

Detecting Core Features of Dementia with Lewy Bodies Using Natural Language Processing

Mohamed Heybe¹, Lucy Gibson^{1,2}, Annabel C. Price^{3,4}, Rudolf N. Cardinal^{3,4}, John T. O'Brien^{3,4}, Robert Stewart^{1,2}, Christoph Mueller^{1,2} - Contact: christoph.mueller@kcl.ac.uk

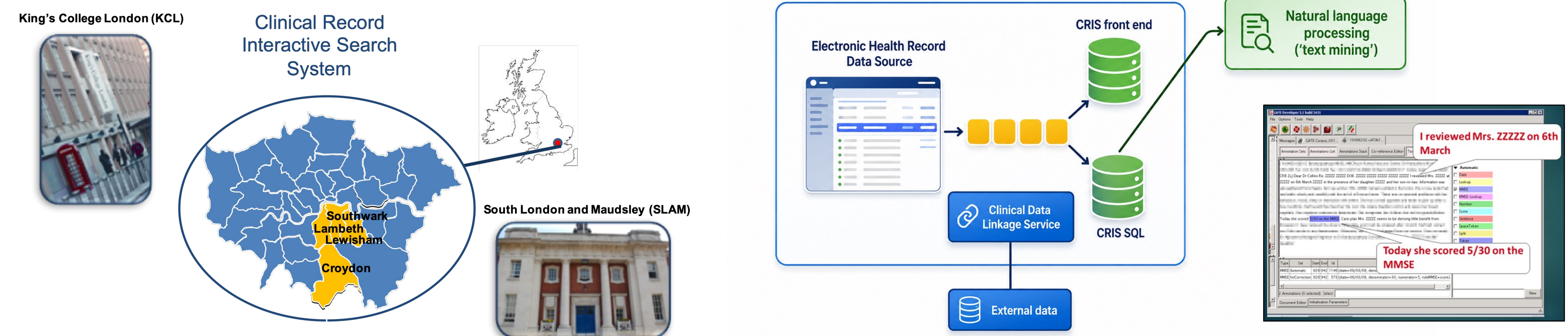
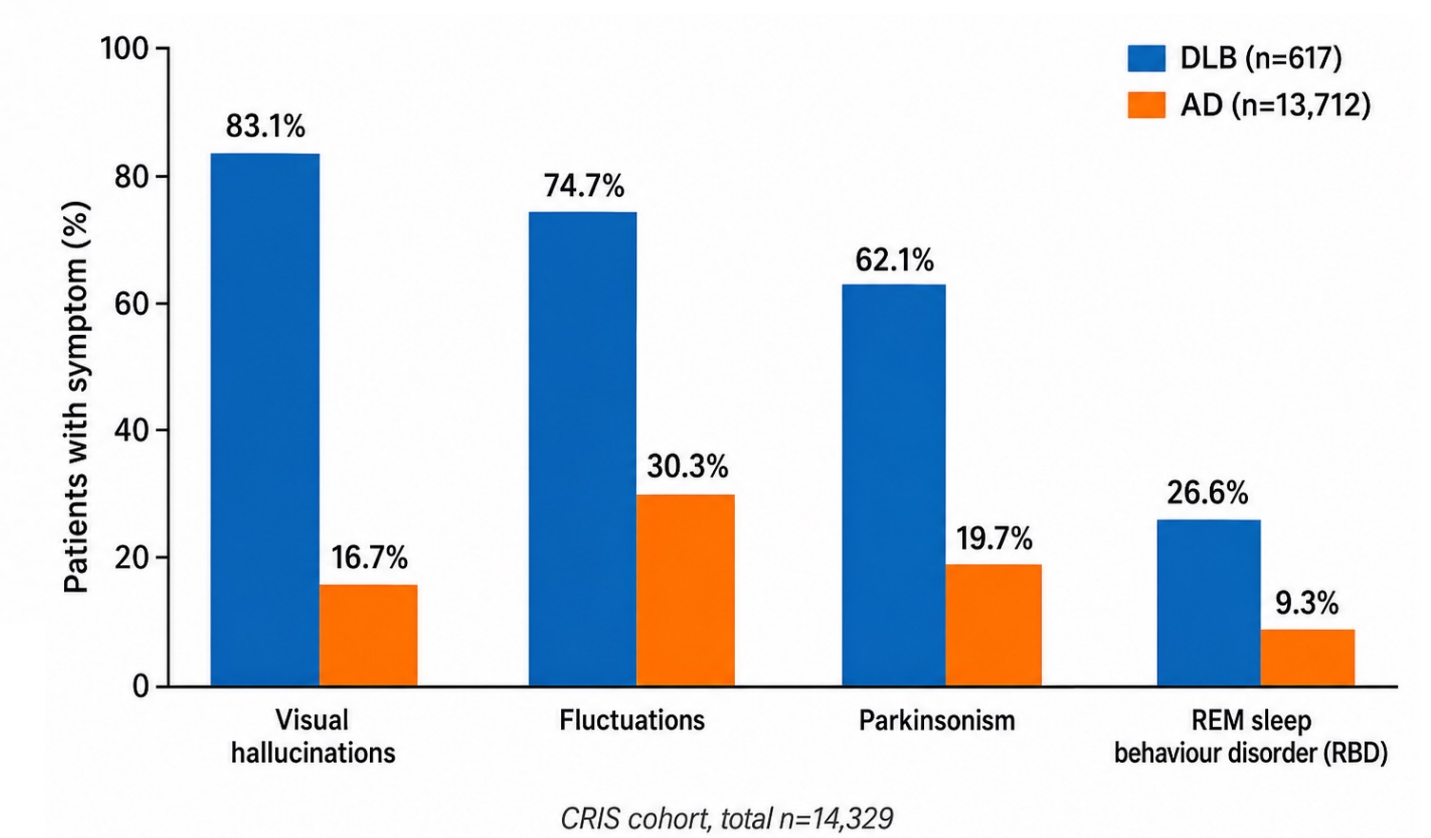
Key points

- Electronic health records (EHRs) provide unique opportunities to study dementia subtypes that are difficult to recruit into prospective studies, such as dementia with Lewy bodies (DLB).
- DLB remains substantially underdiagnosed in routine clinical practice, despite its distinct clinical profile and management implications.
- We developed natural language processing (NLP) algorithms to identify the four core DLB features visual hallucinations, cognitive fluctuations, parkinsonism, and REM sleep behaviour disorder (RBD) from routine clinical free text.
- Among patients later diagnosed with DLB, core symptoms were frequently documented at first presentation to services: visual hallucinations (57%), fluctuations (43%), parkinsonism (35%), and RBD features (12%).
- DLB core features were also common among patients subsequently diagnosed with Alzheimer's disease (AD). Across the full clinical record, 18.7% of patients diagnosed with AD had ≥ 2 DLB core features, potentially meeting criteria for probable DLB.
- Patients with AD and ≥ 2 DLB core symptoms formed a heterogeneous group: clinically closer to DLB in age at referral and diagnosis, but more similar to AD in gender and marital status, suggesting that NLP should not be viewed as diagnostic on its own but as a useful flag for clinical review and possible re-evaluation of AD diagnoses.
- A substantial proportion of patients diagnosed with AD exhibit symptom profiles consistent with DLB, highlighting potential under-recognition in clinical practice.
- NLP-based case finding may help detect previously unrecognised DLB, enabling earlier diagnosis, more appropriate treatment, and reduced exposure to potentially harmful medications.
- Beyond clinical care, NLP could improve identification of DLB research cohorts and facilitate recruitment into observational studies and clinical trials.

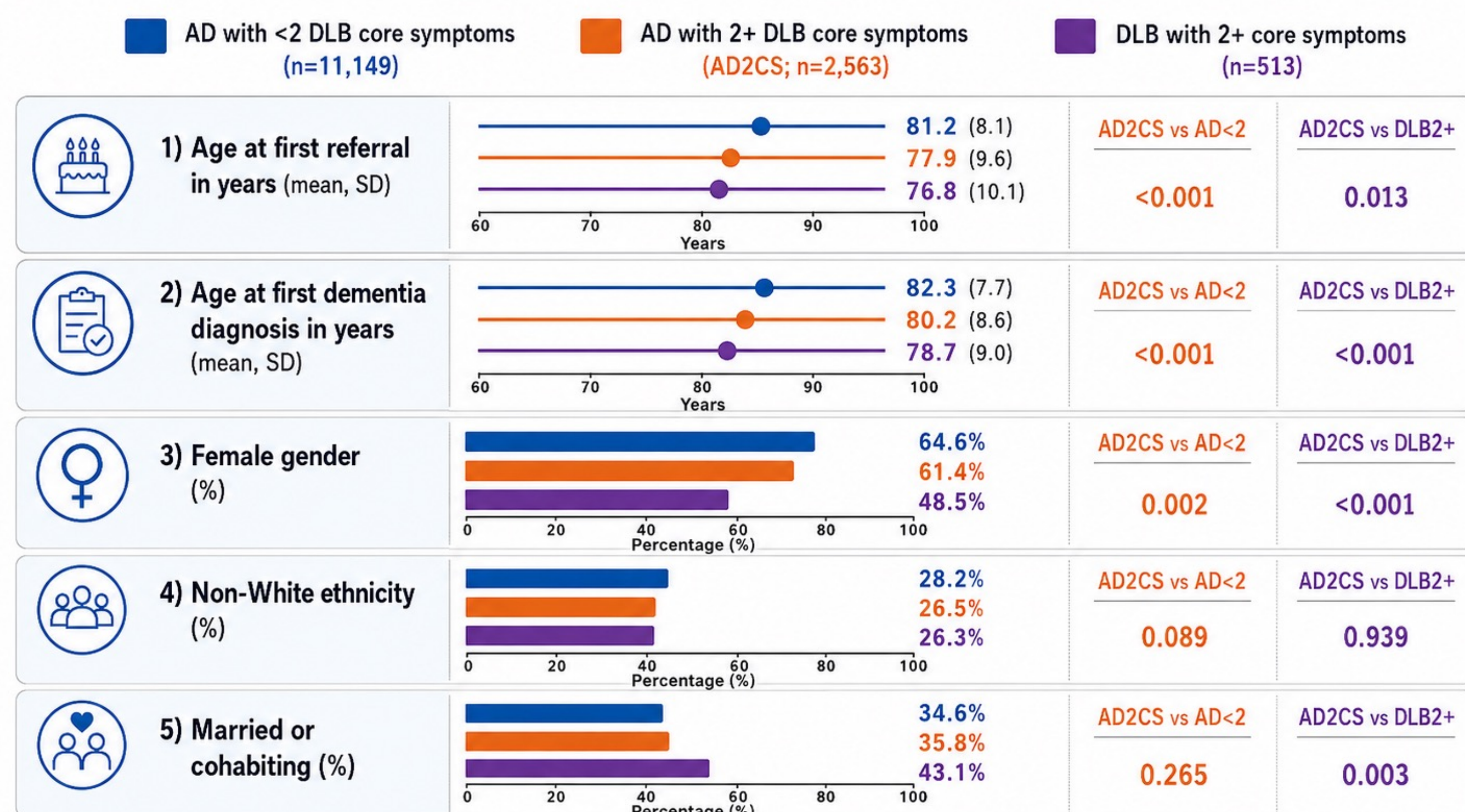
Publication in *npj Aging*:



Prevalence of DLB core symptoms across the electronic health record:



AD with <2 DLB core symptoms vs AD with 2+ DLB core symptoms vs DLB with 2+ core symptoms



Background: Dementia with Lewy bodies (DLB) is a common but frequently underdiagnosed dementia subtype, often misclassified as Alzheimer's disease (AD) due to overlapping cognitive and functional features. Accurate identification is clinically important given DLB's poorer prognosis, differential treatment response, and increased sensitivity to antipsychotic medications. Electronic health records (EHRs) contain rich clinical information, much of it recorded as free text and not routinely used for diagnostic case identification. Natural language processing (NLP) may enable scalable extraction of clinically meaningful symptoms to support improved detection of DLB in routine care.

Methods: Anonymised EHR data were accessed via the Clinical Record Interactive Search (CRIS) platform from South London and Maudsley NHS Foundation Trust. NLP applications were deployed to identify the four core clinical features of DLB (visual hallucinations, cognitive fluctuations, parkinsonism, and REM sleep behaviour disorder) from free-text clinical notes. Patients diagnosed with AD or DLB between 2008 and 2021 were included. Demographic characteristics and prevalence of DLB core features were compared between diagnostic groups, with a focus on patients diagnosed with AD who exhibited two or more DLB core features, consistent with possible or probable DLB.

Results: Of 14,329 patients identified, 617 (4.3%) had a recorded diagnosis of DLB and 13,712 (95.7%) a recorded diagnosis of AD. Median time (interquartile range) from first presentation to a specialist service to subtype diagnosis was 7.4 (1.8-28.8) months for patients with DLB and 4.4 (2.0-13.5) months for patients with AD. All core symptoms detected through NLP were significantly more common in patients with DLB versus AD. At the time of first presentation to specialist services the proportion of the core symptoms in patients with DLB were as follows (in brackets proportions across the whole record): 57.2% (83.1%) with visual hallucinations, 43.3% (74.7%) with fluctuations, 35.2% (62.1%) with Parkinsonism, and 12.3% (26.6%) with RBD. Core symptoms of DLB were also detected in a considerable proportion of patients with AD as follows: 13.4% (30.2%) with fluctuations, 9.3% (19.7%) with Parkinsonism, 7.3% (16.7%) with visual hallucinations, and 4.0% (9.3%) with RBD. Of patients the 13,712 diagnosed with AD, 783 (5.7%) met criteria for probable DLB at the time of first presentation and 2,563 (18.7%) if considering their whole electronic health record. This subgroup showed demographic and clinical characteristics intermediate between typical AD and DLB cases, suggesting a mixture of underdiagnosed DLB and mixed pathology.

¹ Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

² South London and Maudsley NHS Foundation Trust, London, UK

³ Department of Psychiatry, School of Clinical Medicine, University of Cambridge, Cambridge, UK

⁴ Cambridgeshire and Peterborough NHS Foundation Trust, Cambridge, UK