Segmentation Based on Disease Burden for
Primary Care Populations with Cardiometabolic Conditions

Authors: Xiaowei Yan, J.B. Jones, Hannah Husby, Jake Delatorre-Reimer, Farah Refai, Franciscos Sifakis, Karen MacDonald

Background

Population health management depends on segmenting patients with common clinical needs into distinct groups to facilitate targeted services. Historically, segmentation has relied on ad-hoc methods. Longitudinal electronic health records (EHRs) enable data-driven methods to create clinically meaningful segmentation recommendations that explicitly recognize disease burden and severity.

Methods and Settings

We completed a retrospective EHR analysis on adult primary care patients aged ≥35 with >1 cardiometabolic (CM) disease (i.e. hypertension, hyperlipidemia, diabetes) and one related clinical measure (blood pressure, HbA1c, or LDL-C) during the study period (10/1/2015 to 9/30/2017) in a large health care system in Northern California. For each patient, we generated a feature vector that included demographic, disease burden, disease complexity, and biometric measures, and imputed missing data values using a discriminant function method. To generate reliable clusters, we applied a consensus latent class analysis method, and repeated the process for 3-cluster, 4-cluster, and 5-cluster models.

Results

A five-cluster model was selected based on Akaike information criterion and Bayesian information criterion. Among the 182,884 patients identified, 10.6%, 13.6%, 25.4%, 22.3%, and 28.2% were assigned to clusters 1–5, respectively. Prevalence of CM conditions varied significantly, and three clusters had distinct cardiovascular disease (CVD) risk ranges. In cluster 1, the patient mean age was 67 years, average BMI was 32, 70% had all 3 CM conditions, and 88% had high CVD risk. Mean age for cluster 2 was similar to that of cluster 1 (mean=66 years), with most patients having hyperlipidemia (95%) and 38% having all 3 CM conditions. Cluster 3 was the oldest group (mean age=69 years), majority were female (68%) and had a high prevalence of psychiatric conditions (32% anxiety, 28% depression), and had highest care utilization (mean PCP visits in 2 years=8, specialist visits=13). Clusters 4 and 5 were similar in terms of age (59 and 60 years old) and prevalence of diabetes (12% vs. 13%) and CVDs (2%-3%).

Discussion

We used data-driven methods to cluster CM patients into more homogeneous subgroups to facilitate individualization of care. We will conduct clinical validation with clinicians to test distinction of clusters, clinical usefulness of segments, and care protocols that meet needs for each cluster.