



Data-driven, clinico-pathological classification of neurodegenerative diseases

Jurij Rosen¹, Joseph Kambeitz¹, and Frank Jessen^{1,2,3}

¹Department of Psychiatry, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

²Excellence Cluster on Cellular Stress Responses in Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany

³German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany

BACKGROUND

The heterogeneity of neurodegenerative diseases challenges accurate clinical diagnosis and clinico-pathological correlation. Traditional univariate methods often fail to capture the complex relationships between multi-domain clinical symptoms and distinct neuropathological burden.

METHOD

This study utilises data from the National Alzheimer Coordinating Center, a tertiary referral real-world database comprising clinical and post-mortem neuropathological data from both cognitively unimpaired and symptomatic participants. General characteristics and clinical (behavioural, cognitive, and functional) features documented at the final visit prior to death were jointly analysed with pathological findings. A multi-block sparse Partial Least Squares (sPLS) analysis was employed to identify latent clinico-pathological profiles sharing variance between clinical and pathological blocks. The significance, stability, predictive performance, and clinical validity were assessed.

RESULTS

Data preprocessing yielded a final sample of 4,951 participants with 40 clinical and 34 pathological features. The analysis identified distinct clinico-pathological profiles, with the first three yielding the highest correlations (range, $r = 0.41$ – 0.52): Alzheimer's disease (profile 1), Lewy body disease (profile 2), and frontotemporal dementia (profile 3). Clinically, cognitive impairments and the expression of neuropsychiatric symptoms most markedly separated the profiles. Additionally, demographic factors (age, sex, living arrangements) modulated the clinical presentation in profiles 2 and 3, but not profile 1. Pathologically, pontine locus coeruleus degeneration emerged as a disease-specific rather than generalised feature, contributing prominently to profiles 1 and 2 only, despite their divergent clinical presentations. All profiles were independent of vascular factors. With significance, stability, and consistent feature selection frequencies being confirmed, the profiles' prevailing agreement with the clinician-assigned diagnosis validated the identified disease axes.

CONCLUSION

This unbiased sPLS approach autonomously reconstituted three distinct neurodegenerative entities from raw variance, independent of prior clinical labels. By integrating divergent neuropsychiatric symptoms and demographic modulators with their pathological drivers, this data-driven model captures disease complexity more accurately than traditional clinical classifications, potentially enabling earlier diagnosis and treatment.

CONTACT

Jurij Rosen, MD
☎ +49 221 478 1918756
✉ jurij.rosen@uk-koeln.de

GRANT

The NACC database is funded by NIA/NIH Grant U24 AG072122, and multiple grants supported the NIA Alzheimer's Disease Research Centers.

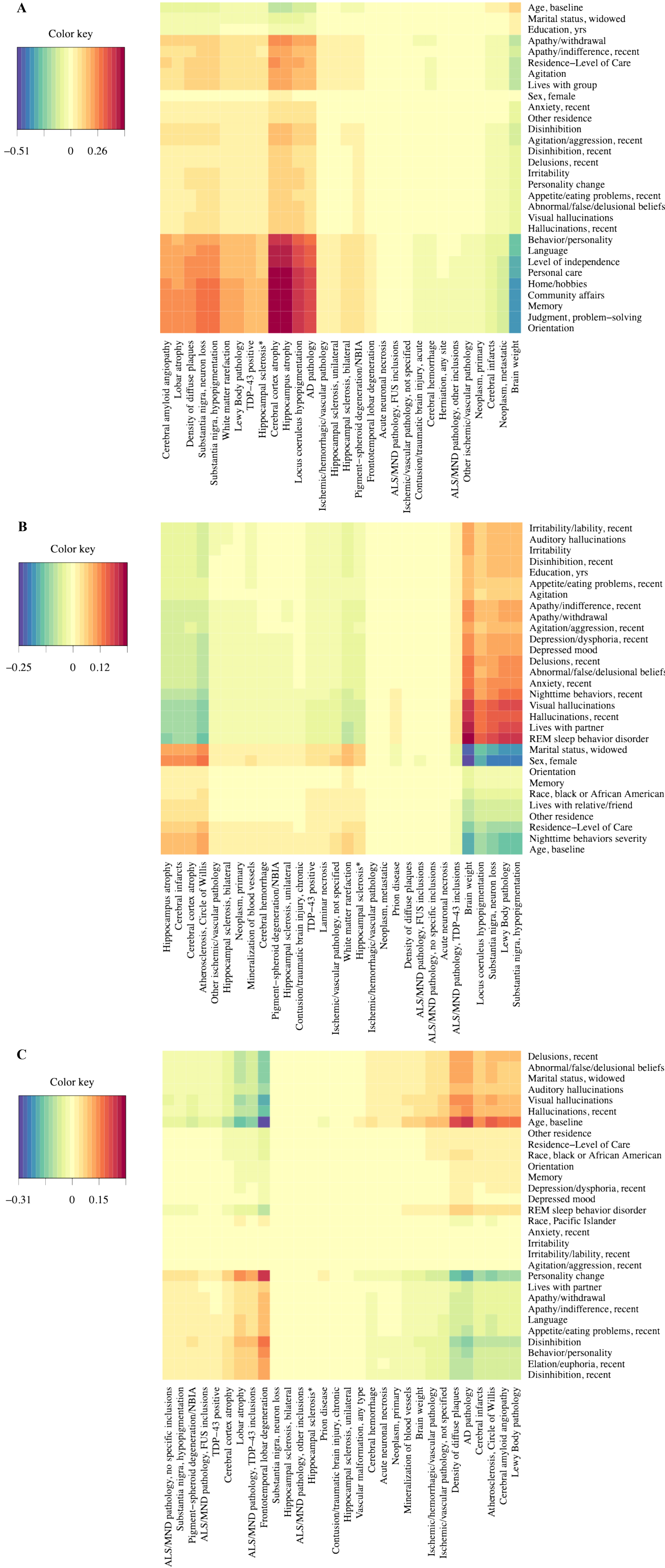


Figure 1: Clustered Image Maps of Clinico-Pathological Correlates. Heatmaps visualising the results of the sPLS analysis, showing the associations between clinical features (vertically) and neuropathological features (horizontally) across the three profiles (A-C) Colors indicate positive and negative correlations reflecting strength and direction of the associations.

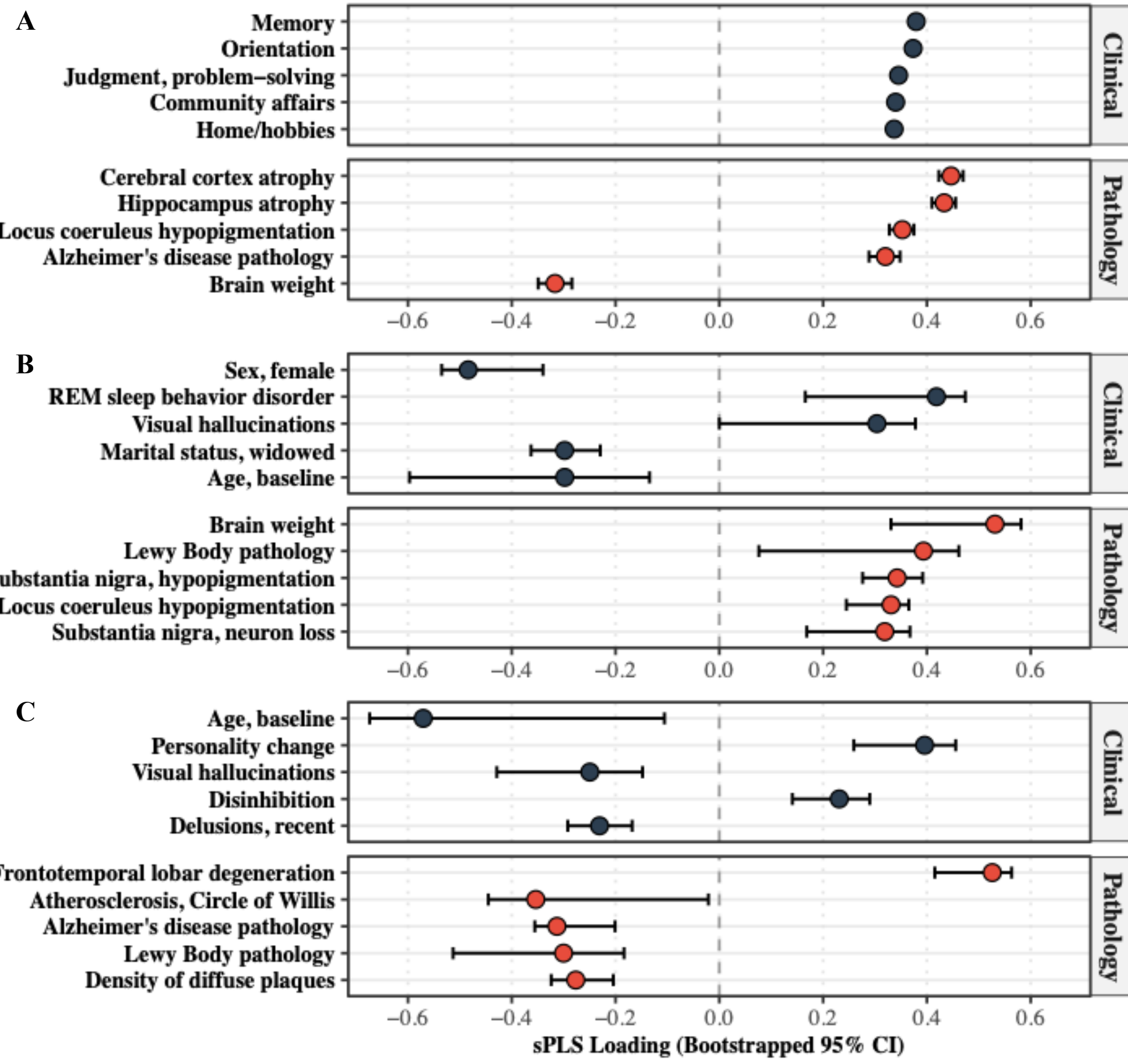


Figure 2: Loadings and stability of predominant features across profiles 1–3 (A-C). Loading weights are shown as point estimates from the final sPLS model, with error bars representing the 95% bootstrapped confidence intervals.

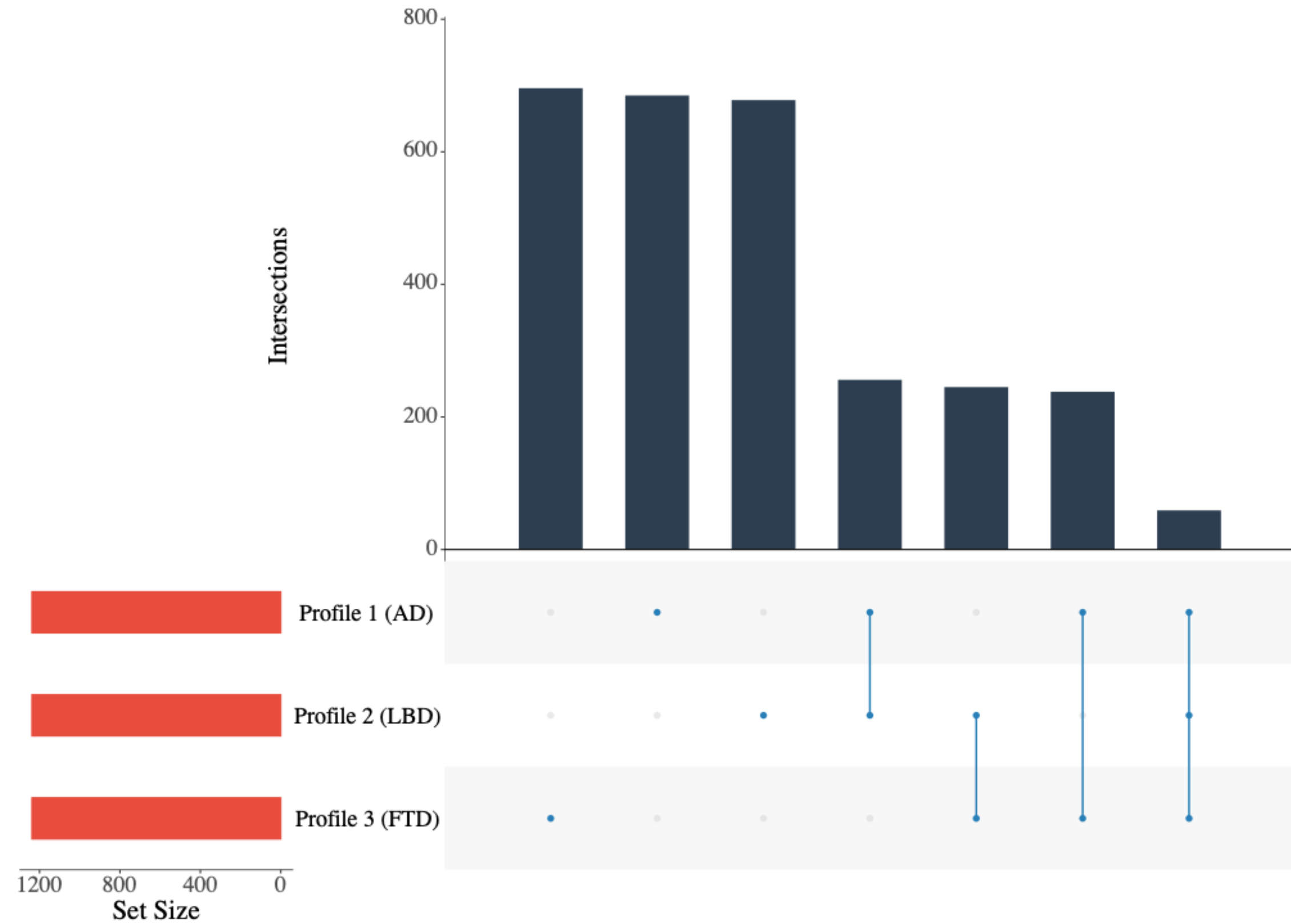


Figure 3: Co-occurrence of clinico-pathological profiles. Participants were classified into profiles 1-3 based on the top quartile (≥ 75 th percentile) of their respective profile scores.

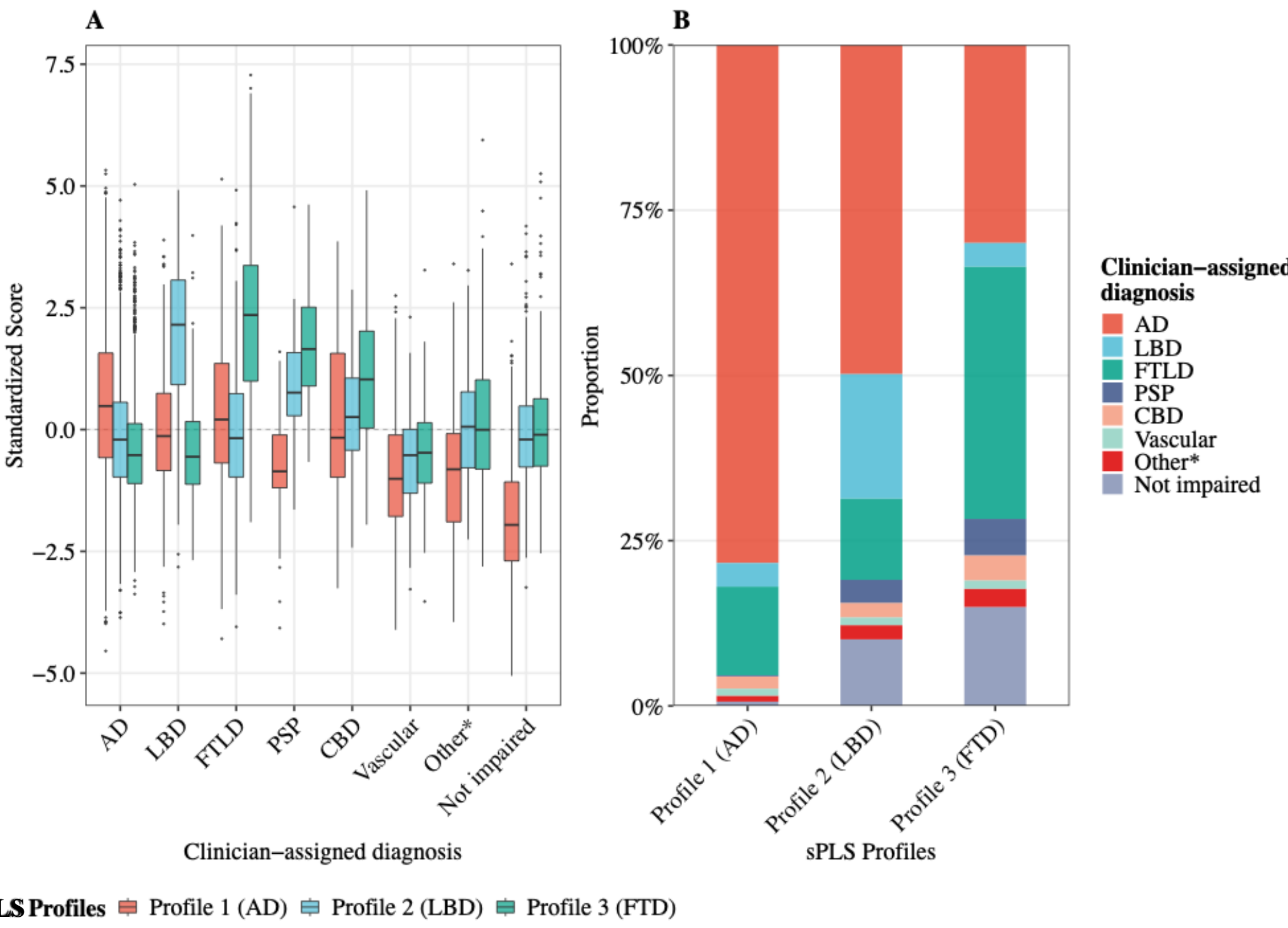


Figure 4: Validation and composition of sPLS profiles across clinician-assigned diagnoses. **A** Boxplots illustrating how diagnoses predict scores across the three profiles. **B** Bar plot displaying the relative proportion of diagnoses within the top quartile of participants expressing each sPLS profile.