Using graph visualization to look at the trajectories of events that lead to readmission

Abbas Shojaee
Yale University, abbas.shojaee@yale.edu

Isuru Ranasinghe
Yale University

Sudhakar Nuti
Yale University

Shu-Xia Li
Yale University

Harlan Krumholz
Yale University

Diagnostic cardiac catheterization

Big Data @ Yale: Visualizing Distinct Trajectories of Healthcare Utilization Patterns Leading to Heart Failure Readmission

Abbas Shojaee MD [1], Shu-Xia Li PhD [1], Isuru Ranasinghe MBChB, MMed, PhD [1], Sudhakar V. Nuti BA [1], Emily Bucholz, MPH [1], Rachel Dreyer PhD [1], Karthik Murugiah MBBS [1], Bobak Mortazavi PhD [1], Harlan M. Krumholz MD, SM [1]

[1] Center for Outcomes Research and Evaluation, Yale University - Yale New Haven Hospital

Background

The leading cause for hospital admission among adults older than 65 is Heart Failure (HF). Readmission is an important outcome for patients, providers, and policymakers and a key measure of healthcare quality. Numerous models for characterizing and predicting readmission risk have been developed with sub satisfactory results. Conventional statistical approaches (i.e. regression analysis) are limited in their ability to assess whether the sequence, timing, venue and intensity of patient's prior healthcare utilization may affect the risk of specific outcome. Accordingly, we visualized the sequences of events prior to hospitalization for heart failure to find common patient trajectories that may be used to identify underlying patterns and to improve prediction of heart failure readmission.

Key Question

Can we improve our understanding of the likelihood of readmission outcomes by looking at sequence, timing, venue and composition of patients' health care utilization before readmission.

Novelty & Innovation

Our method can potentially improve prediction and reveal important patterns associated with outcomes, which can provide insights into underlying mechanisms leading to readmission and targets for interventions.

Method

We identified 254,000 patients and 434,168 admissions with a primary diagnosis of Heart Failure (HF) and included 18,004 patients who had at least 4 admissions using the Agency for Healthcare Research and Quality (AHRQ) Clinical Classification System (CCS) code 108 in the inpatient database of the California Healthcare Cost and Utilization Project (2006-11). Based on availability of time dimension we considered admissions (n=104,035 events) and in-hospital procedures (n=102,596 events) as sequential events; and readmission outcome was defined as one of the three fuzzy readmission periods: prior to 25 days, around 30 days and more than 35 days. Using F# programming language we grouped all admissions of each patient and sorted admissions and the procedures for each patient on the time vector. We mapped sequences of events on the graph network built on QuickGraph library. We then calculated the gap between sequential admissions and connected each sequence of events to correspondent virtual node for readmission outcome. Transitive closure of subgraphs were used to show the direct effect of indirect sequences. We clustered all similar patients, admissions and procedures on the same pathways to reveal frequently observed patient trajectories. Clustering of similar groups was performed based on 40 variables including age, sex, race and all recorded comorbidities using a fuzzy-logic approach on numeric variables. The final graph consisted of 350,657 vertices (unique pathway points) and 487,505 edges. The graph network was then visualized with Gephi and Cytoscape software applications. We filtered out all event with occurrence <3 and in-out degree <2, to identify sequences of events that are more significant. The final graph, consisting of 5,702 events and 6,709 edges, was divided into three sub-graphs based on readmission outcomes.

Figures

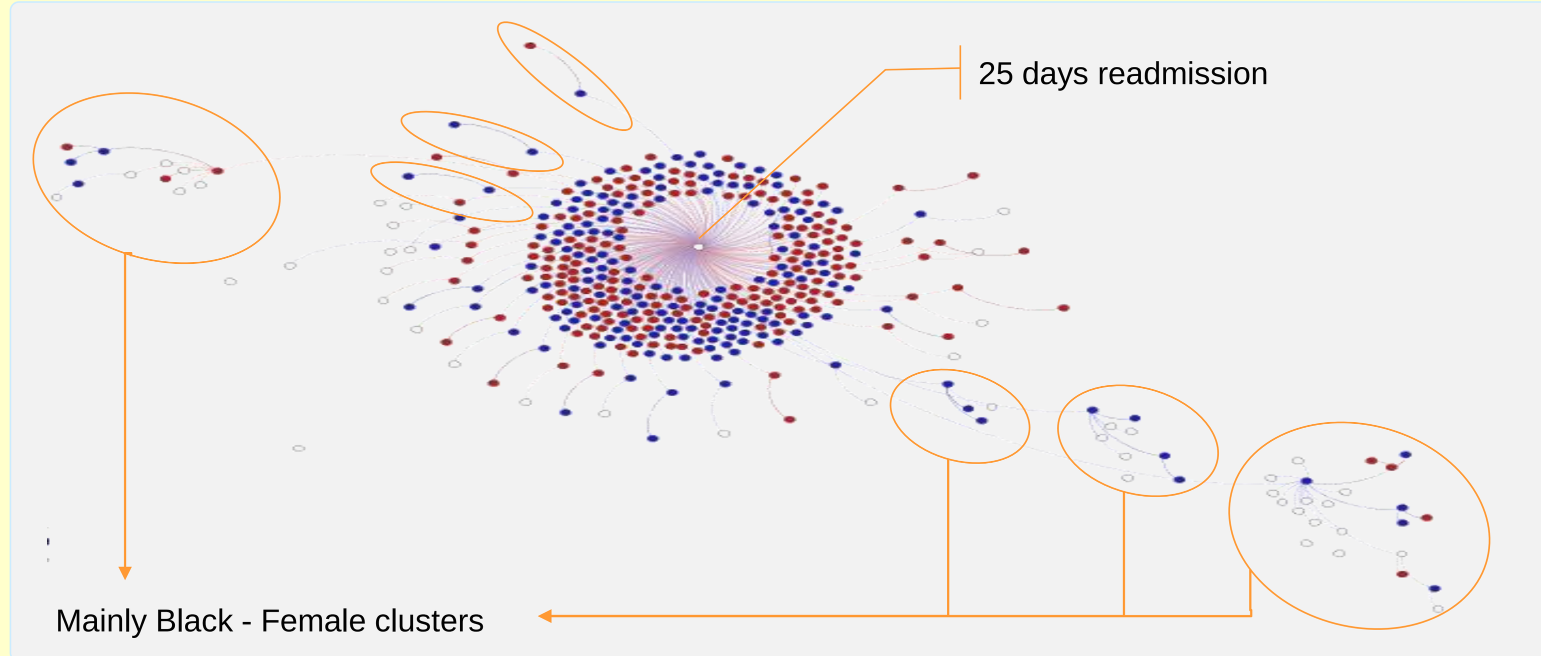


Fig1: Visualization of admissions (blue) and procedures(red). Circles show clusters with dominant black, female, observatory admissions.

Fig 2: Various common pathways toward 30 days readmission (the central point). Node size reflects frequency, admission are in blue and in-hospital procedures are shown in red.

Network diameter = 5,
Average path length: 1.689,
Clustering coefficient:0.021

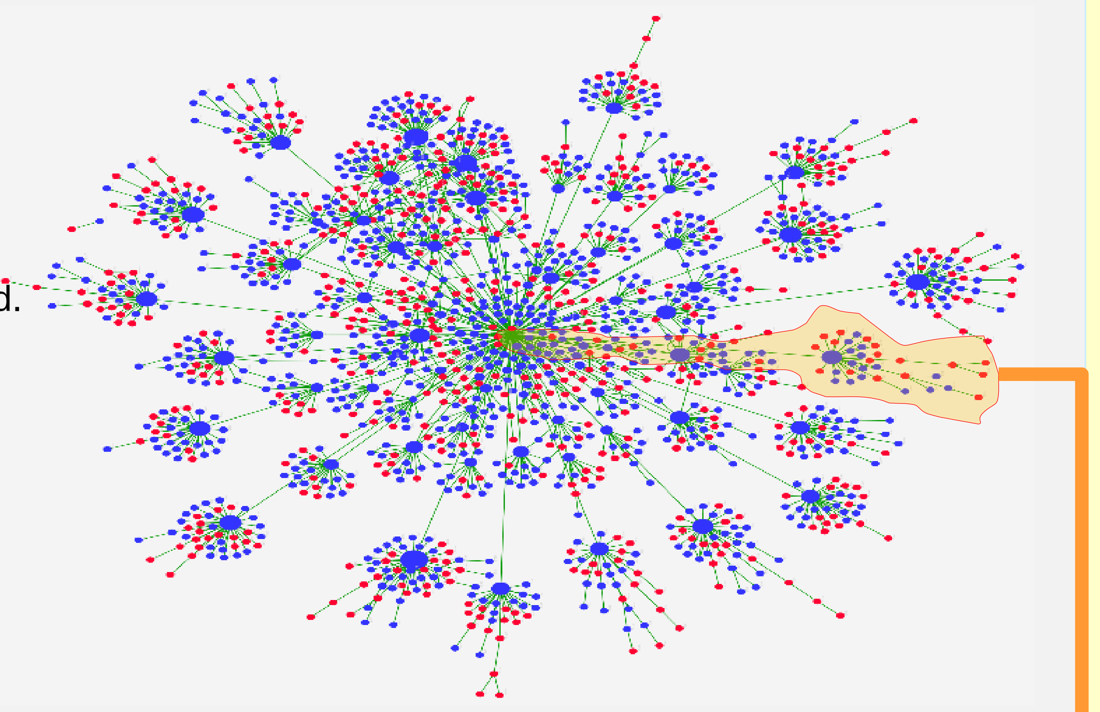
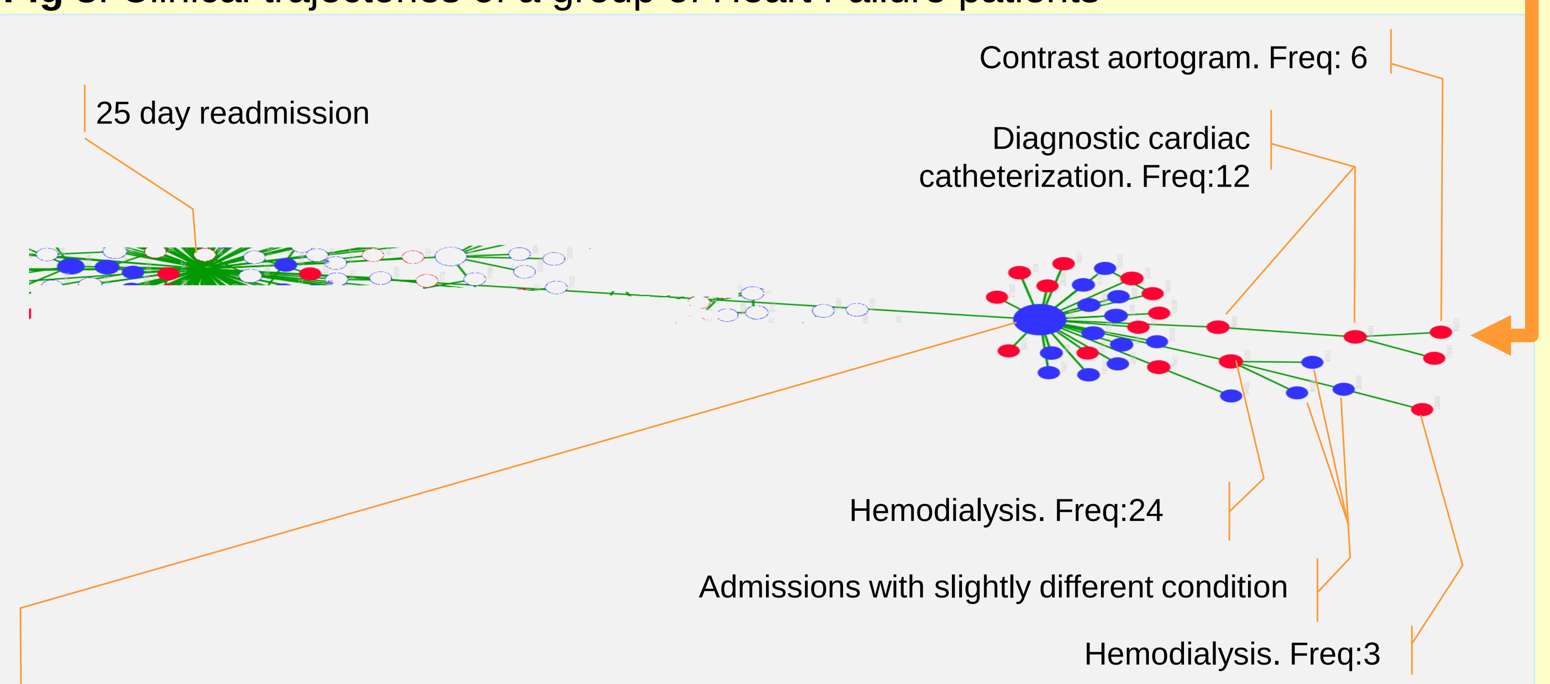


Fig 3: Clinical trajectories of a group of Heart Failure patients



Female / AGE=MidAdge,Senior/ Admission scheduled/ Length of stay=Medium / **Comorbidities**: Cardiac dysrhythmias, Heart valve disorders, Deficiency and other anemia, Chronic obstructive pulmonary disease and bronchiectasis, Pneumonia. Frequency: 324

Results

We identified distinct trajectories of events towards an outcome of readmission among patients with HF. We identified pathways that were more common in particular races/genders (Fig 1). For example longer clinical pathways that involve more admissions and procedures can be found prior to readmission, in women and black patients. Interestingly these clusters contain mostly observational visits.

In sub-graphs for the tree readmission outcomes, the network diameter and average shortest path can reveal differences in clinical pathways. The network diameter was 5 for less than 25-day readmission (Fig 2) compared to 13 for longer than 35 days readmissions, suggesting that longer than 35 days readmission involves a more diverse set of clinical pathways. Also the average shortest path of 1.69 for <25 days readmission comparing to 3.68 for >35 days readmission suggest that a more coherent group of procedures or more limited confounding parameters or both are involved in less than 25 days readmission.

A fuzzy splitting of <25 days, 25-35 and longer than >35-days readmission outcomes shows that disproportionally very few trajectories fall in 25-35 period compared to other groups. This suggests that using shorter time frames for defining readmission may yield more meaningful results when evaluating hospital performance based on readmission rates and predicting readmission.

Limitations:
We did not include emergency and ambulatory medical encounters in our investigation. Addition of these encounters may reveal additional patterns in patient trajectories. The cohort was limited to 6 years of data; longer term follow-up may enhance the results. Similarly clustering of patients with less variables can improve grouping of similar patients and augment the patterns.

Conclusions

Information on specific sequence of healthcare utilization events in heart failure patients may be useful for identifying distinct subpopulations of patients with HF. Knowledge of patient trajectories may help to improve prediction of future readmission which can be used to tailor management to the individual needs of the patient.

By incorporating the multidimensional nature of healthcare utilization data, this approach can improve our understanding of underlying temporal patterns of disease development and progression.