DOI:
10.3233/JAD-230887

Title: Code-based Algorithms for Identifying Dementia in Electronic Health Records: Bridging the Gap between Theory and Practice

This is the Author's Accepted Manuscript version of the article: Chen S, Wang Y, Mueller C. Code-Based Algorithms for Identifying Dementia in Electronic Health Records: Bridging the Gap Between Theory and Practice. J Alzheimers Dis. 2023;95(3):941-943. doi: 10.3233/JAD-230887. PMID: 37718822.

Shanquan Chen, Assistant professor, Ph.D

International Centre for Evidence in Disability, London School of Hygiene & Tropical Medicine, London, United Kingdom, WC1E 7HT shanquan.chen@lshtm.ac.uk

Yuqi Wang, M.Sc

Department of Computer Science, University College London, London, United Kingdom WC1E 6BT yuqiwang000@gmail.com

Christoph Mueller, MD

King's College London, London, United Kingdom

South London and Maudsley NHS Foundation Trust, London, United Kingdom christoph.mueller@kcl.ac.uk

ORCID: 0000-0001-9816-1686

Corresponding author:

Shanquan Chen, Assistant professor, Ph.D

International Centre for Evidence in Disability, London School of Hygiene & Tropical Medicine, London, United Kingdom, WC1E 7HT shanquan.chen@lshtm.ac.uk

Abstract

Code-based algorithms are crucial tools in the detection of dementia using electronic health record (EHR) data, with broad applications in medical research and healthcare. Vassilaki et al.'s study explores the efficacy of code-based algorithms in dementia detection using HER-data, achieving approximately 70% sensitivity and positive predictive value. Despite the promising results, the algorithms fail to detect around 30% of dementia cases, highlighting challenges in distinguishing cognitive decline factors. The study emphasizes the need for algorithmic improvements and further exploration across diverse healthcare systems and populations, serving as a critical step toward bridging gaps in dementia care and understanding.

Keywords: code-based algorithm; electronic health record; Dementia, Alzheimer's disease

Code-based algorithms, especially in the context of electronic health record (EHR) analysis, are computational strategies programmed to recognize and categorize health conditions like dementia using specific codes[1, 2]. These codes can include the International Classification of Diseases (ICD) codes, prescription identifiers, procedural tags, and other universally recognized markers. Their prominence in contemporary healthcare and medical research has surged due to the widespread digitalization of health records and the concomitant data deluge. With EHRs becoming ever more extensive and accessible, they provide an invaluable pool for health data analysis and medical research. They facilitate large-scale observational research and epidemiological studies, from identifying and tracking diseases to understanding healthcare resource utilization and predicting health outcomes in real-world settings[3, 4]. Cohorts assemble from EHR more generalisability to clinical populations than de novo recruited cohorts or trial participant samples which are subject to substantial selection.

Nevertheless, several challenges and uncertainties persist with routinely collected data, one being the way dementia diagnoses are defined and ascertained through algorithms. Their precision is largely reliant on the integrity of the source EHR data, which may be compromised by inconsistent or inadequate coding practices. Sensitivity to shifts in coding protocols over time could introduce inaccuracies in trend analysis. Code-based algorithms usually focus on structured data, such as ICD codes, potentially overlooking pertinent information embedded within unstructured components of EHRs, like nuanced clinical observations or subtle indicators of disease progression[5]. While their application is extensive, our grasp of code-based algorithms is not exhaustive. Strategies to augment their accuracy, robustness, and aptitude for managing diverse and messy EHR data remain ongoing areas of investigation.

Vassilaki et al.'s study represents a valuable contribution to this discussion[6]. The study applied a code-based algorithm to a cohort of 5,316 participants, aiming to identify dementia cases. It compared the algorithm's performance against a reference standard diagnosis provided by the Mayo Clinic Study of Aging (MCSA), evaluated the characteristics of false positives and false negatives, and examined the overall efficiency of the algorithm in diagnosing dementia[6].

The study found that the algorithm identified dementia with a sensitivity and positive predictive value of around 70%, demonstrating reasonably good accuracy[6]. However, the obvious limitation of the algorithm is its failure to detect approximately 30% of dementia cases, whereby the misidentified patients shared similar features. They are often older, more likely to have mild cognitive impairment (MCI) at baseline and have higher comorbidity burden. This highlights the inherent difficulty in distinguishing cognitive decline due to the progression of dementia, natural aging, or physical comorbidities. Such misidentification also reflects the challenges faced by clinicians. Vassilaki et al.'s

study suggests that demographic and health factors can be the potential avenues for future adjustments[6]. While the study did not detail specific ways, it opens the door for potential enhancements through the incorporation of unstructured EHR data via Natural Language Processing (NLP)[7] or the inclusion of laboratory test data[8]. However, these improved algorithms may still face a high percentage of misidentifications, given the extremely similar common features of false positives or false negatives.

Vassilaki et al.'s study further gives insights into the practicalities and hindrances of using code-based algorithms for dementia detection. It underscores the necessity to improve these algorithms to minimize false negatives and positives and illustrates how various demographic and health factors could be harnessed to fine-tune these tools. The practical implications of these scores, such as the undue burden on healthcare systems from false positives and the neglect of needs from false negatives, demand further exploration[9]. This underexplored dimension adds a layer of complexity to the already intricate field of algorithm-driven case finding. The potential distortion of statistical values, such as hazard ratios, due to a 30% error rate, also necessitates careful consideration in clinical research and intervention assessments[5].

Lastly, as Maria Vassilaki's mentioned, we still lack comprehensive knowledge about code-based algorithms' performance across different healthcare systems and populations, considering the variation in EHR practices and population health characteristics[6, 10, 11]. Evidence indicated that such variation exist even within the same setting but between primary care and specialist care[12]. The true potential and limitations of code-based algorithms, therefore, warrant further investigation and validation from different countries or settings.

In conclusion, Maria Vassilaki et al.'s study not only contributes valuable insights into the application and constraints of code-based algorithms in dementia detection but also stimulates further reflection and inquiry into this multifaceted field. It's a step towards bridging the theoretical and practical divides, opening new horizons for future research and improvements in dementia care and understanding.

Authors' contributions

SC searched the literature and wrote the manuscript. YM and CM was a major contributor in writing the manuscript and substantially revised it. The authors read and approved the final manuscript.

Funding

SC was supported by the PENDA, funded by the UK Foreign, Commonwealth and Development Office. CM is part-funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Data Availability

Not applicable.

Ethics approval and consent to participate

Not applicable.

Conflict of Interest

The authors have no conflict of interest to report. SC is an Editorial Board Member of this journal, but was not involved in the peer-review process nor had access to any information regarding its peerreview.

Acknowledgments

The authors have no acknowledgments to report.

References

[1] Riedel O, Braitmaier M, Langner I (2023) Dementia in health claims data: The influence of different case definitions on incidence and prevalence estimates. *Int J Methods Psychiatr Res* **32**, e1947.

[2] McCarthy EP, Chang CH, Tilton N, Kabeto MU, Langa KM, Bynum JPW (2022) Validation of Claims Algorithms to Identify Alzheimer's Disease and Related Dementias. *J Gerontol A Biol Sci Med Sci* **77**, 1261-1271.

[3] Ponjoan A, Garre-Olmo J, Blanch J, Fages E, Alves-Cabratosa L, Martí-Lluch R, Comas-Cufí M, Parramon D, Garcia-Gil M, Ramos R (2020) Is it time to use real-world data from primary care in Alzheimer's disease? *Alzheimer's Research & Therapy* **12**, 60.

[4] Ilomäki J, Lai EC-C, Bell JS (2020) Using clinical registries, administrative data and electronic medical records to improve medication safety and effectiveness in dementia. *Current Opinion in Psychiatry* **33**.

[5] Sauer CM, Chen LC, Hyland SL, Girbes A, Elbers P, Celi LA (2022) Leveraging electronic health records for data science: common pitfalls and how to avoid them. *Lancet Digit Health* **4**, e893e898.

[6] Maria Vassilaki, Sunyang Fu, Luke R. Christenson, Muskan Garg, Ronald C Petersen, Jennifer St. Sauver, Sohn S (2023) Characterizing Performance Gaps of a Code-Based Dementia Algorithm in a Population-Based Cohort of Cognitive Aging. *Journal of Alzheimer's Disease* **xx**, xx.

[7] Chen S, Price AC, Cardinal RN, Moylett S, Kershenbaum AD, Fitzgerald J, Mueller C, Stewart R, O'Brien JT (2022) Association between antidementia medication use and mortality in people diagnosed with dementia with Lewy bodies in the UK: A retrospective cohort study. *PLoS Med* **19**, e1004124.

[8] McKeith IG, Boeve BF, Dickson DW, Halliday G, Taylor JP, Weintraub D, Aarsland D, Galvin J, Attems J, Ballard CG, Bayston A, Beach TG, Blanc F, Bohnen N, Bonanni L, Bras J, Brundin P, Burn D, Chen-Plotkin A, Duda JE, El-Agnaf O, Feldman H, Ferman TJ, Ffytche D, Fujishiro H, Galasko D, Goldman JG, Gomperts SN, Graff-Radford NR, Honig LS, Iranzo A, Kantarci K, Kaufer D, Kukull W, Lee VMY, Leverenz JB, Lewis S, Lippa C, Lunde A, Masellis M, Masliah E, McLean P, Mollenhauer B, Montine TJ, Moreno E, Mori E, Murray M, O'Brien JT, Orimo S, Postuma RB, Ramaswamy S, Ross OA, Salmon DP, Singleton A, Taylor A, Thomas A, Tiraboschi P, Toledo JB, Trojanowski JQ, Tsuang D, Walker Z, Yamada M, Kosaka K (2017) Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium. *Neurology* **89**, 88-100.

[9] Zhu CW, Ornstein KA, Cosentino S, Gu Y, Andrews H, Stern Y (2019) Misidentification of Dementia in Medicare Claims and Related Costs. *J Am Geriatr Soc* **67**, 269-276.

[10] Hayat S, Luben R, Khaw KT, Wareham N, Brayne C (2022) Evaluation of routinely collected records for dementia outcomes in UK: a prospective cohort study. *BMJ Open* **12**, e060931.

[11] Festa N, Price M, Moura L, Blacker D, Normand SL, Newhouse JP, Hsu J (2022) Evaluation of Claims-Based Ascertainment of Alzheimer Disease and Related Dementias Across Health Care Settings. *JAMA Health Forum* **3**, e220653.

[12] Davis KAS, Mueller C, Ashworth M, Broadbent M, Jewel A, Molokhia M, Perera G, Stewart RJ (2021) What gets recorded, counts: dementia recording in primary care compared with a specialist database. *Age Ageing* **50**, 2206-2213.