

AI-Enabled Conversational Agent for History Taking in Alzheimer's Disease and Related Dementias: A Pilot Study

Andrew G. Breithaupt, MD¹; Arin Nelson¹; Oz Alon¹; Jeanne Powell, PhD¹; James Finch, PhD²;
Nayoung Choi²; Howard Rosen, MD³; James Lah, MD, PhD, Allan Levey, MD, PhD; Jinho D. Choi, PhD²

¹Department of Neurology, Emory University School of Medicine, Atlanta, GA ²Emory Natural Language Processing Research Lab, Atlanta, GA ³Department of Neurology, University of California, San Francisco, San Francisco, CA

This research was supported by a Next Generation Research Grant from the American Brain Foundation as well as Ekai and the Alzheimer's Association.

BACKGROUND

The Challenge

- Early ADRD diagnosis requires comprehensive patient histories, yet time constraints limit thorough symptom exploration
- Memory clinic wait times exceed one year in many regions

Our Approach

- We developed a voice-interactive conversational agent powered by a large language model (LLM) to conduct structured patient interviews
- The agent systematically covers ~70 symptoms across cognitive, behavioral, motor, and functional domains recommended by dementia specialists

METHODS

Study Design

- Within-subjects pilot study: 51 adults with suspected ADRD from a cognitive neurology clinic
- Each participant completed both an agent-led interview and a separate clinician interview (blinded to agent results)
- 37 of 51 participants had both interviews recorded and transcribed

Symptom Labeling

- 35 high-yield symptoms identified by dementia specialists and labeled through consensus: PRESENT, ABSENT, AMBIGUOUS, PAST HISTORY, or NOT DISCUSSED
- Gold standard established by reviewing both transcripts when interviews disagreed
- Inter-rater reliability: Cohen's κ = 0.71–0.74 (substantial agreement)

Outcome Measures

- Sensitivity & specificity for symptom detection: all 35 symptoms
- Systematic coverage & ambiguity : 21 core symptoms ideally asked at every visit
- User experience surveys using the Acceptability E-Scale

I. Clinical Data Acquisition



Participants (51 interviewed by agent, 37 with recorded clinician interview)

Characteristic	Value
Age, years (mean ± SD)	72.6 ± 9.1
Female / Male	62.7% / 37.3%
Race (White / Black / Asian Indian/ Not Reported)	74.5% / 21.6% / 2.0% / 2.0%
Education, years (mean ± SD)	15.6 ± 2.1

RESULTS

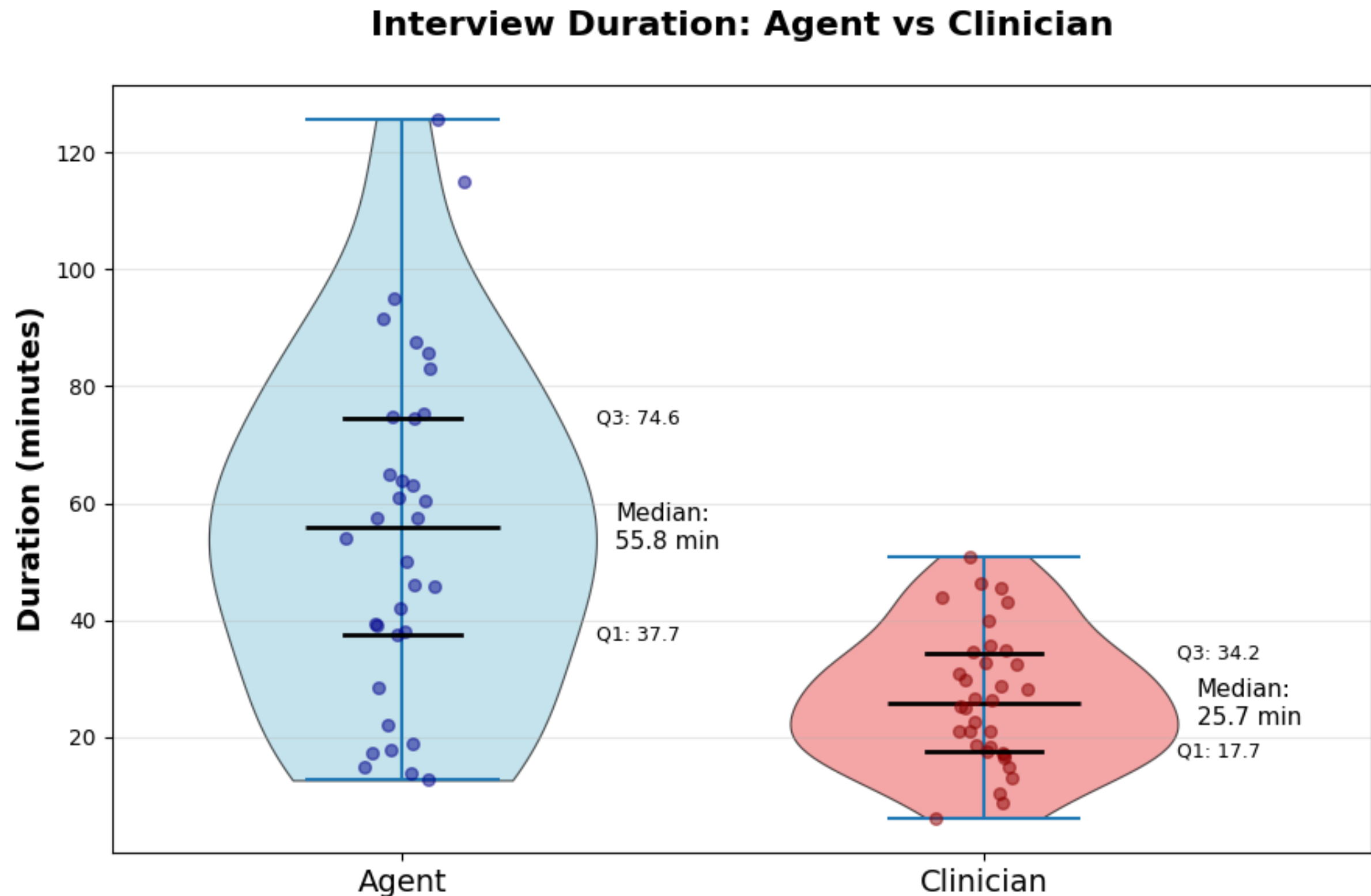
37 participants had transcribed interviews with both the conversational agent and the clinician for comparison

Participants are spending about twice as much time for history taking overall with the agent compared to the clinician.

Measure	Agent Interview	Clinician Interview
Median interview duration	56 min (IQR 38–75)	26 min (IQR 18–34)

More time spent despite 10 of 37 agent interviews ending prematurely

- 6 terminated by agent (misinterpreted user to say bye then stopped the interview)
- 2 terminated by patients (time burden; repeated freezing)
- 2 terminated by staff (time constraints, patient needed to go to their clinic visit)



The agent has promising accuracy in identifying symptoms (n = 37)

84%

Sensitivity

95% CI: 78–88%

98%

Specificity

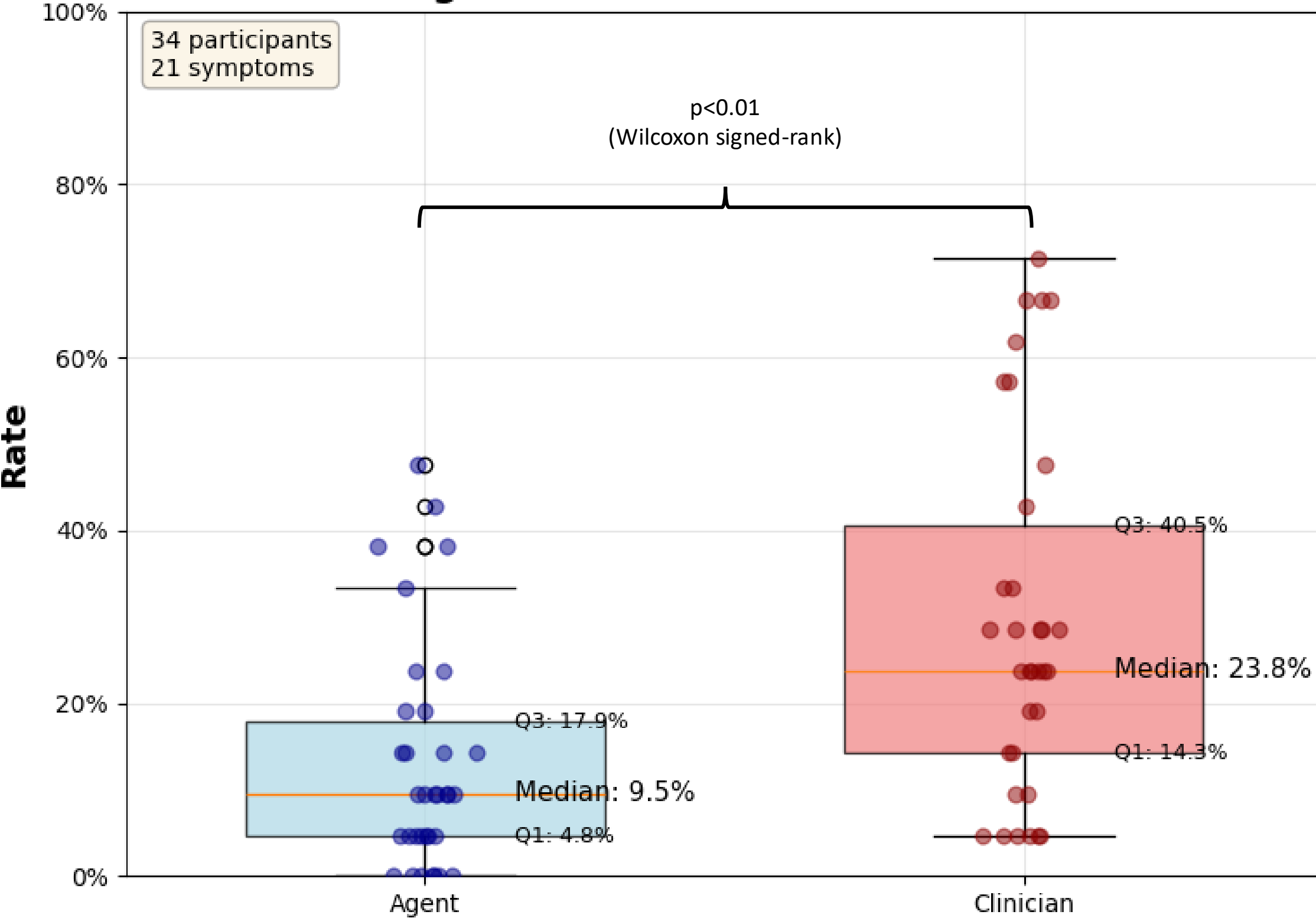
95% CI: 93-99%

- Informants reported less to the agent, potentially reducing sensitivity
 - Patient-to-informant turn ratio shifted from ~1.5:1 (clinician) to ~3:1 (agent)
- Some informants reported withholding information to avoid upsetting their loved one during the agent research interview.

34 of 37 participants above had new patient diagnostic interviews where the clinician collected comprehensive histories.

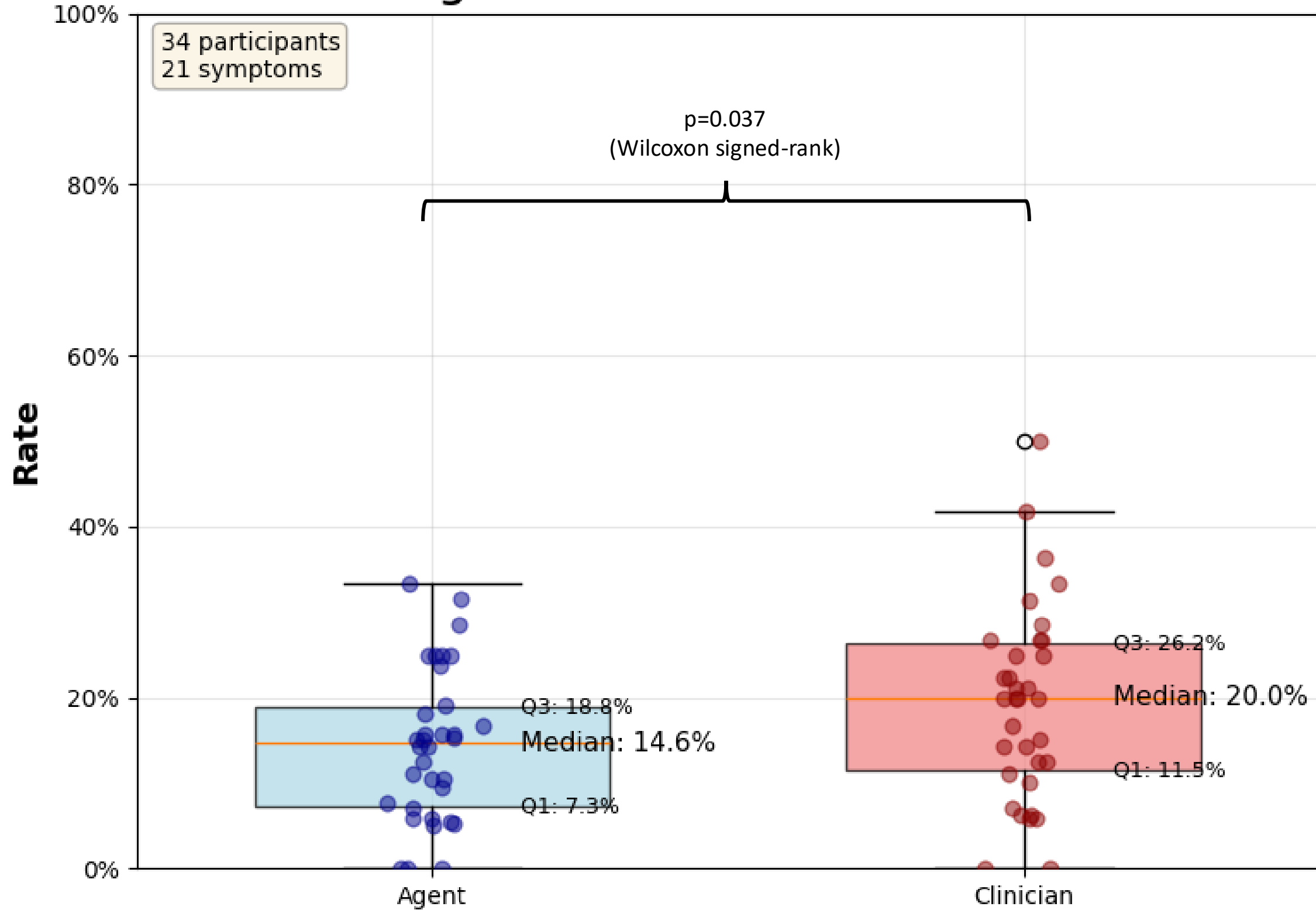
The agent is covering more symptoms than the clinician

Distribution of Per-Participant Not Discussed Rates: Agent vs. Clinician Interviews



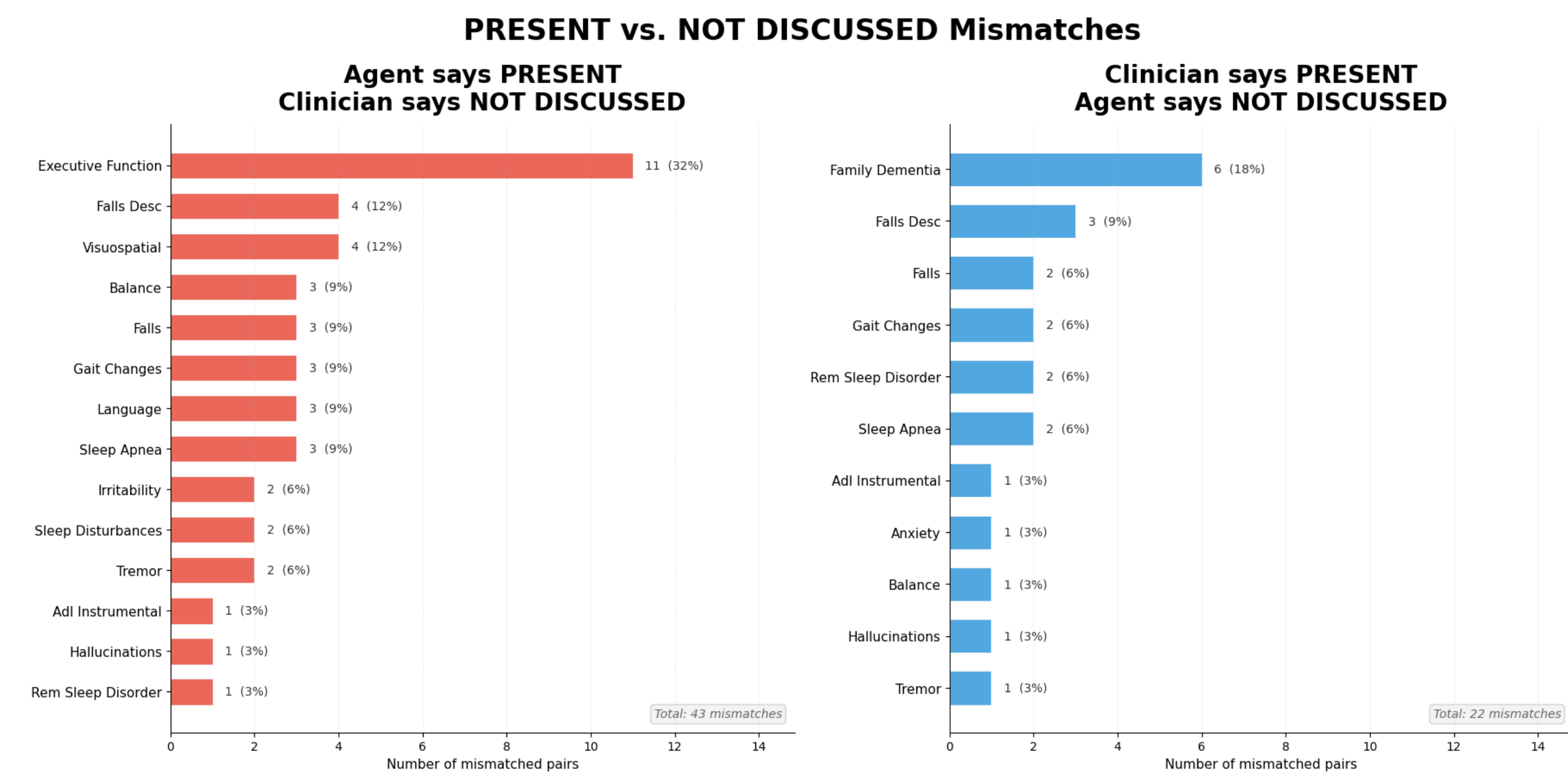
The agent does as well as clinicians in clarifying answers

Distribution of Per-Participant Ambiguity Rates: Agent vs. Clinician Interviews



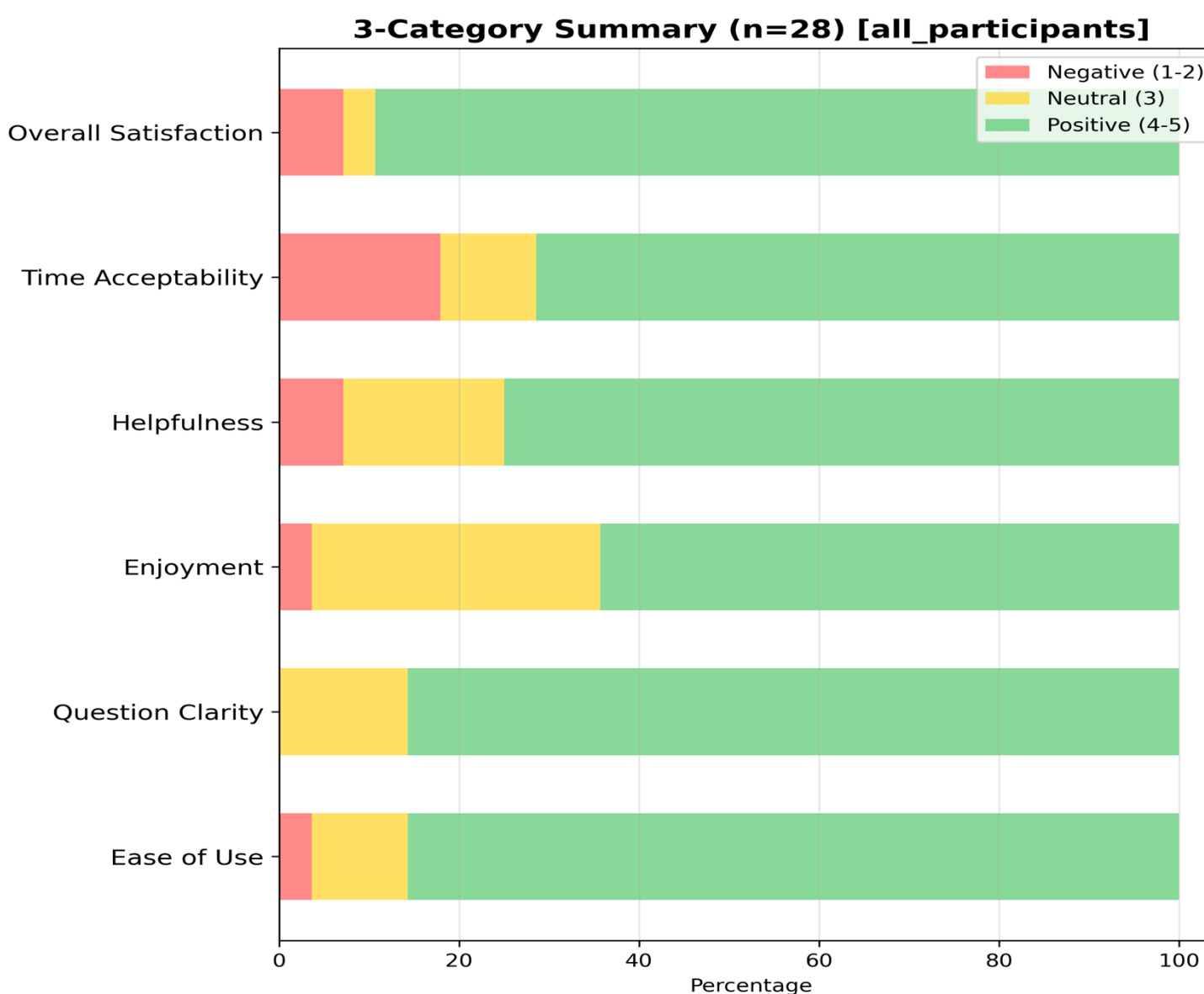
RESULTS (continued)

The agent more frequently identified present symptoms not discussed in clinician interview than vice versa



Overall satisfaction reached 89% (28/51 completed survey)

- “actually quite fun”, “enjoyable” (P9, P12)
- “I found this to be very thorough and felt it understood my information more than most doctors. It was affirming and accurate with the majority of my patient history. It was patient and took time to listen.”
- “some of the questions should be broken down as opposed to 2 or 3 questions” (P23);
- “It locked up twice and was pretty slow a couple of other times. I tried my patience when it was working well..” (P27)



CONCLUSIONS

- LLM-powered conversational agents can systematically elicit complex ADRD symptoms with greater systematic symptom coverage and less ambiguity than clinicians by spending twice as much history taking time
- These preliminary results suggest agents have potential to serve as structured pre-visit tools to aid clinicians with diagnostic efficiency and increase access to earlier ADRD evaluation

Limitations & Future Directions

- Investigating implementation hurdles
- Investigate separate informant interviews to address disclosure asymmetry
- Assess whether additional agent-elicited symptoms change clinician diagnosis or management
- Evaluate diagnostic accuracy and management decisions using AI-generated HPIs