Identifying Time Trajectories in Risk Factors Documented in Clinical Notes and Predicting Hospitalizations and Emergency Department Visits during Home Health Care

Jiyoun Song, PhD, RN[1]; Se Hee Min, PhD, RN[1]; Sena Chae, PhD, RN[2]; Kathryn H. Bowles, PhD, RN[3,4]; Margaret V McDonald, MSW[3]; Mollie Hobensack, BSN, RN[1]; Yolanda Barrón, MS[3]; Sridevi Sridharan, MSc[3]; Anahita Davoudi, PhD[3]; Sungho Oh, PhD, RN[4]; Lauren Evans, DrPH[3]; Maxim Topaz, PhD, RN[1,3]

[1]Columbia University School of Nursing, New York, NY; [2]University of Iowa College of Nursing, Iowa City, IA; [3]Center for Home Care Policy & Research, VNS Health, New York, NY; [4]University of Pennsylvania School of Nursing, Philadelphia, PA

Background.

Home health care (HHC) provides personalized healthcare and social support services to patients in their homes. Despite efforts to reduce negative outcomes, over 20% of HHC patients are hospitalized or visit the emergency department (ED) within 60 days of starting HHC. To identify risk factors for hospitalizations or ED visits in HHC, researchers have used standardized assessments and clinical notes processed by natural language processing (NLP). However, previous studies have not examined the changes in risk factors over time which could be related to negative outcomes. Clustering techniques can aid in analyzing time trajectories of risk factors by grouping similar patterns together, enabling the identification of representative patterns in longitudinal time series data. To address this knowledge gap, this study aimed to: 1) identify a clustering in temporal patterns of risk factors documented in clinical notes using advanced data-driven analysis, including dynamic time warping and unsupervised clustering analysis, 2) examine the association between the clustering in temporal patterns and hospitalizations or ED visits.

Methods.

This study analyzed data from patients who received at least two home visits from a non-profit organization in the northeastern U.S. between 2015 and 2017. The study included 57,572 patients with 73,350 HHC episodes and 551,681 home visits. We used structured data from the Outcome and Assessment Information Set– a standardized assessment tool mandated by the Centers for Medicare and Medicaid Services– and other assessment items extracted from the electronic health record (EHR), and unstructured data from clinical notes. In the previous study, our team identified 31 risk factors (e.g., “Income,” “Circulation,” and “Medication regimen”) for hospitalizations or ED visits in HHC. Then, we developed a rule-based NLP system to automatically identify these risk factors in clinical notes. After evaluating our NLP performance (average F-score = 0.84), we applied it to our entire sample of 2.3 million HHC clinical notes, and created binary indicators for their presence during a HHC episode.

During HHC visits, risk factors documented in clinical notes generated time series data of varying lengths (i.e., different frequencies and intervals). While standard time series data analytical techniques assume time spans between observations to be constant, our study collected data at varying intervals and frequency of visits. Thus, we applied the dynamic time warping method, which aligns two-time series sequences to calculate their pattern similarity. Then, we applied unsupervised hierarchical clustering to group similar patterns together. Lastly, we used multivariate logistic regression analysis to investigate the association between clustering in temporal patterns and hospitalizations or ED visits while adjusting for sociodemographic (i.e., age, gender, ethnicity, type of insurance and living conditions such as living alone), chronic diseases, and daily living function at the episode level.

Results.

During the study period, 8,227/73,350 (11.2%) of HHC episodes resulted in a hospitalization or ED visits. Based on expert analysis of denogram – an illustration of the cluster arrangement produced by hierarchical clustering to visualize and classify taxonomic relationships, the temporal pattern of risk factors documented in clinical notes was best represented by six clusters: *Cluster 1* had no documented risk factors, *Cluster 2* had a steep decrease in risk factors, *Cluster 3* had a decrease in risk factors, *Cluster 4* had a steep increase in risk factors, *Cluster 5* initially had a decrease but rebounded and increased over time, and *Cluster 6* had consistently documented risk factors over time. The clusters showed that patients had different patterns of documented risk over time. After adjusting for sociodemographic factors, chronic diseases, and daily living function, the odds of hospitalization or ED visits were significantly higher in all clusters compared to *Cluster 1* (No Risk Factors). The adjusted odds ratios were 1.27 in *Cluster 2* (Steeply Decreased), 1.89 in *Cluster 3* (Decreased), 2.95 in *Cluster 4* (Steeply Increased), 2.47 in *Cluster 5* (Decreased and Rebound Increased), and 2.27 in *Cluster 6* (Steadily Present) (all p-value <0.001).

Conclusion.

Our study aimed to identify the temporal patterns of risk factors based on the Omaha System documented in HHC clinical notes and explore the relationship between these patterns and hospitalizations or ED visits in HHC. To achieve this, we used advanced data science methods such as dynamic time warping and unsupervised hierarchical clustering to identify hidden patterns of risk factors over time. We found that patients had different patterns of documented risk over time, and the risk of hospitalization or ED visits was higher in clusters with documented risk factors than in *Cluster 1* (No Risk Factors). Our results validate the information in clinical notes as a highly reliable risk indicator for hospitalizations or ED visits. Healthcare providers can use this information to identify at-risk patients and implement interventions to prevent adverse outcomes. By analyzing a patient's health status over time, risk factor trajectory assessment provides a more accurate reflection of their overall health status. Based on our findings, HHC could develop time-integrated cluster-based risk prediction models for early warning systems to identify at-risk patients and improve patient outcomes.