Developing a risk prediction model for the identification of new Alzheimer's Disease (AD) in elderly patients

Verma V, Brooks L, Krebs B, Sharma S, Markan R, Chawla S, Gupta A, Ashra P, Nayyar A , Gaur A, Kukreja I, Dawar V

Optum Life Sciences

RWD121

Introduction

- Alzheimer's disease (AD) is a progressive, neurodegenerative disease that creates complex challenges in patients' everyday life and leads to significant economic burden on patients and caregivers. An estimated 6.7 million[1] Americans aged 65 and older are living with Alzheimer's Disease in 2023; 73%[1] are aged 75 or older. This translates to ~10.7% of US population with age 65 and older affected by the disease. There has been a substantial increase in deaths (more than doubled between 2000 to 2019) related to Alzheimer's Disease as well. Alzheimer's and other dementias are estimated to cost $345 billion[1] to the US economy, in the year 2023.
- Diagnosis of an AD patient is difficult, especially in its early stages. Integrating clinical data (EHR data) with claim data would help identify patient more accurately and early during their prognosis phase, leading to better patient outcomes

Objective

Aim of this study is to develop a risk prediction model, for identification of new Alzheimer's patients in elderly age group, that utilizes integration of EHR data with claim data.

Methods

- A retrospective study using Optum® de-identified Market Clarity Dataset (linked claims and EHR of patients) was done among elderly (≥60 years) in the US.
- Patients with at least 2 outpatient claims/encounters at least 30 days apart OR at least one inpatient diagnosis recorded between 2019 and 2021 and not having any claim/encounter history for AD in 3-year Pre index period were selected.
- 3-years of continuous pre-index eligibility was ensured in which risk factors and symptoms were identified as predictor variable.
- Diagnoses were only considered for claims model while both diagnoses and symptoms (SDS terms) were considered for claims + EHR model
- Patients who had AD diagnosis were considered as case group and control was created by random sampling. To reduce the confounding effects of age, gender, region, race/ethnicity, the study matched cases to controls at a 1:2 ratio using propensity score matching (PSM) method.
- Data was randomly divided into 70:30 ratio for model training and testing purposes. A comparison of training and testing accuracy was conducted for each model to eliminate overfitting.
- Supervised ML techniques like Random Forest, XG Boost and logistic were executed to develop algorithms that predict occurrence of prodromal AD Cases.

Results

Fig 1: Logistic Regression Result

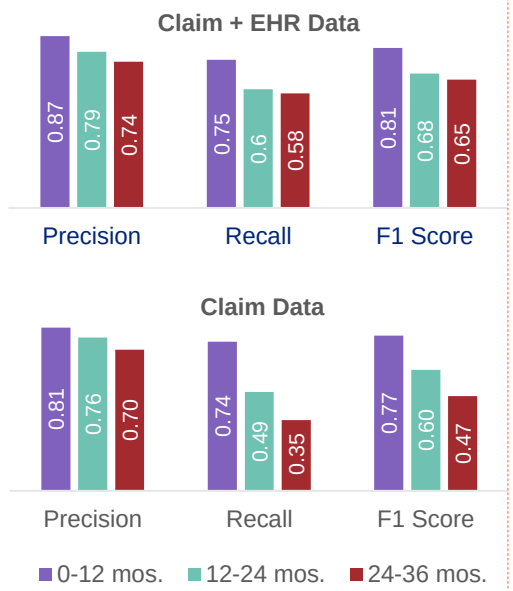
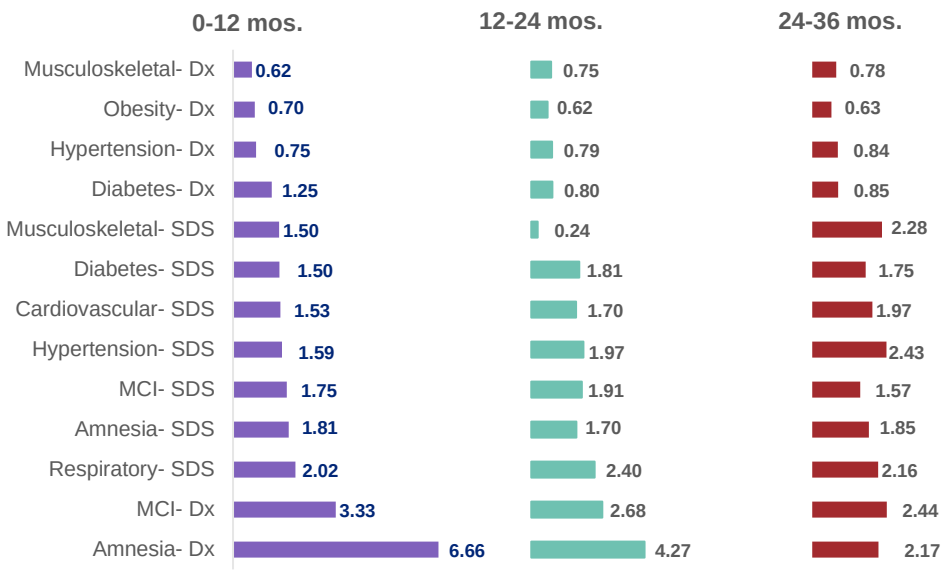


Fig 2: ODDS Ratio



- ~8K patients were identified to build prediction model.
- AUROC for Claim + EHR data vs. claims data ranged from 76%-84% vs 65%-83% for Logistic, 78%-87% vs 65%-83% for XG Boost, and 79%-86% vs 64%-84% for Random Forest, respectively.
- Maximum precision was observed in 0-12 months time frame (Fig 1). Percentage gain in precision was 18% for integrated data (Claim + EHR) (Fig 1) as compared to 16% for claim data (Fig 1) at 24-36 months and 0-12 months pre-index time frame.
- Fig 2 display odd ratio of three different time frame by utilizing integrated data (Claim + EHR)
- Odds ratio (OR) indicates that the probability of having AD in patients with risk factors like amnesia (OR: 6.66), MCI (OR: 3.33), hypertension symptoms (1.56) and cardiovascular symptoms (1.53) was higher in integrated data (Fig 3).

Conclusion

- Providers investigates clinical symptoms ahead of the diagnosis of AD during early phases of disease. Hence, by integrating clinical data (EHR) with claim data we can enable providers with more information that will aid them to predict disease more accurately during early phase and will further aid in early disease intervention.

Optum