# Advancing Symptom Science through Bipartite Networks: Application to Precision Stroke Rehabilitation

**Brian Downer PhD[1], Irma Alvarado PhD[2], Monique Pappadis PhD[1], Weibin Zhang PhD[3], Timothy Reistetter PhD[4], Lorraine Evangelista PhD[2], Yong-Fang Kuo PhD[3], Suresh K. Bhavnani PhD[3]**

**[1]School of Health Professionals, [2]School of Nursing, [3]Preventive Medicine and Population Health, University of Texas Medical Branch, Galveston, TX; [4]School of Health Professionals, University of Texas Health, San Antonio, TX**

## Introduction

Stroke is a leading cause of disability and mortality in the developed world,[1] which disproportionally affects older adults, a growing segment of the US population.[2] Unfortunately, nearly a third of stroke survivors experience post-stroke depression,[3] which can significantly impact rehabilitation and recovery. Because stroke rehabilitation can help individuals regain independence and improve quality of life,[4] there is an urgent need to design precision rehabilitation approaches to prevent or manage post-stroke conditions such as depression.

A critical input into the design of stroke rehabilitation is patient-reported symptoms, a primary focus of the emerging field of symptom science. This field has revealed the potential advantages of identifying and designing interventions to target groups of frequently co-occurring symptoms referred to as *symptom clusters.*[5] For example, cancer patients often report the symptom cluster of fatigue, pain, and sleep disturbance, which 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).[5] Such symptom cluster research can therefore enable a holistic understanding of patients, leading to the design of targeted interventions.

While analysis of symptom clusters has revealed important insights in cancer, few studies have focused on conditions such as stroke. Furthermore, such research has mainly used unipartite methods (e.g., k-means clustering, hierarchical clustering, and factor analysis) to identify how symptoms co-occur. However, these methods ignore how those frequently co-occurring symptoms form patient subgroups, critical for analyzing their association with outcomes. Here we demonstrate how bipartite networks[6] can be used to (1) analyze how depression symptoms co-occur to form stroke patient subgroups, enabling analysis of their association to physical outcomes, and (2) interpret the results to infer the processes underlying the clusters, and to design targeted interventions.

## Method

***Data*.** We accessed stroke patient data from the Health and Retirement Study (HRS),[7] which is a longitudinal, nationally representative survey of nearly 20,000 U.S. adults aged 50 and older. The data consisted of all HRS participants with self-reported stroke (n=1187) from 2010, of which 798 had at least one of 8 depressive symptoms (*not happy*, *did not enjoy life*, *felt depressed*, *felt sad*, *felt lonely*, *couldn't get going*, *everything effort*, and *restless sleep*) derived from the Center for Epidemiologic Studies Depression Scale (CES-D); covariates included age, gender, race/ethnicity (white/Caucasian, minority), education, and self-reported health conditions common in stroke survivors (e.g., Diabetes); outcomes included having limitations in one or more activities of daily living (ADLs), and in one or more instrumental ADLs (IADLs).

***Analysis*. (1) Bicluster Identification.** (a) Represented the data as a bipartite network (Fig. 1), where nodes (circles and triangles) represented either patients or symptoms, and the edges (lines) connecting the nodes represented the presence or absence of a symptom, (b) used bicluster modularity[6] to identify the number and boundaries of patient-symptom biclusters and the degree of biclustering (Q), and (c) measured the significance of Q by comparing it to a distribution of Q generated from 1000 random permutations of the network; **(2) Outcome Association.** Used multivariable logistic regression (with age, gender, race, education, and self-reported health conditions as covariates) to measure the OR of the patients in each bicluster to ≥1 ADLs and to ≥1 IADLs, compared to (a) patients that had no depression symptoms (n=389), and (b) patients in the rest of the biclusters; **(3) Visualization.** (a) Used *Kamada-Kawai*[6] to layout the network, and *ExplodeLayout*[8] to separate the identified biclusters for improving their interpretability; and **(4) Interpretation.** Conducted semi-structured interviews with two clinicians that had stroke rehabilitation experience to interpret the results, and used Grounded Theory to analyze the interview transcripts.

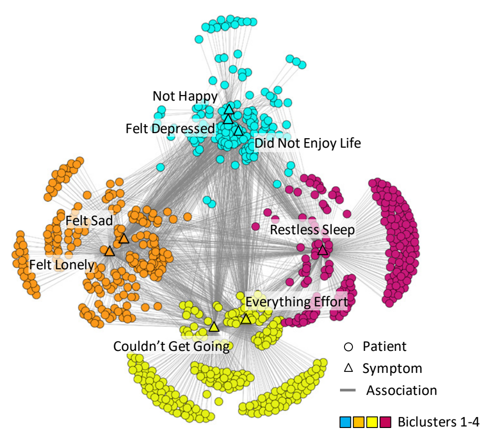


**Fig. 1.** Bipartite network showing how 8 depressive symptoms co-occurred across 4 subgroups of stroke patients.

## Results

***Symptom Co-occurrence*.** As shown in Fig. 1, the bipartite network analysis identified four biclusters consisting of subgroups of stroke patients and their most frequently co-occurring symptoms. This biclustering had significant modularity (Q=0.26, z=3.03, p<.01, two-tailed). The visualization also revealed that within each bicluster, there were patients on the periphery with only one symptom, and patients on the inner part with many symptoms. Furthermore, the proportion of patients with more than one symptom in each bicluster was significantly different ($\chi^2$ (3, N=798) = 106.2, p<.05) across the biclusters (Yellow=50.9%, Pink=51.0%, Orange=81.7%, Blue=89.1%). Therefore, while the bipartite network analysis revealed symptom heterogeneity *across* stroke patients, it also revealed a second level of heterogeneity *within* each bicluster, revealing the complexity of how symptoms co-occur within and across patient subgroups. Such insights are difficult to derive from unipartite clustering methods currently being used because they do not cluster patients and symptoms simultaneously.

***Association to ADLs and IADLs*****.** Compared to the patients with no symptoms, patients in each of the 4 clusters had a significantly higher association with having one or more ADL limitations. However, they differed in odds for ADL ranked as follows: (1) Blue Cluster (ADL OR=4.08, 95% CI=2.66-6.29), (2) Yellow Cluster (ADL OR=3.67, 95% CI=2.49-5.45), (3) Orange Cluster (ADL OR=2.99, 95% CI=2.01-4.47), and (4) Pink Cluster (ADL OR=2.62, 95% CI=1.74-3.96), with an identical ranking order, and similar significance for IADL. Furthermore, the Blue Cluster had significantly higher ORs for ≥1 IADLs (OR=1.78, 95% CI=1.23-2.60), and the Pink Cluster had significantly lower ORs for ≥1 IADLs (OR=0.56, 95% CI=0.38-0.81), compared to the rest of the patients with ≥1 symptoms.

***Interpretation of Results*****.** (a) *Overall Symptom Co-occurrence.* The two participants agreed that the 4 symptom clusters were clinically meaningful (Blue Cluster implied general discontent symptoms, Yellow Cluster implied apathy symptoms, Orange Cluster implied psychosocial symptoms, and Pink Cluster implied sleep disturbance). (b) *Potential Processes Resulting in IADLs.* Because patients in the blue cluster with general discontent symptoms had the highest risk for ≥1 IADL limitations compared to the controls, in addition to the significantly higher risk compared to rest of the patients with ≥1 symptoms, it implied that those symptoms interfere in a patient's ability or desire to engage in activities such as household tasks, managing finances, and social integration. Furthermore, because a lack of engagement in such instrumental daily activities often has a negative impact on the initial symptoms, further exacerbating their increased risk for ≥1 IADL limitations. In contrast, because the pink bicluster had almost half of the patients with only restless sleep, this group of patients could have a relatively lower risk of limitations in IADLs. Future analysis should compare the OR between those with only sleep disturbance, to those with sleep disturbance and other symptoms. (c) *Implications for Intervention Design.* The overall results suggest that clinicians and researchers should provide patients that have general discontent symptoms, with therapies to compensate or create new values for activities, and to increase participation in IADLs. In contrast, patients with sleep restlessness could benefit from targeted sleep hygiene strategies to prevent them from developing more risky symptoms such as those related to general discontent, which could require more involved stroke rehabilitation.

## Current and Future Research

The results demonstrated the feasibility of using bipartite network analysis and visualizations to (1) identify the symptom-patient biclusters, and (2) interpret the results to infer the processes underlying the biclusters, and to design potential interventions targeted to the respective patient subgroups. However, the bipartite network visualization analysis revealed a second level of heterogeneity within and across biclusters that requires further investigation, such as which combinations of intra- and inter-cluster symptoms are significantly associated with limitations in ADLs and IADLs. Furthermore, as only two clinicians interpreted the results reported here, we are using our semi-structured interview protocol to solicit clinical interpretations from other clinicians involved in stroke rehabilitation, including physicians, nurses, rehabilitation professionals, and behavioral health professionals. Finally, our future analyses will include a broader range of symptoms and comorbidities relevant to stroke, which should enable a deeper understanding of heterogeneities pertinent to the design of precision rehabilitation in stroke, in addition to the application of our approach to a wide range of other conditions.

**Acknowledgments.** Funded by the School of Nursing and School of Health Professionals, UTMB, and the UTMB CTSA (UL1 TR001439).

## References


1. World Stroke Organization. Facts and Figures about Stroke. 2019; https://www.world-stroke.org/component/content/article/16-forpatients/84-facts-and-figures-about-stroke.
2. Centers for Disease Control and Prevention: Stroke Facts. 2017; https://www.cdc.gov/stroke/facts.htm. Accessed June, 4, 2019.
3. Hackett ML, Pickles K. Part I: frequency of depression after stroke: an updated systematic review and meta-analysis of observational studies. *International journal of stroke : official journal of the International Stroke Society.* 2014;9(8):1017-1025.
4. Winstein CJ, Stein J, Arena R, et al. Guidelines for Adult Stroke Rehabilitation and Recovery: A Guideline for Healthcare Professionals From the American Heart Association/American Stroke Association. *Stroke.* 2016;47(6):e98-e169.
5. Fan G, Filipczak L, Chow E. Symptom clusters in cancer patients: a review of the literature. *Current oncology (Toronto, Ont).* 2007;14(5):173-179.
6. Newman MEJ. *Networks: An Introduction.* Oxford, United Kingdom: Oxford University Press; 2010.
7. Sonnega A, Faul JD, Ofstedal MB, Langa KM, Phillips JW, Weir DR. Cohort Profile: the Health and Retirement Study (HRS). *Int J Epidemiol.* 2014;43(2):576-585.
8. Bhavnani SK, Chen T, Ayyaswamy A, et al. Enabling Comprehension of Patient Subgroups and Characteristics in Large Bipartite Networks: Implications for Precision Medicine. *Proceedings of AMIA Joint Summits on Translational Science.* 2017:21-29.