Identifying clinical subgroups of Alzheimer's Disease (AD) patients from Optum's de-identified Market Clarity Database using machine learning techniques

Khan S , Markan R , Paul A , Khandelwal H, Nayyar A , Roy A, Dawar V, **Verma V** , Brooks L

Optum Life Sciences

RWD12

Introduction

- Alzheimer's is a neurodegenerative disease that is highly heterogeneous with varied symptoms (e.g., anxiety, depression, behavioral issues, etc.) and rate of progression
- The causes of AD and other dementias are not completely understood, but researchers believe they include a combination of various factors. Alzheimer's patients often suffer from other comorbidities such as cardiovascular disease, diabetes, etc.
- AD patients differ in disease management, which significantly increases the cost burden.
- Identifying clinical subgroups thereby helps to design personalized treatment and intervention programs according to their unique need

Objective

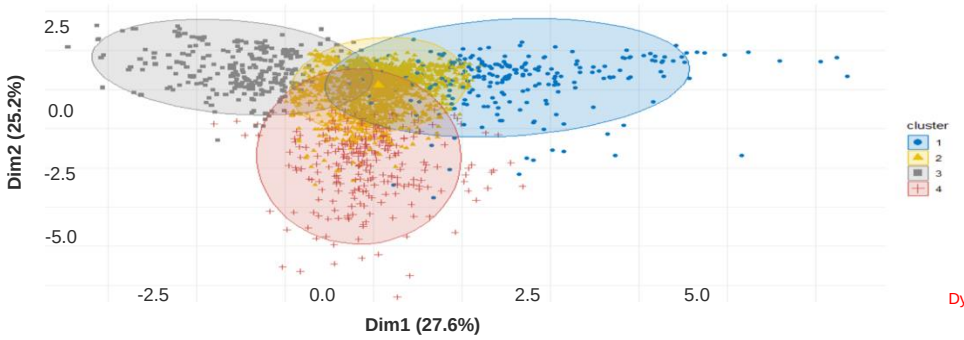
To identify clinically relevant patient segments of AD based on their signs & symptoms, comorbid conditions and other patient characteristic using different clustering techniques

Method

- AD patients were identified from Optum® de-identified Market Clarity Dataset, which links medical, and pharmacy claims with EHR data using ICD 9 and ICD 10 codes
- Patients with continuous eligibility (1-year pre and 2 years post index) > 60 years with at least 2 primary diagnosis of AD (irrespective of outpatient and inpatient) falling 30 days apart between 01 Jan 2018 and 31 Dec 2019 were considered
- Pre-index period: 1 year prior to diagnosis & post Index period: 2 years after diagnosis
- Patients were clustered into clinically relevant subgroups based on multimorbidity's, symptoms, demography and other patient characteristic. Patients with no symptoms, associated co-morbidities and prior claim of AD in the pre-index period were excluded from the analysis
- For subgroups, ML techniques like K-Means with multiple correspondence analysis (MCA) was used. Silhouette score was used to determine the optimal number of clusters (K)

Results

Fig1: Distribution of Alzheimer's patients' subgroups using K-Means clustering technique



- Among 4,463 identified AD patients, mean age was 80 years and predominant population were female patients (65%)

Fig3: Individual feature and related % of Alzheimer's subgroups

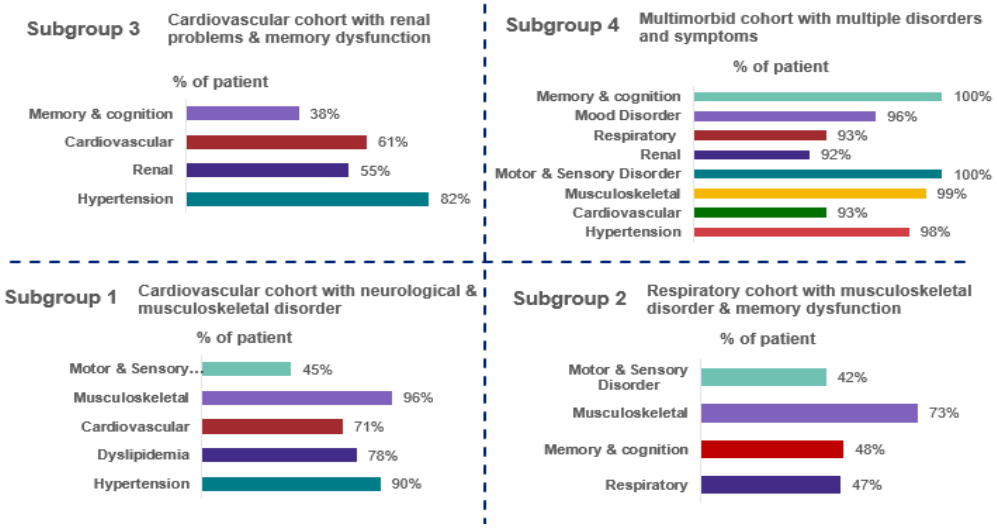
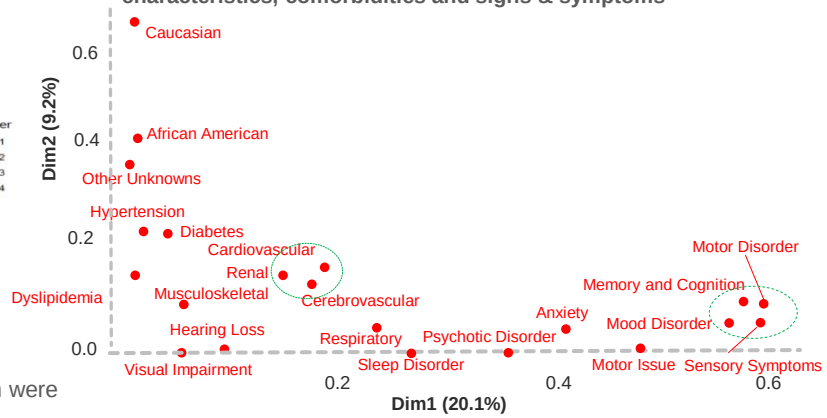


Fig2: 2-dimensional bi-plot of patient subgroups based on patient's characteristics, comorbidities and signs & symptoms



- K-Means clustering with Multiple Correspondence Analysis (MCA) gave most consistent subgroups based on distance measures cosine and Eigen values **[Fig.1]**
- Multiple iterations were performed to optimize the number of subgroups (K value) and 4 distinct subgroups were identified in the process
- 73% of sample showing similar characteristics, clubbed into one subgroup **[Fig.2]**
- Hypertension and Musculoskeletal were identified as prominent risk factor and showed a significant presence in most of the clusters.
- Mood disorder', 'Memory and cognition' and 'Neurological issues' were prominent symptoms where as 'Cardiovascular', 'Renal', 'Cerebrovascular' were prominent comorbidities observed. **[Fig. 3]**

Conclusions

- This study can be leveraged for personalized and targeted healthcare intervention among different clusters
- If clusters have similar genetic makeup increasing their risk of developing AD, early intervention in such cases can be beneficial in slowing disease progression as well as reducing the overall cost of care.

Optum