Background:
Speech and language biomarkers (SLBs) are increasingly studied for early detection of cognitive decline[1]. However, many commonly used speech tasks produce limited and predictable responses[2], and may be constrained by task-specific parameters. This can lead to samples that have formulaic lexical content and limited structural variety, reducing the sensitivity of higher-order linguistic measures.

Methods:
We introduce an autobiographical recall task embedded within the **California Cognitive Assessment Battery (CCAB)**. Similar to the CDR Event Recall Task, participants are prompted to verbally describe a significant event from the prior two weeks, with instructions automatically presented and audio collected remotely. Speech samples are then transcribed using a **Consensus Automatic Speech Recognition (CASR)** pipeline that integrates outputs from multiple ASR systems to reduce transcription errors relative to any single engine. Automated pipelines extract acoustic, timing, lexical, semantic, and syntactic features. Planned analyses examine cross-sectional associations between SLBs and CCAB test performance, as well as within-subject change as longitudinal data accrue.

Results:
Initial results indicate that the event recall task reliably elicits longer (m =268 words) and more linguistically diverse spontaneous speech than picture description or logical memory recall paradigms (m=180 words and m=83 words, respectively). Increased response length and structural variability are expected to improve the reliability and interpretability of lexical diversity, semantic coherence, and syntactic complexity measures, which are known to be sensitive to sample length. Normative data collection for this task is ongoing (n=343), and adds to the CCAB normative dataset, which exceeds 2,000 participants across complementary verbal and non-verbal tasks.

Conclusions:
The CCAB event-recall task elicits ecologically valid speech data optimized for automated language analysis. Integration within CCAB allows speech derived measures to be interpreted alongside established cognitive outcomes, while the CASR pipeline provides more accurate transcripts than any single ASR engine. Together, this infrastructure supports scalable cognitive screening and longitudinal monitoring, complementing established speech-based tasks while addressing key methodological shortcomings in speech-based digital biomarkers.

---

[1] Kovac, D., Novakova, L., Mekyska, J., Novotny, K., Brabenec, L., Klobusiakova, P., & Rektorova, I. (2025). Digital speech biomarkers for assessing cognitive decline across neurodegenerative conditions. *Computers in Biology and Medicine*, *198*, 111251.

[2] Boschi, V., Catricala, E., Consonni, M., Chesi, C., Moro, A., & Cappa, S. F. (2017). Connected speech in neurodegenerative language disorders: a review. *Frontiers in psychology*, *8*, 269.