An Evidence-Oriented Large Language Model Pipeline for Comparing Subjective–Objective Cognitive Correspondence in HIV and Alzheimer's Disease

Yiran Yuan[1], Joyce M. Lee[1], Carolyn Kuz[2], Xinglong Shi[3], Chelsea D. Voytek[2], Robert Scavilla[4], Cagla Akay-Espinoza[3], Corey McMillan[5], Lily A. Brown[2], Dennis Kolson[5], Shana D. Stites[6]

[1]College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA, [2]Department of Psychiatry, University of Pennsylvania, [3]Department of Oral Medicine, University of Pennsylvania, [4]Fourfront, LLC, [5]Department of Neurology, University of Pennsylvania, [6]Department of Psychiatry and Division of Geriatric Medicine, University of Pennsylvania

Introduction: Cognitive concerns are common among people with Human Immunodeficiency Virus (PWH), but how well self-reported cognition corresponds to objective performance—and how discrepancies might differ from those in people with Alzheimer's disease (PWAD), where impaired insight has been characterized—has not been systematically synthesized. Rather than using question-answering approaches, we compared evidence patterns across two heterogenous literature. We hypothesized that, in older PWH, subjective cognitive concerns align more consistently with objective cognition than in PWAD, where insight can be limited.
Methods: PubMed was searched through July 2026. PWH queries paired HIV/PLWH/AIDS with subjective cognition or IADL self-report and objective tests; PWAD queries paired AD, MCI, or dementia with insight, informant discrepancy, anosognosia, or SCD. Reviews, meta-analyses, and comments were excluded. After deduplication and one forward-citation pass, 139 PWH and 1,504 PWAD records were screened. Eligible studies were coded at the study level (Tier-1: paired measures; Tier-2: reported association or explicit dissociation). For each study, we classified how the relationship between patients' subjective cognitive reports and their objective test performance is described: as "alignment" (subjective reports move in step with objective performance) or as "reduced awareness/under-reporting" (patients report fewer or milder problems than their objective impairment indicates). Screening was benchmarked on a 27-study HIV gold set. Quantitative subsets were directionally harmonized and analyzed using Fisher-z transformed correlations with inverse-variance weighting.
Results: Retrieval recovered 93% (25/27) gold-set studies; screening achieved 100% precision (21/21) and 78% recall (21/27). In Tier-1 studies, alignment appeared in 7/14 PWH versus 4/13 AD studies; reduced awareness/under-reporting in 0/14 versus 2/13. Across five independent runs, Layer-1 screening showed 100% agreement (Fleiss' $\kappa = 1.00$). For Tier- pairs with extractable r/ρ, the harmonized pooled association was positive in PWH and negative in PWAD, indicating partial alignment in PWH versus opposite-direction pairing in PWAD; the small number of independent effects limited power to establish a significant between-group difference.
Conclusions: An evidence-oriented LLM pipeline can reproduce expert-curated HIV literature and structure cross-disease comparison, offering a scalable and auditable framework for evidence synthesis and informing interpretation of cognitive concerns in HIV research.