Dementia Prediction Using Routinely Collected Health Data

Luis H. John[1], Jan A. Kors[1], Jenna M. Reps[2], Peter R. Rijnbeek[1]

[1]Erasmus University Medical Center, Rotterdam, The Netherlands

[2]Janssen Research and Development, Raritan, NJ, United States

EUROPEAN HEALTH DATA & EVIDENCE NETWORK

Background

Dementia is a general term to describe various brain diseases (e.g. Alzheimer, vascular dementia, Lewy body dementia, and frontotemporal lobe dementia) that affect cognition and may lead to mental degradation. All types of dementia are observed to be chronic and progressive.

We performed a literature review of prognostic dementia risk models and identified common drawbacks to be relatively small sample sizes, a limited set of manually selected predictors, as well as reliance on non-routinely collected information (e.g. years of education, pesticide exposure). In addition, prediction is performed across a wide range of age groups, generally specified as *above the age of 65*. We believe these factors may prevent wider adoption of the proposed models.

To overcome this, we aim to predict 5-year dementia risk using routinely collected health data from various databases. Our approach evaluates predictive performance in a real-world setting, without pre-selecting predictors or imputing data.

Methods

We defined several target cohorts by stratifying patients above the age of 60 into 5-year age groups. The index date was defined as the first outpatient visit since entering the age group (before 1 January 2013). As additional inclusion criteria, we require 5 years of follow up time and the absence of a condition occurrence of dementia or other cognitive impairments any time prior. Patients also should not have had a condition occurrence of a disease indicating subtypes of dementia (e.g. Pick’s disease, parkinsonism, Huntington’s chorea).

Prediction is performed using LASSO logistic regression. We specified a time-at-risk of 5 years and a washout period of 3 years before the index date. Several thousand predictors have been included covering demographics, condition occurrences, drug exposures, procedures performed, measurements, observations, and device exposures. The following table summarizes the study populations across four databases of interest. IBM MDCR is a U.S. claims database that holds patient data of retirees, Iqvia Germany DA comprises patient data of primary care interventions in Germany, Optum databases hold U.S. claims data including lab analyte results and administrative data, as well as socioeconomic information such as ethnicity, income, education, marital status.

Age groups	IBM MDCR		Iqvia Germany DA		Optum Extended SES		Optum Panther	
	Records	Events	Records	Events	Records	Events	Records	Events
60 – 64	15,282	733 (4.8%)	147,815	1,556 (1.1%)	291,563	6325 (2.2%)	1,039,545	11,889 (1.1%)
65 – 69	308,948	13,718 (4.4%)	131,690	3,295 (2.6%)	294,022	12,301 (4.2%)	866,669	21,426 (2.5%)
70 – 74	478,317	29,366 (6.1%)	169,559	8,366 (4.9%)	363,263	26,014 (7.2%)	700,056	36,658 (5.2%)
75 – 79	406,050	50,440 (12.4%)	116,177	11,277 (9.7%)	375,272	59,316 (15.8%)	544,754	57,197 (10.5%)
80 – 84	300,908	67,687 (22.5%)	59,979	11,214 (18.7%)	316,637	96,127 (30.4%)	640,323	148,613 (23.2%)
85 – 89	161,971	60,340 (37.3%)	24,698	7,924 (32.1%)				
90 – 94	55,018	29,841 (54.2%)	5,680	2,731 (48.1%)				
95 – 99	10,907	7,595 (69.6%)						

Results

A prediction model has been developed for each age group within each database. We report discrimination ability using the area under the Receiver Operating Characteristic curve (AUROC) as well as model calibration using the calibration slope.

Age groups	MDCR		Iqvia Germany DA		Optum Extended SES		Optum Panther	
	AUROC	Cal. Slope	AUROC	Cal. Slope	AUROC	Cal. Slope	AUROC	Cal. Slope
60 – 64	0.83	1.22	0.77	1.05	0.77	1.14	0.74	1.16
65 – 69	0.74	1.03	0.75	1.04	0.74	1.05	0.71	1.10
70 – 74	0.74	1.03	0.73	0.91	0.74	1.06	0.69	1.07
75 – 79	0.74	1.03	0.74	1.01	0.73	1.04	0.67	1.03
80 – 84	0.74	1.00	0.73	0.94	0.74	1.00	0.68	1.04
85 – 89	0.74	1.01	0.70	1.02				
90 – 94	0.75	1.00	0.66	0.96				
95 – 99	0.74	1.07						

Preliminary results showed reasonable discrimination performance and calibration, with the MDCR models showing highest average performance and stability across all age groups. By direct comparison to some models in literature, which do not use routinely collected health data (AUROC between 0.73 and 0.89) our model at this point in time underperform slightly. [1-4]

Conclusions

We developed various prediction models using routinely collected health data from four different databases. We believe that our models, which do not rely on manual predictor selection or data imputation and use data commonly available in observational health databases, will be easier to implement and update, which may result in a wider adoption in clinical practice.

By stratifying patients into age groups we observed our models’ discrimination performances for Iqvia Germany DA and Optum Panther to decline in older populations. Future work will focus on studying the frequency of variables selected by each model in these different age groups to deepen our understanding of dementia prediction.

Next steps will focus on externally validating the prediction models in the respective other databases as well as further improving model performances.

1. Barnes, D. E., Covinsky, K. E., Whitmer, R. A., Kuller, L. H., Lopez, O. L., & Yaffe, K. (2010). Dementia risk indices: A framework for identifying individuals with a high dementia risk. *Alzheimer's & dementia: the journal of the Alzheimer's Association*, *6*(2), 138.
2. Ebbert, M. T., Ridge, P. G., Wilson, A. R., Sharp, A. R., Bailey, M., Norton, M. C., ... & Kauwe, J. S. (2014). Population-based analysis of Alzheimer’s disease risk alleles implicates genetic interactions. *Biological psychiatry*, *75*(9), 732-737.
3. Vuoksimaa, E., Rinne, J. O., Lindgren, N., Heikkilä, K., Koskenvuo, M., & Kaprio, J. (2016). Middle age self-report risk score predicts cognitive functioning and dementia in 20–40 years. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *4*, 118-125.
4. Mura, T., Baramova, M., Gabelle, A., Artero, S., Dartigues, J. F., Amieva, H., & Berr, C. (2017). Predicting dementia using socio-demographic characteristics and the Free and Cued Selective Reminding Test in the general population. *Alzheimer's research & therapy*, *9*(1), 21.

Contact: l.john@erasmusmc.nl