노인재활

게시일시 및 장소 : 10 월 18 일(금) 08:30-12:20 Room G(3F)

질의응답 일시 및 장소 : 10 월 18 일(금) 10:00-10:45 Room G(3F)

P 1-19

EHR Based Prediction of Future Incidence of Alzheimer's Disease Using Machine Learning

Han Eol Cho[1*], Ji Hwan Park[2], Jong Hun Kim[3], Melanie Wall[4], Yaakov Stern[4], Hyunsun Lim[5], Shinjae Yoo[2], Justin Byun[1], Gun Jae Lee[6], Jiook Cha[4,7†], Hyoung-Seop Kim[6†]

Yonsei University College of Medicine, Department of Rehabilitation Medicine, Gangnam Severance Hospital and Rehabilitation Institute of Neuromusular Disease[1], Brookhaven National Laboratory, Computational Science Initiative[2], National Health Insurance Service Ilsan Hospital, Department of Neurology[3], Columbia University, Department of Psychiatry[4], National Health Insurance Service Ilsan Hospital, Research and Analysis Team[5], National Health Insurance Service Ilsan Hospital, Department of Physical Medicine and Rehabilitation[6], Columbia University, Data Science Institute[7]

Background

Accurate prediction of future incidence of Alzheimer's disease may facilitate intervention strategy to delay disease onset. Existing AD risk prediction models require collection of biospecimen (genetic, CSF, or blood samples), cognitive testing, or brain imaging. Conversely, EHR provides an opportunity to build a completely automated risk prediction model based on individuals' history of health and healthcare. We tested machine learning models to predict future incidence of AD using administrative EHR in individuals aged 65 or older.

Methods

We obtained de-identified EHR from Korean elders age above 65 years old (N=40,736) collected between 2002 and 2010 in the Korean National Health Insurance Service database system. Consisting of Participant Insurance Eligibility database, Healthcare Utilization database, and Health Screening database, our EHR contain 4,894 unique clinical features including ICD-10 codes, medication codes, laboratory values, history of personal and family illness, and socio-demographics. Our event of interest was new incidence of AD defined from the EHR based on both AD codes and prescription of anti-dementia medication. Two definitions were considered: a more stringent one requiring a diagnosis and dementia medication resulting in n=614 cases ("definite AD") and a more liberal one requiring only diagnostic codes (n=2,026; "probable AD"). We trained and validated a random forest, support vector machine, and logistic regression to predict incident AD in 1,2,3, and 4 subsequent years using the EHR available since 2002. The length of the EHR used in the models ranged from 1,571 to 2,239 days. Model training, validation, and testing was done using iterative (5 times), nested, stratified 5-fold cross validation.

Results

Average duration of EHR was 1,936 days in AD and 2,694 days in controls. For predicting future incidence of AD using the "definite AD" outcome, the machine learning models showed the best performance in 1 year prediction with AUC of 0.781; in 2 year, 0.739; in 3 year, 0.686; in 4 year, 0.662. Using "probable AD" outcome, the machine learning models showed the best performance in 1 year prediction with AUC of 0.730; in 2 year, 0.645; in 3 year, 0.575; in 4 year, 0.602. Important clinical features selected in logistic regression included hemoglobin level (b=-0.902), age (b=0.689), urine protein level (b=0.303), prescription of Lodopin (antipsychotic drug) (b=0.303), and prescription of Nicametate Citrate (vasodilator) (b=-0.297).

Conclusion

This study demonstrates that EHR can detect risk for incident AD. This approach could enable risk-specific stratification of elders for better targeted clinical trials.

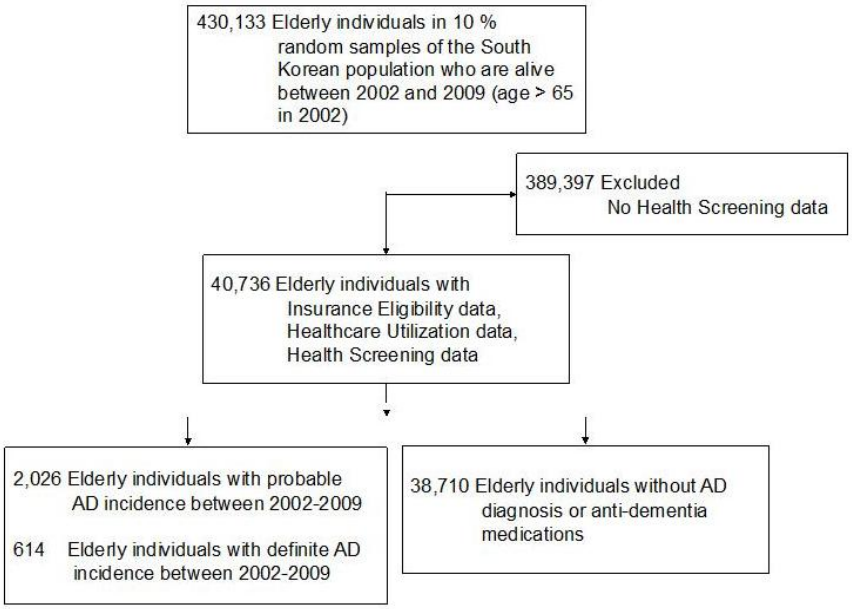

Figure 1. Consort Diagram

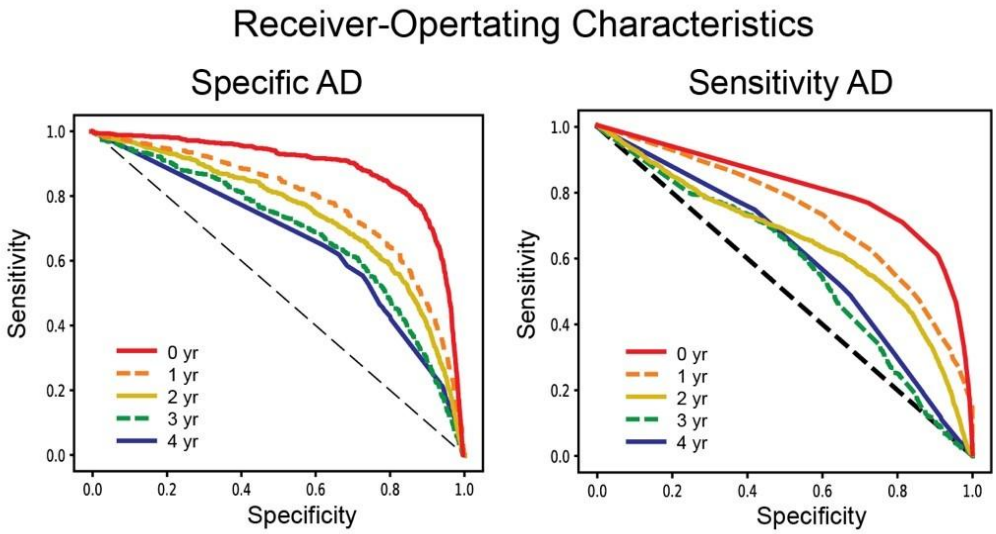

Figure 2. Receiver-Operating Characteristic plots are shown for 0,1,2,3,4-year prediction. Incident AD was defined based on ICD-10 AD codes and anti-dementia medication for AD, "Definite AD", or based on AD codes only, "Probable AD".

Table 1. Top ten features and weights from logistic regression (0-yr prediction).

Type of data	Name	b value
health checkup	hemoglobin	-0.902
demography	age	0.689
health checkup	urine protein	0.303
medication	Zotepine (antipsychotic drug)	0.303
medication	Nicametate Citrate (vasodilator)	-0.297
disease code	other degenerative disorders of nervous system in diseases classified elsewhere	-0.292
disease code	disorders of external ear in diseases classified elsewhere	-0.274
medication	Tolfenamic acid 200mg (pain killer)	-0.266
disease code	adult respiratory distress syndrome	-0.259
medication	Eperisone Hydrochloride (antispasmodic drug)	0.255