme Trust for funding this work. Thank you to Sam Baillie for his love of glmmTMB and DHARMA statistical packages. Special thanks to Laura Nevay (LCHS) for her steadfast commitment to fostering an inclusive and accessible research environment.