

## Clinical phenotypes of patients with acute stroke: a secondary analysis

MOHAMMADREZA HAJIESMAEILI <sup>\*1</sup>, NAVID NOORAEI <sup>\*2</sup>, NASSER MALEKPOUR ALAMDARI <sup>2</sup>,  
BEHRUZ FARZANEGAN BIDGOLI <sup>3</sup> SANAZ ZARGAR BALAYE JAME <sup>4</sup>,  
NADER MARKAZI MOGHADDAM <sup>2,4</sup>, MOHAMMAD FATHI <sup>2</sup>

<sup>1</sup> Critical Care Quality Improvement Research Center, Loghman Hakim Hospital,  
Shahid Beheshti University of Medical Sciences, Tehran, Iran

<sup>2</sup> Critical Care Quality Improvement Research Center, Shahid Modarres Hospital,  
Shahid Beheshti University of Medical Sciences, Tehran, Iran

<sup>3</sup> Critical Care Quality Improvement Research Center, Dr. Masih Daneshvari Hospital,  
Shahid Beheshti University of Medical Sciences, Tehran, Iran

<sup>4</sup> Department of Health Management and Economics, Faculty of Medicine,  
Aja University of Medical Sciences, Tehran, Iran

\*These authors are co-first authors and had equal contribution

### Abstract

**Introduction:** Stroke is a leading cause of mortality worldwide and a major cause of disability having a high burden on patients, society, and caregiving systems. This study was conducted to investigate the presence of clusters of in-hospital patients with acute stroke based on demographic and clinical data. Cluster analysis reveals patterns in patient characteristics without requiring knowledge of a predefined patient category or assumptions about likely groupings within the data.

**Methods:** We performed a secondary analysis of open-access anonymized data from patients with acute stroke admitted to a hospital between December 2019 to June 2021. In total, 216 patients (78; 36.1% men) were included in the analytical dataset with a mean (SD) age of 60.3 (14.4). Many demographic and clinical features were included in the analysis and the Barthel Index on discharge was used for comparing the functional recovery of the identified clusters.

**Results:** Hierarchical clustering based on the principal components identified two clusters of 109 and 107 patients. The clusters were different in the Barthel Index scores on discharge with the mean (SD) of 39.3 (29.3) versus 62.6 (29.4);  $t(213.87) = -5.818$ ,  $P < 0.001$ , Cohen's  $d(95\%CI) = -0.80 (-1.07, -0.52)$ . A logistic model showed that age, systolic blood pressure, pulse rate, D-dimer blood level, low-density lipoprotein, hemoglobin, creatinine concentration, the National Institute of Health Stroke Scale value, and the Barthel Index scores on admission were significant predictors of cluster profiles (all  $P \leq 0.029$ ).

**Conclusion:** There are two clusters in hospitalized patients with acute stroke with significantly different functional recovery. This allows prognostic grouping of hospitalized acute stroke patients for prioritization of care or resource allocation. The clusters can be recognized using easily measured demographic and clinical features.

**Keywords:** Acute Stroke; Phenotype; Functional Recovery; Clustering; Barthel Index.

### What is new? What is important?

- There were two clusters in patients with acute stroke.
- Cluster 1 had an ischemic stroke and higher LDL, D-dimer, and NIHSS.
- Cluster 2 had a non-ischemic stroke and a higher prevalence of DM.
- Cluster 2 patients were older, and had a lower pulse rate, and higher SBP, Hb, and Cr.
- Functional recovery on discharge was higher in cluster 2.

### INTRODUCTION

Cerebrovascular accidents are the second leading cause of death worldwide[1]. As a major cause of disability, stroke has a high burden on

patients, society, and caregiving systems as well [1, 2]. A variety of risk factors are associated with higher incidence, mortality, and disability following stroke [3]. Predicting the outcome of acute stroke has been focused on in previous studies [4].

Commonly, a set of demographic and clinical risk factors has been investigated to risk-stratify patients using regression and classification techniques [5, 6]. Researchers tried to increase the predictive ability of clinical models by applying statistical methods and machine learning algorithms [4–6].

A recent systematic review enumerated methods used for the prediction of functional recovery and prognosis after stroke [6]. Linear and logistic regressions were the most commonly used techniques for data analysis. Other less frequently applied methods included: random forest, k-nearest neighbors, neural network, parallel cascade identification, and support vector machine [6]. However, the results of the unsupervised analysis of clinical stroke data are still limited. Unsupervised analysis recognizes patterns within data sets without the need to predict predefined target labels. Unsupervised machine learning methods are primarily used to discover phenotypes of patients in a variety of diseases [7, 8]. Cluster analysis is an unsupervised method for segmenting heterogeneous data into distinct subsets (i.e., clusters) based on inherent patterns or similarities, thereby facilitating the identification of latent structures within complex medical data. It has the potential to elucidate relations among variables, thereby providing insight into the intricacies of disease entities [7]. Studies on the cluster analysis of stroke patients are more focused on specific subjects such as linguistic deficits, botulinum toxin therapy, sedentary behavior, gait, and sleep apnea in patients with stroke [9–13]. In a recent nationwide Chinese study, 92 biomarkers (including imaging data) were used to cluster patients with stroke into four phenotypes [14].

Meanwhile, we sought to identify a smaller number of clusters using routine and easy-to-measure biomarkers. This study was conducted to investigate the presence of clusters of in-hospital patients with acute stroke based on demographic and clinical data. Our hypothesis was that patient clusters would differ in the outcome of stroke. This approach would help to recognize groups of patients with different prognoses based on practical rules of thumb.

## MATERIAL AND METHODS

### DESIGN AND SETTING

Qu *et al.* studied 216 patients with acute stroke who were admitted to the Shenzhen Second People's Hospital between December 2019 to

June 2021 [5]. They compared the accuracies of generalized regression neural networks and logistic regression models to identify an optimal model for predicting acute stroke outcomes. They also studied the association of biomarkers with an outcome variable for acute stroke. The outcome variable was the Barthel Index (BI) on discharge. It was reported that the generalized regression neural network showed a better performance than the logistic regression [5]. The primary study was approved by the Ethics Committee of Shenzhen Second People's Hospital. Our study was a secondary analysis of patients' data from Qu's study using an unsupervised learning method. Patients were clustered and then, the resulting subgroups were compared in their prognoses using BI as the outcome variable. Our study was carried out following the Declaration of Helsinki. We did not perform measurements on participants. Also, we did not have access to any patients' identifying information.

### ETHICAL CONSIDERATIONS

Our study was carried out following the Declaration of Helsinki. We did not carry out measurements on participants nor did we have access to identifying information. This is a secondary analysis of open access anonymized data. The primary study was approved by the Ethics Committee of Shenzhen Second People's Hospital.

### DATA AVAILABILITY

The dataset is publically accessible at <https://doi.org/10.1371/journal.pone.0267747.s001> (Published: May 11, 2022; Last access date: December 22, 2022).

### CANDIDATE VARIABLES AND MEASUREMENTS

Patients with acute stroke were included if they were  $\geq 18$  years; had an ischemic (infarction of the central nervous system) or hemorrhagic (spontaneous, non-traumatic bleeding) clinical and imaging diagnosis of stroke; and the onset was within 90 days. Patients with missing BI scores at discharge did not enter the study. The analytic dataset included data on patients' age (years), sex, body mass index (BMI; kg/m<sup>2</sup>), current smoking, alcohol use, systolic (SBP) and diastolic (DBP) blood pressures in mmHg, and previous history of blood hypertension (HTN), pulse rate (beat/min), history of ischemic stroke, coronary artery disease (CAD), diabetes mellitus (DM), atrial fibrillation

(AF). Biochemical assessments included serum D-dimer ( $\mu\text{g/mL}$ ), hemoglobin (Hb;  $\text{g/L}$ ), creatinine (Cr;  $\mu\text{mol/L}$ ), and albumin (Alb;  $\text{g/L}$ ), serum sodium (Na;  $\text{mmol/L}$ ), potassium (K;  $\text{mmol/L}$ ), and lipid profile measurements encompassing total (TC), low-density (LDL), and high-density lipoprotein (HDL) cholesterol, and triglyceride (TG) in  $\text{mmol/L}$ . Also, the National Institute of Health Stroke Scale (NIHSS) for the severity of initial neurological damage was recorded for each patient [15, 16]. The NIHSS ranges from 0 to 42 with higher scores representing more severe stroke. The scale assesses the level of consciousness, gaze, visual fields, facial palsy, extremity strength, ataxia, sensory function, language, dysarthria, extinction, and inattention. Severe neurological impairment at the initial stage has negative effects on functional recovery from stroke [17]. Functional recovery is suggested to be monitored in patients with stroke [17]. In Qu's study, the BI was used for evaluating functional recovery [5]. One professional physiotherapist measured the BI at baseline and on discharge [5]. The BI was initially developed to assess daily activities in patients with neuromuscular disorders [18]. However, it is commonly used as a reliable and valid assessment for assessing physical functioning in critically ill patients [18–20]. The index assesses feeding, dressing, mobility (walking on a level surface and ascending/descending stairs or an incline), and personal hygiene (bathing, grooming, toileting, and control of bodily functions) [21]. The BI is categorized into 0–39 (total functional dependence), 40–59 (partially dependent), 60–84 (independent), and 85–100 (total functional independence) [21]. The satisfactory functional outcome has been suggested as  $\text{BI} \geq 60$  [21].

#### DATA ANALYSES

Cluster analysis has been used extensively for analyzing clinical data [22]. We carried out a cluster analysis of the data using the measured risk factors of stroke. Data were assessed for the presence of duplicate rows and missing data. Also, the dataset was searched for the presence of correlated variables and if the absolute value of the correlation coefficients were  $> 0.5$ , one variable from each pair was excluded. The clustering tendency was confirmed with the Hopkins statistic. The data were normalized and hierarchically clustered using factor analysis of mixed (continuous and categorical) data. We profiled the resulting clusters by selecting the most important predictors

of cluster membership through a random forest algorithm. Then, a stepwise binary logistic regression with 10-fold cross-validation was carried out and adjusted odds ratios, model  $\chi^2$ , Nagelkerke  $R^2$ , Akaike information criterion, Hosmer-Lemeshow test result, Brier's calibration score, and model performance metrics were reported. Results are presented as mean (SD) for continuous variables, and as absolute numbers (%) for categorical data. The means of the continuous variables were compared using paired and independent t-tests. A  $\chi^2$  test was used for testing differences among the study clusters for categorical variables. The level of significance was set at two-tailed  $\alpha = 0.05$ . The analysis and visualization were carried out using R software version 4.0.2. for windows (R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>).

## RESULTS

### SAMPLE

In total, 216 patients were included in the analytical dataset. There were no missing data or duplicate rows. Table 1 shows the characteristics of the study sample at the baseline. The mean (SD) of BI on discharge was 50.8 (31.5). There was a statistically significant difference between discharge and baseline BI, mean difference (95%CI) = 10.8 (7.9, 13.7), paired sample t-test (215 degrees of freedom) = 7.23,  $P < 0.001$ , Cohen's  $d$  (95%CI) = 0.50 (0.35, 0.63). This suggests a functional improvement on discharge compared with the baseline. Of the study candidate variables, current smoking and alcohol consumption, SBP and DBP, and LDL and TC were the highly correlated pairs of features with Pearson's coefficients of 0.51, 0.55, and 0.84, respectively. We excluded alcohol, DBP, and TC from further analysis.

### CLUSTERING

The outcome variable (BI on discharge) was excluded and the data were scaled. The Hopkins statistic was 0.764 indicating a high cluster tendency of patients' data. We clustered patients based on a factor analysis of mixed data. Figure 1.A illustrates the scatter diagram of the first and second principal components. Hierarchical clustering based on the principal components best suggested the presence of two clusters (Figure 1.B). Clusters 1 and 2 included 109 and 107 patients, respectively. The

scatter diagram of the clustered patients (Figure 1.C) shows that one of the clusters is more compact than the other. There was a highly significant difference with a large effect size between the two clusters in BI scores on discharge, with a mean (SD) of 39.3 (29.3) versus 62.6 (29.4) for clusters 1 and 2, respectively;  $t(213.87) = -5.818$ ,  $P < 0.001$ ,

Cohen's  $d$  (95%CI) =  $-0.80$  ( $-1.07$ ,  $-0.52$ ). Also, the satisfactory functional outcome (BI  $\geq 60$ ) was significantly different between the two clusters, 25 (22.9%) versus 58 (54.2%),  $\chi^2(1) = 21.012$ ,  $P < 0.001$ , unadjusted odds ratio (95%CI) =  $0.25$  ( $0.14$ ,  $0.45$ ). This showed that patients in cluster 2 experienced better functional recovery on discharge.

Table 1

Patients' characteristics. Figures represent mean (SD) for continuous, and number (%) for categorical variables

| Characteristic           | (N = 216)    |
|--------------------------|--------------|
| Age (year)               | 60.3 (14.4)  |
| BMI (kg/m <sup>2</sup> ) | 24.0 (4.7)   |
| SBP (mmHg)               | 130.0 (18.2) |
| DBP (mmHg)               | 79.4 (10.2)  |
| Pulse (beat/min)         | 78.5 (11.4)  |
| D-dimer ( $\mu$ g/mL)    | 1.6 (3.3)    |
| TC (mmol/L)              | 3.6 (0.9)    |
| TG (mmol/L)              | 1.3 (0.6)    |
| HDL (mmol/L)             | 1.3 (4.2)    |
| LDL (mmol/L)             | 2.2 (0.9)    |
| Hb (g/L)                 | 122.8 (19.5) |
| Na (mmol/L)              | 139.1 (4.9)  |
| K (mmol/L)               | 4.0 (0.4)    |
| Cr ( $\mu$ mol/L)        | 64.2 (26.1)  |
| Alb (g/L)                | 39.1 (4.7)   |
| NIHSS (score)            | 8.2 (9.2)    |
| BI on Admission(score)   | 40.0 (29.5)  |
| Men (%)                  | 78 (36.1)    |
| Smoker (%)               | 73 (33.8)    |
| Alcohol (%)              | 75 (34.7)    |
| Ischemic (%)             | 100 (46.3)   |
| Hemorrhagic (%)          | 116 (53.7)   |
| CAD (%)                  | 41 (19.0)    |
| HTN (%)                  | 172 (79.6)   |
| DM (%)                   | 69 (31.9)    |
| AF (%)                   | 17 (7.9)     |

BMI, Body Mass Index; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; TC, Total Cholesterol; TG, Triglyceride; HDL, High-Density Lipoprotein; LDL, Low-Density Lipoprotein; Hb, Hemoglobin; Na, sodium; K, potassium; Cr, Creatinine; Alb, Albumin; NIHSS, National Institute of Health Stroke Scale; BI, Barthel Index; CAD, Coronary Artery Disease; HTN, blood Hypertension; DM, Diabetes Mellitus; AF, Atrial Flutter

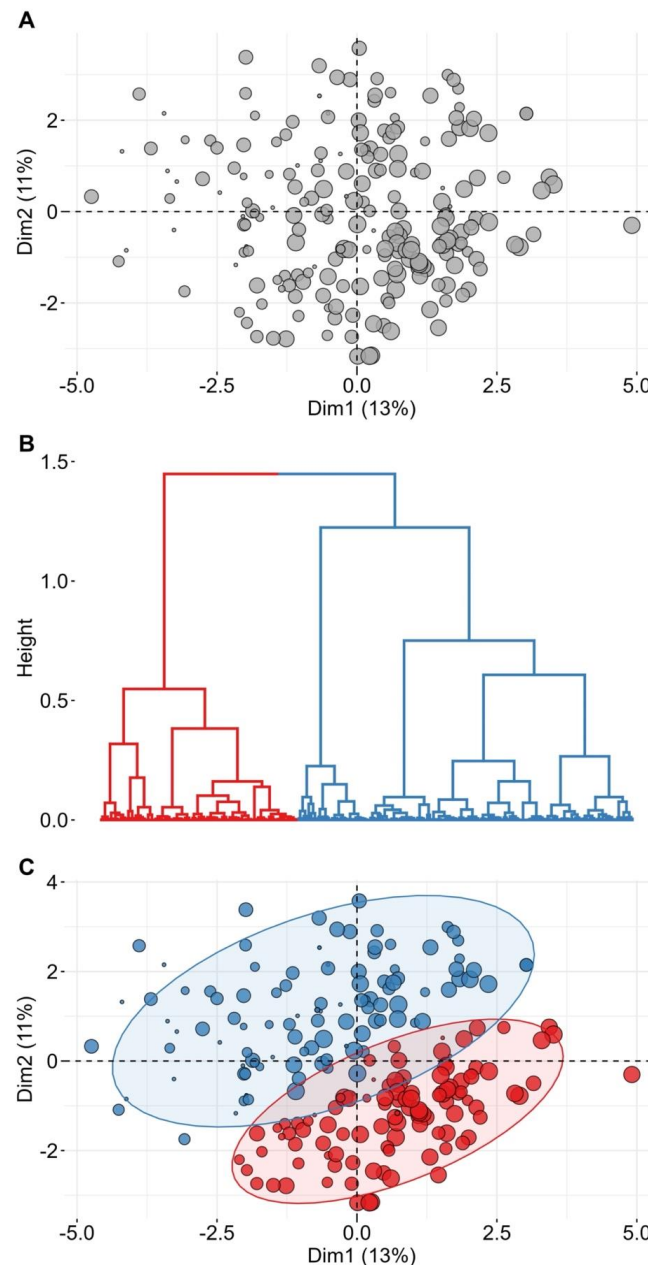


Figure 1. (A) The scatter plot of the contribution of each patient to the first two principal components based on a factor analysis of mixed data. The size of the circles is proportional to the Barthel index on discharge. More commonly, patients with lower indices are located at the upper left region of the scatter plot. (B) The dendrogram for hierarchical clustering. (C) Cluster illustration of patients. The colored concentration ellipses represent patients' clusters.

#### CLUSTER PROFILING

We used a random forest to primarily select important variables for cluster profiling (Figure 2). Then, the selected variables were incorporated into a logistic model to predict cluster membership. However, model diagnostics revealed heavy collinearities for ischemic and DM variables (variance inflation factor = 340.074 and 132.902, respectively). Pairwise comparisons suggested large

effects for the difference between the two clusters in ischemic [88 (80.7%) vs. 12 (11.2%),  $P < 0.001$ , unadjusted odds ratio = 33.17] and DM [8 (7.3%) vs. 61 (57.0%),  $P < 0.001$ , unadjusted odds ratio = 0.06]. Also, the random forest implied the large influence of the two features on models. We excluded ischemic and DM variables and made a stepwise logistic model with 10-fold cross-validation. For the remaining variables, the maximum variance

inflation factor was 2.343. Table 3 shows model fitness and accuracy for cluster profiling and Table 2 presents the final model specification. The model was relatively accurate and well-fitted. Overall, the profiling implied that the two clusters are

significantly different in all but serum K concentration. Both clusters plausibly showed an inverse relation between NIHSS and BI on discharge (both Pearson's correlation coefficient  $r = -0.33$ ,  $P < 0.001$ ).

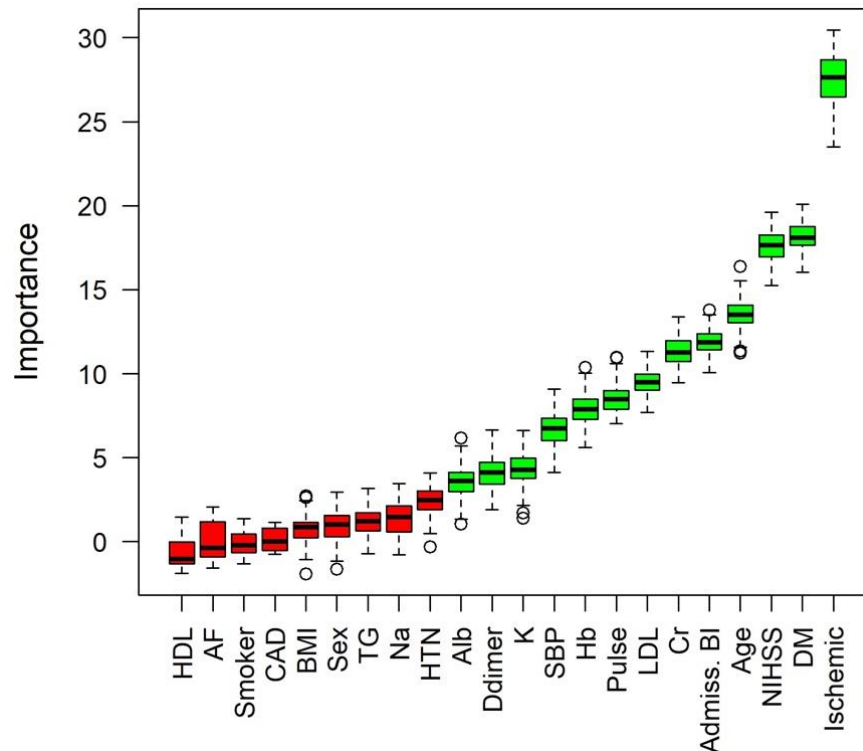


Figure 2. Variable importance for predicting mortality rate. The green boxplots illustrate the confirmed selected features. HDL, High Density Lipoprotein; AF, Atrial Flutter; CAD, Coronary Artery Disease; BMI, Body Mass Index; TG, Triglyceride; Na, sodium; HTN, history of blood Hypertension; Alb, Albumin; SBP, K, potassium; Systolic Blood Pressure, Hb, Hemoglobin; LDL, Low Density Lipoprotein; Cr, Creatinine; BI, Barthel Index; NIHSS, National Institute of Health Stroke Scale; DM, Diabetes Mellitus.

Table 2

The fitness and accuracy of the final logistic model for cluster membership

| Model characteristic         |                                        |
|------------------------------|----------------------------------------|
| Fitness                      | $\chi^2(10) = 203.750$ , $P < 0.001^*$ |
| Nagelkerke R <sup>2</sup>    | 0.81                                   |
| Akaike information criterion | 117.67                                 |
| Hosmer-Lemeshow test         | $\chi^2(8) = 13.372$ , $P = 0.100$     |
| Brier calibration score**    | 0.11                                   |
| Observed                     |                                        |
| Predicted                    | 93                                     |
|                              | 17                                     |
| Predicted                    | 16                                     |
|                              | 90                                     |
| Accuracy (%) (95% CI)        | 84.7 (79.2, 82.2)                      |
| Sensitivity (%)              | 85.3                                   |
| Specificity (%)              | 84.1                                   |
| Positive Predicted Value (%) | 84.6                                   |
| Negative Predicted Value (%) | 84.9                                   |
| C-statistic (%)              | 97.0                                   |

\* significant at  $< 0.05$ .

\*\* the calibration plot had an intercept of 0 and a slope of 1.

Table 3

Logistic model for cluster membership

|                         | Cluster 1 (N = 109) | Cluster 2 (N = 107) | Adjusted OR (95%CI) | P       |
|-------------------------|---------------------|---------------------|---------------------|---------|
| Age (year)              | 55.0 (14.8)         | 65.7 (11.9)         | 1.16 (1.1, 1.23)    | <0.001* |
| SBP (mmHg)              | 125.5 (18.3)        | 134.6 (16.9)        | 1.05 (1.02, 1.09)   | 0.004*  |
| Pulse (beat/min)        | 82.0 (10.4)         | 74.9 (11.2)         | 0.93 (0.87, 0.98)   | 0.008*  |
| D-dimer (µg/mL)         | 2.1 (4.3)           | 1.0 (1.7)           | 0.71 (0.52, 1)      | 0.029*  |
| LDL (mmol/L)            | 2.5 (1.0)           | 1.8 (0.7)           | 0.17 (0.07, 0.36)   | <0.001* |
| Hb (g/L)                | 116.9 (19.0)        | 128.9 (18.1)        | 1.06 (1.03, 1.10)   | 0.001*  |
| K (mmol/L)              | 3.9 (0.4)           | 4.0 (0.4)           | 4.05 (0.95, 19.68)  | 0.068   |
| Cr (µmol/L)             | 55.6 (20.9)         | 72.9 (28.0)         | 1.07 (1.04, 1.1)    | <0.001* |
| NIHSS (score)           | 12.4 (10.6)         | 3.8 (4.3)           | 0.8 (0.7, 0.88)     | <0.001* |
| BI on Admission (score) | 28.5 (26.2)         | 51.8 (28.2)         | 1.03 (1.01, 1.06)   | 0.011*  |

## DISCUSSION

We conducted the study to cluster in-hospital patients with acute stroke based on demographic, laboratory, and clinical data. Following a factor analysis of mixed data two clusters of patients were identified. A simple profiling scheme showed that cluster 2 membership was more associated with a non-ischemic stroke and a history of DM. Overall, patients in cluster 2 were a little older and had a higher mean SBP, Hb, Cr, and BI on admission than cluster 1. Also, cluster 2 patients had lower pulse rates, D-dimer, LDL, and NIHSS than patients in cluster 1. The clusters were different in post-stroke functional recovery. Our results showed that the mean BI on discharge was higher in cluster 2. This means that cluster 2 membership was associated with a better outcome of treatment for stroke. Both groups showed a similar negative association of NIHSS on admission with BI on discharge. Comparisons of the adjusted odds ratios for the resulting model imply that LDL, D-dimer, and NIHSS more influenced cluster membership. Briefly, cluster 2 members can be best described as patients with DM and non-ischemic stroke, lower LDL, D-dimer, and NIHSS. Cluster 1 patients are those with ischemic stroke and higher LDL, D-dimer, and NIHSS.

Uncertainties still need to be addressed concerning ischemic versus non-ischemic stroke recovery and the optimum rehabilitation strategy [23]. A previous study of 1064 individuals with acute stroke showed that patients with intra-cerebral hemorrhage had higher post-treatment improvement in functional status than those with cerebral

infarction [24]. Also, the initial severity of the disability, age, and duration of therapy were identified as significant predictors of functional outcomes after rehabilitation [24]. A recent study on 489 individuals also confirmed that patients with intra-cerebral hemorrhage have greater functional recovery than their ischemic stroke counterparts receiving post-acute care rehabilitation [23]. In our study, the clustering algorithm accurately separated patients with ischemic (cluster 1) from non-ischemic (cluster 2) strokes. Our cluster 2 patients showed a greater mean BI index after the end of therapy. Furthermore, we confirmed that age and NIHSS were significantly different between the clusters.

A recent systematic review showed the insufficiency of evidence regarding the effect of diabetes on motor and cognitive post-stroke recovery [25]. However, DM was suggested to have some influence on the quality of life after a stroke [25]. Also, a previous study of 395 patients with acute stroke implies that DM has no influence on motor and functional outcomes within the acute and post-acute phases after stroke [26]. However, a recent study of 208 patients with stroke suggested that DM is a risk factor for both stroke and delayed recovery after stroke [27]. Kissela *et al.* studied the effects of DM on post-stroke outcome and mortality [28]. They reported that DM unfavorably affects both outcomes and mortality of stroke for patients who are less than age 65 [28]. We did not have the quality of life or long-term post-stroke data. However, our clustering did not give supportive evidence for the association of DM with worse functional recovery. Most of our patients with DM

were cluster 2 members. The mean age for cluster 2 patients was >65 years. This was compatible with the results of Kissela's study. In a large prevalence study, Munoz-Rivas *et al.* included 173,979 discharges of patients admitted with hemorrhagic stroke [29]. They found a significant association of DM with the incidence of hemorrhagic stroke but not with in-hospital mortality [29]. The association of DM with hemorrhagic stroke might be another reason for the high percentage of DM in our cluster 2 (mostly non-ischemic stroke) patients.

Our cluster analysis yielded reasonable results concerning serum LDL concentration. The concentration of LDL had an adjusted odds ratio of <1 for cluster 2 participants. This means a higher serum LDL is more representative of cluster 1 than cluster 2 patients. The reason is that most of the patients with ischemic stroke were members of cluster 1. The magnitude of the odds ratio implies that LDL concentration is a strong discriminator of the patient clusters. It should be noticed that patients with low LDL concentrations are still at risk of hemorrhagic stroke [30]. Similarly, we detected an odds ratio of <1 for D-dimer concentration showing an association of higher D-dimer levels with cluster 1 membership. In a recent study, a high D-dimer level was associated with the risks of acute ischemic, but not hemorrhagic stroke [31]. Moreover, high D-dimer concentrations on admission predicted 90-day poor functional outcomes of patients with acute ischemic stroke [31].

The rationale for the analysis in this study was that cluster analysis has the potential to identify patterns within the demographic and clinical data of hospitalized patients with acute stroke, particularly for grouping of prognostic importance. This allowed exploration of potential clusters by assessing associations between patient characteristics independent of clinical interpretation. However, the clusters differed in terms of functional recovery. This means that patients experience different outcomes of stroke treatment depending on their cluster membership. We reported the clinical implications of this clustering, taking into account

easily measured demographic and clinical characteristics. Incorporating these findings into clinical decision making will enable healthcare professionals to prioritize patient care and better allocate resources, thereby improving the quality of healthcare.

## LIMITATIONS

We incorporated features that are easily measured upon admission of the patient. This allowed for providing applicable recommendations for stratifying patients concerning the possible outcomes of treatment. Meanwhile, we had some limitations in our study. Our sample size was enough to reveal differences between the clusters in some basic patients' features. A larger sample size with plenty of details is needed to allow for discovering more knowledge from the data. Particularly, we were unable to detect the effects of different treatment measures on post-stroke functional recovery. Also, we had the data of BI on two occasions, admission and discharge. There might be interesting patterns recognizable in the longitudinal assessment of BI scores that we did not address. In addition, we did not have access to post-discharge or readmission data. However, our approach and specific findings could serve as a basis for further clinical and basic studies.

## CONCLUSION

In a cluster analysis of in-hospital patients with acute stroke based on a factor analysis of mixed data, two subgroups of patients were accurately identified. Cluster 1 patients were those with ischemic stroke and higher LDL, D-dimer, and NIHSS. Cluster 2 included older patients with a non-ischemic stroke and a history of DM who had a higher mean SBP, Hb, Cr, and BI on admission; and a lower pulse rate, D-dimer, LDL, and NIHSS at baseline compared with patients in cluster 1. In general, functional recovery on discharge was better for cluster 2.

---

**Introducere:** Accidentul vascular cerebral (AVC) este o cauză principală de mortalitate la nivel mondial și o cauză majoră de dizabilitate, având o povară mare pentru pacienți, societate și sistemele de îngrijire. Acest studiu a fost realizat pentru a investiga prezența grupurilor de pacienți în spital cu AVC acut pe baza datelor demografice și clinice. Analiza tip cluster dezvăluie modele în caracteristicile pacientului fără a necesita cunoașterea unei categorii de pacienți predefinite sau ipoteze despre grupările probabile în cadrul datelor.

---

**Metode:** Am efectuat o analiză secundară a datelor anonimizate (open-access) de la pacienții cu AVC acut internați într-un spital între decembrie 2019 și iunie 2021. În total, 216 pacienți (78; 36,1% bărbați) au fost incluși în setul de date analitice cu o medie (SD) a vârstei de 60,3 (14,4). Multe caracteristici demografice și clinice au fost incluse în analiză și a fost utilizat indicele Barthel la externare pentru a compara recuperarea funcțională a grupurilor identificate.

**Rezultate:** Gruparea ierarhică bazată pe componentele principale a identificat două grupuri de 109 și 107 pacienți. Grupurile au fost diferite în scorurile indicelui Barthel la externare, cu media (SD) de 39,3 (29,3) față de 62,6 (29,4);  $t(213,87) = -5,818$ ,  $P < 0,001$ ,  $d$  Cohen (95%CI) =  $-0,80 (-1,07, -0,52)$ . Un model de regresie logistică a arătat că vârsta, tensiunea arterială sistolică, frecvența pulsului, nivelul sanguin D-dimeri, lipoproteinele cu densitate joasă, hemoglobina, concentrația creatininei, valoarea Scalei National Institute of Health Stroke Scale (NIHSS) pentru accident vascular cerebral și scorurile indicelui Barthel la internare au fost predictorii semnificativi ale profilurilor hrupurilor generate (toate  $P \leq 0,029$ ).

**Concluzie:** Există două grupuri la pacienții spitalizați cu AVC acut cu recuperare funcțională semnificativ diferită. Acest lucru permite gruparea prognostică a pacienților spitalizați cu AVC acut pentru prioritizarea îngrijirii sau alocarea resurselor. Grupurile pot fi recunoscute folosind caracteristici demografice și clinice ușor de măsurat.

**Correspondence to:** Mohammad Fathi, M.D., FCCM, Critical Care Quality Improvement Research Center, Shahid Modarres Hospital, Saadat Abad St. Yadegare Imam Highway, Tehran, Islamic Republic of Iran, Postal code: 1998734383  
Tel: (+98) 21 22074087-98  
Fax: (+98) 21 22074101  
Email: m.fathi@sbmu.ac.ir

**Declaration of interest:** The authors declare that they have no competing interests.

**Funding support:** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Author Contributions:** **MH** and **NN** guided development of the study protocol and contributed to the interpretation of the results. **NMA** and **BFB** participated in protocol development, literature review, and planning. **SZBJ** and **NMM** contributed to the design, visualization, and statistical analysis. **MF** contributed to the concept and interpretation of the results, supervised the research, and helped with administrative affair. All the authors participated in drafting and its final approval.

## REFERENCES

1. KATAN M., LUFT A. *Global Burden of Stroke*. Semin Neurol. 2018; **38**(2): p. 208-211.
2. LEKANDER I., WILLERS C., VON EULER M., LILJA M., SUNNERHAGEN K S., PESSAH-RASMUSSEN H, et al., *Relationship between functional disability and costs one and two years post stroke*. PLoS One. 2017; **12**(4): p. e0174861.
3. O'DONNELL M J., XAVIER D., LIU L., ZHANG H., CHIN S L., RAO-MELACINI P., et al., *Risk factors for ischaemic and intracerebral haemorrhagic stroke in 22 countries (the INTERSTROKE study): a case-control study*. Lancet. 2010; **376**(9735): p. 112-23.
4. QIN H., CHEN Y., LIU G., TURNBULL I., ZHANG R., LI Z., et al., *Management characteristics and prognosis after stroke in China: findings from a large nationwide stroke registry*. Stroke Vasc Neurol. 2021; **6**(1): p. 1-9.
5. QU S., ZHOU M., JIAO S., ZHANG Z., XUE K., LONG J., et al., *Optimizing acute stroke outcome prediction models: Comparison of generalized regression neural networks and logistic regressions*. PLoS One. 2022; **17**(5): p. e0267747.
6. CAMPAGNINI S., ARIENTI C., PATRINI M., LIUZZI P., MANNINI ACARROZZA M C. *Machine learning methods for functional recovery prediction and prognosis in post-stroke rehabilitation: a systematic review*. J Neuroeng Rehabil. 2022; **19**(1): p. 54.
7. MOHAMMADI T., MOHAMMADI B. *Neuropsychological features in children and adults with congenital heart disease: an exploratory data analysis*. Psychol Health Med. 2022; p. 1-14.
8. MOHAMMADI T., MOHAMMADI B. *Drawing clinical pictures of heart failure with high mortality risk*. Informatics in Medicine Unlocked. 2021; **26**: p. 100752.
9. YAN Z., WEI D., XU S., ZHANG J., YANG C., HE X., et al., *Determining levels of linguistic deficit by applying cluster analysis to the aphasia quotient of Western Aphasia Battery in post-stroke aphasia*. Sci Rep. 2022; **12**(1): p. 15108.

10. HARA T., NIIMI M.,YAMADA N., SHIMAMOTO Y., MASUDA G., HARA H., *et al.*, *Prognosis prediction of the effect of botulinum toxin therapy and intensive rehabilitation on the upper arm function in post-stroke patients using hierarchical cluster analysis.* Disabil Rehabil. 2022; **44**(22): p. 6815-6823.
11. HENDRICKX W., RIVEROS C., ASKIM T., BUSSMANN J B J., CALLISAYA M L.,CHASTIN S F M., *et al.*, *An Exploration of Sedentary Behavior Patterns in Community-Dwelling People With Stroke: A Cluster-Based Analysis.* J Neurol Phys Ther. 2021; **45**(3): p. 221-227.
12. CHEN C Y., CHEN C L. *Recognizable Clinical Subtypes of Obstructive Sleep Apnea After Ischemic Stroke: A Cluster Analysis.* Nat Sci Sleep. 2021; **13**: p. 283-290.
13. KIM H., KIM Y H., KIM S JCHOI M T. *Pathological gait clustering in post-stroke patients using motion capture data.* Gait Posture. 2022; **94**: p. 210-216.
14. DING L., MANE R.,WU Z., JIANG Y., MENG X., JING J., *et al.*, *Data-driven clustering approach to identify novel phenotypes using multiple biomarkers in acute ischaemic stroke: A retrospective, multicentre cohort study.* EclinicalMedicine. 2022; **53**: p. 101639.
15. CHEN J., LIU S.,WU M., DAI L.,WANG J., XIE W., *et al.*, *Twenty-four-hour National Institute of Health Stroke Scale predicts short- and long-term outcomes of basilar artery occlusion after endovascular treatment.* Front Aging Neurosci. 2022; **14**: p. 941034.
16. KHAN M S A., AHMAD S., GHAFOR B., SHAH M H., MUMTAZ H., AHMAD W., *et al.*, *Inpatient assessment of the neurological outcome of acute stroke patients based on the National Institute of Health Stroke Scale (NIHSS).* Ann Med Surg (Lond). 2022; **82**: p. 104770.
17. LEE E Y., SOHN M K., LEE J M., KIM D Y., SHIN Y I., OH G J., *et al.*, *Changes in Long-Term Functional Independence in Patients with Moderate and Severe Ischemic Stroke: Comparison of the Responsiveness of the Modified Barthel Index and the Functional Independence Measure.* Int J Environ Res Public Health. 2022; **19**(15).
18. DOS REIS N F., FIGUEIREDO F., BISCARO R R M., LUNARDELLI E BMAURICI R. *Psychometric Properties of the Barthel Index Used at Intensive Care Unit Discharge.* Am J Crit Care. 2022; **31**(1): p. 65-72.
19. WANG Y C., CHANG P F., CHEN Y M., LEE Y C., HUANG S L., CHEN M H., *et al.*, *Comparison of responsiveness of the Barthel Index and modified Barthel Index in patients with stroke.* Disabil Rehabil. 2022; p. 1-6.
20. LIU X., ZHOU M., ZHAO J., GAO Y., WANG Y., ZHOU J., *et al.*, *Functional Independence and Disability Evaluation in Stroke Patients: Optimal Cutoff Scores for a Pictorial-Based Longshi Scale, Barthel Index, and Modified Rankin Scale.* Front Neurol. 2022; **13**: p. 710852.
21. NAKIBUUKA J., SAJATOVIC M., NANKABIRWA J., SSENDIKADIWA C., FURLAN A J., KATABIRA E., *et al.*, *Early mortality and functional outcome after acute stroke in Uganda: prospective study with 30 day follow-up.* Springerplus. 2015; **4**: p. 450.
22. MOHAMMADI T., MOHAMMADI B. *The long-term prognostic value provided by Coronary CT angiography.* Eur J Intern Med. 2022.
23. CHU C L., CHEN Y P., CHEN C C P., CHEN C K., CHANG H N., CHANG C H., *et al.*, *Functional Recovery Patterns of Hemorrhagic and Ischemic Stroke Patients Under Post-Acute Care Rehabilitation Program.* Neuropsychiatr Dis Treat. 2020; **16**: p. 1975-1985.
24. KELLY P J., FURIE K L., SHAFQAT S., RALLIS N., CHANG YSTEIN J. *Functional recovery following rehabilitation after hemorrhagic and ischemic stroke.* Arch Phys Med Rehabil. 2003; **84**(7): p. 968-72.
25. YANG S., BOUDIER-REVERET M., KWON S., LEE M YCHANG M C. *Effect of Diabetes on Post-stroke Recovery: A Systematic Narrative Review.* Front Neurol. 2021; **12**: p. 747878.
26. NANNETTI L., PACI M., BACCINI M., RINALDI L ATAITI P G. *Recovery from stroke in patients with diabetes mellitus.* J Diabetes Complications. 2009; **23**(4): p. 249-54.
27. CHATURVEDI P., SINGH A K., TIWARI VTHACKER A K. *Diabetes mellitus type 2 impedes functional recovery, neuroplasticity and quality of life after stroke.* J Family Med Prim Care. 2020; **9**(2): p. 1035-1041.
28. KISSELA B., AIR E. *Diabetes: impact on stroke risk and poststroke recovery.* Semin Neurol. 2006; **26**(1): p. 100-7.
29. MUNOZ-RIVAS N., MENDEZ-BAILON M., HERNANDEZ-BARRERA V., DE MIGUEL-YANES J M., JIMENEZ-GARCIA R., ESTEBAN-HERNANDEZ J., *et al.*, *Type 2 Diabetes and Hemorrhagic Stroke: A Population-Based Study in Spain from 2003 to 2012.* J Stroke Cerebrovasc Dis. 2016; **25**(6): p. 1431-43.
30. WU Z., HUANG Z., LICHTENSTEIN A H., LIU Y., CHEN S., JIN Y., *et al.*, *The risk of ischemic stroke and hemorrhagic stroke in Chinese adults with low-density lipoprotein cholesterol concentrations < 70 mg/dL.* BMC Med. 2021; **19**(1): p. 142.
31. YUAN B., YANG T., YAN T., CHENG WBU X. *Relationships Between D-Dimer Levels and Stroke Risk as Well as Adverse Clinical Outcomes After Acute Ischemic Stroke or Transient Ischemic Attack: A Systematic Review and Meta-Analysis.* Front Neurol. 2021; **12**: p. 670730.