How Symptom Clusters in Elderly Stroke Patients are Longitudinally Associated with Cognitive and Physical Outcomes

Co-PIs: Suresh K. Bhavnani PhD¹, Timothy Reisttetet PhD²
Co-Is: Lorraine Evangelista RN PhD³, Ming Tzeng RN PhD³, Yong-Fang Kuo PhD⁴, Mukaila Raji MD⁵, Monique Pappadis⁶ Brian Downer PhD⁶

¹Institute for Translational Sciences, University of Texas Medical Branch, Galveston; ²School of Health Professions, University of Texas Health Sciences Center, San Antonio; ³School of Nursing, ⁴Preventative Medicine & Community Health, ⁵Division of Geriatrics, ⁶Division of Rehabilitation Sciences University of Texas Medical Branch, Galveston,

FOA: Symptom Cluster Characterization in Chronic Conditions (R01)
A. RESEARCH STRATEGY

A.1 Introduction

**Stroke: A Growing Worldwide Epidemic.** Stroke is a leading cause of disability, and the second leading cause of mortality in the developed world. This condition disproportionately affects the elderly, a growing and vulnerable segment of the US population. Depending on the area of the brain affected, two-thirds of stroke patients have long-term physical, cognitive and psychosocial impairments resulting in loss of independence, and a negative quality of life. Because stroke rehabilitation can increase the chances of recovering lost function, much attention in recent years has focused on designing targeted interventions to help elderly stroke patients prevent post-stroke conditions such as depression and a subsequent stroke, with the goal of increasing independence, and improving overall quality of life.

**Current Hurdles in Stroke Rehabilitation.** A critical input into the design of stroke rehabilitation is the symptoms reported by stroke patients. For example, stroke patients often report multiple symptoms such as feeling depressed, having pain, and encountering sleep disturbance. However, current standard-of-care in stroke rehabilitation tends to address such symptoms individually. For example, depressive symptoms are often addressed by a psychiatrist through depression treatment, and pain is addressed by a neurologist through opioids. Such an approach can lead to a proliferation of treatments, complicating symptom management, and potentially leading to unintended negative interactions.

Recent research in cancer suggests the potential advantages of treating frequently co-occurring symptoms as a whole. For example, cancer patients often report fatigue, pain, and sleep disturbance co-occurring together as a symptom cluster. Such a symptom cluster could have a common cause (e.g., inflammation caused by a tumor), or could result from a cascade (e.g., chemotherapy causing pain, resulting in sleep disturbance, causing fatigue). Identification of such symptom clusters and their underlying mechanisms has led to the design and testing of interventions that address the symptom cluster as a unit, with the potential of reducing complexities in symptom management, improving communication between provider and patient about expected symptoms, and increasing treatment compliance leading to better outcomes.

However, while symptom cluster research has made important strides in cancer, much less is known about how they manifest in other chronic conditions. For example, while a few studies in non-cancer chronic conditions such as HIV disease and heart failure have investigated symptom clusters, little is known about how they manifest in stroke, a leading cause of disability in the elderly. To address this gap, the National Institute of Nursing Research (NINR) has selected symptom cluster research in chronic conditions to be a high-priority research area, with the goal of accelerating symptom science.

**New Opportunities to Accelerate Symptom Cluster Research.** As reported by several reviews, the primary goals of symptom cluster research include identification of symptom clusters, how they change over time, how they impact critical outcomes, and what are their underlying mechanisms, with the goal of designing and testing interventions. To address these goals, symptom cluster research has primarily used unipartite methods such as hierarchical cluster analysis (HCA), principal component analysis (PCA), factor analysis (FA), and K-means. As shown in Fig. 1A, the typical output of such unipartite methods are de novo symptom clusters (e.g., Sad/Lonely, and Restless Sleep/Everything Effort), which are then used as a priori clusters to define patient subgroups, with the goal of analyzing their association to outcomes.

While such unpartite approaches have enabled important insights into symptom clusters, new bipartite machine learning methods can automatically identify and visualize both symptom clusters and patient subgroups simultaneously. For example, bipartite networks can automatically output a quantitative and visual description of patients and symptoms. The quantitative output provides the number, size, and statistical significance of biclusters of patients and symptoms. The visual output displays the quantitative information of the biclusters through a network diagram critical for comprehending and inferring the mechanisms in each bicluster. As shown in Fig. 1B, a network consists of nodes (circles and triangles) and edges (lines connecting the circles to triangles), which represent the association between patients and their symptoms.

A key advantage of a bipartite network visualization is that besides showing the symptom clusters and patient subgroups, it also reveals the relationships within and across patient subgroups. For example, the visualization in Fig. 1B shows that
patients in the left subgroup have a more homogenous profile compared to patients in the right subgroup. Furthermore, three patients in the right subgroup share a symptom that occurs most frequently in the left subgroup (shown by the darker edges between the subgroups), whereas none of the patients in the left subgroup share a symptom frequently occurring in the right subgroup. Such relationships could enable researchers to infer for example that the symptom-related mechanisms and the interventions in the right subgroup involve complex interactions, and which could overlap with the left subgroup. Furthermore, the same approach could be used to analyze how the symptom biclusters change over time, and how they predict outcomes such as physical and cognitive outcomes.

Co-PI Bhavnani has pioneered the use of bipartite networks to represent such patient-characteristic networks and used it to identify and interpret patient subgroups in a wide range of clinical and molecular datasets. More recently, his team has demonstrated its use in PCOR to analyze readmission data. These projects have shown that patient-characteristic networks enable clinician and researcher stakeholders to (1) comprehend the nature of patient subgroups, (2) infer disease processes in each patient subgroup, and (3) design potential interventions targeted to those patient subgroups. Furthermore, because patients within a subgroup are relatively more homogeneous compared to patients outside that subgroup, regression models designed for each patient subgroup (such as through stratified regression models) have the potential for achieving higher prediction accuracy compared to models that do not take subgroups into consideration.

In this project, we will use bipartite networks to automatically identify symptom clusters and patient subgroups, how they change over time, and their longitudinal association to outcomes critical in stroke. Furthermore, we will compare them to current unipartite approaches to understand their strengths and limitations. These analyses will be conducted using data from the Health and Retirement Study (HRS), which is currently the largest longitudinal dataset of elderly patients.

A.2 Significance (To Be Expanded)

- First cross-sectional and longitudinal symptom cluster study in a representative sample of elderly stroke patients in the US. Will provide recommendations for designing strategies for stroke rehabilitation.
- First study that will test significance, replication, and longitudinal trends using the same dataset
- First study to compare commonly used methods (k-means, hierarchical cluster analysis, Principal Component Analysis, Factor Analysis) to analyze symptom clusters (one of the reviews recommends using multiple methods). Will provide recommendations for which methods to use for answering different research questions.

A.3 Innovation (To Be Expanded)

- First symptom cluster study to use bipartite methods (representing both patients and symptoms simultaneously) to analyze symptom clusters (except my study https://www.ncbi.nlm.nih.gov/pubmed/21085743)
- First symptom cluster study to use a multidisciplinary team-science approach (involving geriatrics, stroke rehabilitation, nursing, gerontology, biostatistics and data mining stakeholders) to interpret symptom clusters through qualitative (focus groups), and quantitative methods.

A.4 Background

Symptoms Clusters in Cancer and Other Chronic Diseases

- Frameworks and Theories
  - Theory of Unpleasant Symptoms: Causes, Characteristics, and Outcomes of symptom clusters
- Symptom Clusters in Cancer
  - Frequent symptoms clusters in Cancer (e.g., fatigue, etc. caused by chemotherapy)
  - Causes of symptom clusters
  - Outcomes of symptom clusters
  - Interventions tested based on symptom clusters
- Common clusters in COPD, and Heart Failure
  - Frequent symptoms
  - Causes
  - Outcomes
  - Interventions tested based on symptom clusters
- Clusters common across chronic diseases and clusters unique to each disease
- Symptoms Reported in Stroke and Critical Outcomes
Methods used to Analyze Symptom Clusters

- *De Novo Methods* (Unipartite Methods)
  - K-means
  - Hierarchical Clustering
  - Principal Component Analysis
  - Factor Analysis
  - Limited research in comparing them on the same dataset
    - Sometimes they give similar results, sometimes different

- *A Priori Methods*
Bipartite Visual Analytical Methods Useful to Analyze Symptom Clusters

- Introduction to Bipartite Networks
- Quantitative methods to identify and analyze biclusters
- Modularity and Significance
- Replication
- Longitudinal changes
- Layout algorithms: Fruchterman Reingold, ExplodeLayout
### A.5 Approach

**Overview of Analytical Approach**

<table>
<thead>
<tr>
<th>Aims</th>
<th>Data</th>
<th>Quantitative Analysis</th>
<th>Qualitative Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Aim-1: Identify Symptom Clusters in Elderly Stroke Patients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Identify symptoms in stroke patients commonly associated with cognitive and physical outcomes</td>
<td>HRS Wave 3 (2010)</td>
<td>Cross-sectional Measures: # of clusters, degree of clusteredness (Modularity), significance (permutation test), replication (Rand Index), replication significance (permutation test)</td>
<td>Focus Group: Solicit symptoms common in stroke patients</td>
</tr>
<tr>
<td></td>
<td>Patients: 1187 with complete symptom data</td>
<td>Visualization: Fruchterman Reingold, ExplodeLayout</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Symptoms: 15 symptoms (8 depression, 2 memory 1 pain, 2 vision, 1 hearing, 1 sleep)</td>
<td>Cross-sectional Measures for Comparison: # of clusters, degree of clusteredness, significance, replication, and replication significance.</td>
<td>Focus Group: Rank symptom clusters across (blinded) methods by clinical significance, infer mechanisms resulting in outcomes, and rehab implications</td>
</tr>
<tr>
<td>B. Identify symptom clusters and patient subgroups using bipartite networks</td>
<td></td>
<td>Hypothesis: Bipartite method will identify clusters that have higher significance in replication compared to other methods.</td>
<td></td>
</tr>
<tr>
<td>C. Compare symptom clusters between bipartite and unipartite methods</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Aim-2: Analyze Association between Symptom Clusters and Outcomes** | | | |
| A. Predict association between patient subgroups & outcomes using bipartite network analysis | HRS Wave 3 (2010) | Measures: Predictive accuracy (discrimination, calibration) of regression models to predict each outcome based on symptom clusters identified through bipartite network analysis | Focus Group: Rank symptom clusters with significant outcomes from different methods (blinded by method) by clinical significance; infer potential mechanisms in each subgroup resulting in outcomes |
| | Patient Subgroups: All subgroups in each method identified through bipartite networks in Aim-1 | | |
| | Outcomes: 16 outcome measures (6 ADL, 5 IADL, 4 cognitive, 1 diagnoses post-stroke) | Comparison Measures: Rank methods based on accuracy of predicting outcomes by each symptom cluster identified by bipartite and unipartite methods. | |
| | Demographics: Age, gender, race/ethnicity | Hypothesis: Subgroups identified through bipartite method will have higher accuracy in predicting outcomes compared to other methods. | |
| | Treatment Status: Medications, Rehab | | |
| | Time Since Last Stroke | | |

| **Aim-3: Develop Longitudinal Models to Characterize and Predict Outcomes based on Symptom Clusters** | | | |
| A. Analyze symptom clusters over time using bipartite method | HRS Waves 3-5 (2010, 2012, 2014) | Longitudinal Measures: Longitudinal plots showing change in # of clusters, degree of clusteredness and its significance, and significant differences in Rand Index. | Focus Group: Rank longitudinal models from different methods (blinded by method) by clinical significance; and suggest modifications to variables to improve predictive accuracy |
| | Patients: 765 with all 15 symptoms across all three waves | Visualization: Network layouts highlighting cluster splits, merges, deletions and additions over 3 waves. | |
| | Outcomes: 17 outcome measures (6 ADL, 5 IADL, 27 cognitive, 1 mortality, 1 diagnoses post-stroke) | | |
| B. Predict outcomes based on symptom clusters using bipartite method | | Longitudinal Prediction Measures: Predictive accuracy (discrimination, calibration) of regression models to predict each outcome based on symptom clusters at baseline; Covariates: time, time since | |
| | | | |
Data

- Overall Goal and Description of HRS (with a focus on elderly population who are close to or at retirement, and whom the researchers wish to be independent).
- Waves and Years (Table 1 showing number of patients in each wave, percentage of those that were in the previous year, and percentage of replenished sample).

Sample

- Cross-sectional Analysis
  - Independent Variables (Symptoms): 15 symptoms related to depression, memory, pain, vision, hearing and sleep (Table 2 showing all variables divided into Independent Variables [Symptoms], Dependent Variables [Outcomes], and Covariates [Demographics, Variables about Stroke, Medication, Rehab status])
  - Dependent Variables (Outcomes, and Demographics). Distribution shows 30% had ADL problems, which matches literature
  - 1187 stroke patients from 2010 with complete symptom and outcome data
- Longitudinal Analysis
  - Independent Variables (same as above)
  - Dependent Variables (Outcomes, Demographics, Mortality, and Post-Stroke Diagnoses, Medication, and Rehab status)
  - Description of distribution related to all symptom, ADL, IADL, and cognitive variables (we can drop these later but good to have them here).

Team-Science Approach

- Multidisciplinary teams focusing on data management, quantitative analysis, and qualitative analysis.
- Visual Analytics used as a boundary object to span disciplinary boundaries for qualitative interpretation of clusters, mechanisms, and targeted interventions

Preliminary Results

- Bipartite network of depression symptoms in stroke patients from HRS.
Specific Aims

Aim-1: Identify Symptom Clusters in Stroke Patients. This aim will be achieved by evaluating a range of conventional and advanced visual analytical clustering approaches to identify symptom clusters based on clinical meaningfulness, in addition to statistical significance and replicability.

A. Identify symptoms in stroke patients commonly associated with cognitive and physical outcomes
B. Analyze symptom clusters and patient subgroups using bipartite method
C. Compare symptom clusters between bipartite and unipartite methods

Contingency plan if clusters are not significant

Aim-2: Analyze the Association between Symptom Clusters and Cognitive and Physical Outcomes. This aim will be achieved by comparing how symptom clusters identified from each of the above methods, are associated with cognitive and physical function.

A. Predict association between patient subgroups & outcomes using bipartite network analysis
B. Compare prediction of patient subgroup outcomes across bipartite and unipartite methods

Aim-3: Develop Longitudinal Models to Predict how Symptom Clusters are Associated with Cognitive and Physical Outcomes. This aim will be achieved by developing predictive models that take into consideration the change of symptom clusters over time, and their association with cognitive and physical decline.

A. Analyze symptom clusters over time using bipartite method
B. Predict outcomes based on symptom clusters using bipartite method
C. Compare longitudinal models across unipartite and bipartite methods
Limitations

- Limitations in HRS data
- Focus on one bipartite method

Timeline (4 years)