Partitioning Stroke Patients, Determining Related Factors, and Comparing Derived Clusters Based on 12-Month Health Outcomes

Ali Soroush, MD; Payam Sariasi, MD; Nadya Baharirad, BS; Nasim Shams-Alizadeh, PhD; Saeid Komasi, MSc*

1Lifestyle Modification Research Center, Imam Reza Hospital, Kermanshah University of Medical Sciences, Kermanshah, Iran
2Department of Neurology, Imam Reza Hospital, Kermanshah University of Medical Sciences, Kermanshah, Iran
3Clinical Research Development Center, Imam Reza Hospital, Kermanshah University of Medical Sciences, Kermanshah, Iran

Abstract

**Background:** (i) Cluster analysis and partitioning samples based on cardio-cerebrovascular histories and length of stay (LOS); (ii) Determining related demographic and medical factors in individual clusters; and (iii) Comparing clusters based on 12-month health outcomes.

**Methods:** The statistical population of the study included 2,293 stroke patients hospitalized in Imam Reza hospital of Kermanshah city from January 1, 2015, to December 31, 2016. After a one-year follow-up, the data collection window was closed on December 31, 2017. The patients’ data were extracted from the electronic hospital information system (HIS). Two-step cluster analysis (TSCA), chi-square, Fisher exact, Kruskal-Wallis, and Mann-Whitney U tests, as well as multinomial logistic regression analysis were the analysis methods.

**Results:** This model suggested five distinct clusters: the patients (i) without any cardio-cerebrovascular history and LOS = 5 days (36.2%); (ii) without any cardio-cerebrovascular history and LOS = 6 days (21.6%); (iii) with cerebrovascular history and LOS = 6 days (18.6%); (iv) with cardio-cerebrovascular history and LOS = 6 days (16.1%); and (v) with cardio-cerebrovascular history and LOS = 6 days (7.5%). Hypertension, diabetes, and smoking were respectively the most significant modifiable risk factors, while sex, cerebrovascular diseases in the family, and age were respectively the most significant non-modifiable risk factors in high-risk clusters and LOS = 6 days. Compared to Cluster 1 (reference), during a one-year follow-up, a larger number of members in Clusters 3 and 5 were readmitted and/or expired.

**Conclusion:** Considering the modifiable risk factors identified in the current study, providing programs for preventing readmission and potential death caused by stroke for Clusters 3 and 5 seems essential.

**Keywords:** Clustering, Hospitalization, Medical history taking, Mortality, Patient readmission, Stroke

Cite this article as: Soroush A, Sariasi, P; Baharirad N, Shams-Alizadeh N, Komasi S. Partitioning stroke patients, determining related factors, and comparing derived clusters based on 12-month health outcomes. Arch Iran Med. 2019;22(12):708–715.

Received: March 17, 2019, Accepted: September 18, 2019, ePublished: December 1, 2019

Introduction

Stroke is one of the most serious health threats in the world, such that it is considered the second cause of mortality and the third cause of disability.1,2 It imposes a heavy financial burden on the health system.3 Although in recent decades, stroke and deaths from it are decreasing in the US, 975 thousand strokes occur annually in this country, leading to death in 57%–68% of cases.4,5 Stroke is a common disease in Iran, as well; recent reports show a prevalence of 139 per 100,000 people, leading to death in 11.4%–40.6% of cases.6,5 The results of two review studies show that the prevalence of stroke in Iran ranges from 23 to 139 cases per 100,000 people.6,4 These studies introduce factors such as aging, hypertension, diabetes, dyslipidemia, cardiovascular diseases (CVDs), and a history of stroke as the main risk factors for increasing the likelihood of stroke.6,7,9

Stroke, as a chronic disease, is sometimes accompanied by readmission and mortality. Based on recent reports, gender, aging, diabetes, history of stroke and/or cerebrovascular complications, CVDs, type of stroke, and longer length of stay (LOS) in the hospital are the most important measures for predicting readmission and mortality caused by stroke.10,11 Besides histories related to cardio-cerebrovascular diseases, LOS is one of the factors related to mortality, and it can increase the likelihood of mortality 1.45 times.12,13 Moreover, readmission increases the risk of mortality.10,11

While the abovementioned studies have addressed the risk factors of readmission and mortality among stroke patients, there are few studies on partitioning patients based on their cardio-cerebrovascular history as well as LOS. On the other hand, recent studies have shown that cluster analysis is a good method for categorizing and identifying high-risk stroke patients.14,15 Therefore,
systematic categorization of stroke patients for identifying groups in risk of readmission and mortality is likely to be effective in planning and delivering timely service to these patients. Based on these considerations, the current study was carried out in order to realize three objectives: (i) cluster analysis and partitioning the participants based on cardio-cerebrovascular history and LOS; (ii) determining related demographic and medical factors in individual clusters; and (iii) comparing clusters based on 12-month health outcomes.

Materials and Methods

Design and Context
In this existing data study, we evaluated stroke patients hospitalized in Imam Reza hospital of Kermanshah city (Iran) from 2015 to 2016. Kermanshah city, the capital of Kermanshah province in western Iran, is located 326 miles west of Tehran. According to the 2016 census, the total population of Kermanshah province was 1,952,434 people (988,015 men). More than 947,000 people of this province are living in Kermanshah city. Imam Reza hospital of Kermanshah city is a general public hospital with 841 beds and more than 20 health wards. The neurology ward is one of the most active departments of this hospital, providing healthcare services for a large number of hospitalized stroke patients.

Participants and Data Collection
The statistical population of the study includes all patients diagnosed with stroke who were hospitalized in Imam Reza hospital of Kermanshah city from January 1, 2015, to December 31, 2016. The initial sample of this existing data study included 2,347 individuals, whose information was extracted from the electronic hospital information system (HIS). Of these, 54 individuals were eliminated from the sample due to missing records. The final sample included 2,293 patients. From February to April 2018, the patients’ data were extracted by a nurse using the electronic HIS and recorded in the research forms designed for the current study. The electronic HIS includes the current medical records of the patients as well as all their outpatient and hospitalization records. In the next stage, the data related to the readmission of patients and in-hospital mortality within one year of their release from the hospital was recorded. Therefore, the data collection window for the study was closed on December 31, 2017. The data for the study were extracted under the supervision of a neurologist and the accuracy of the information for each patient was later rechecked by another nurse. The recorded data included gender, age, histories related to hypertension, diabetes, hyperlipidemia, smoking, drug addiction, CVDs, cerebrovascular diseases (CeVDs), and CeVDs in the family, LOS in the hospital, type of release (recovery/death), readmission, waiting time for readmission, LOS after readmission, and type of release after readmission (recovery/death).

Data Analysis
The data related to continuous variables were reported as mean and standard deviation, while the discrete data were reported as values (percentages) or median (interquartile range). Before performing the main analysis, the patients’ records were first coded as no (=0)/yes (=1). The LOS score was entered into the analysis in the form of mean and standard deviation. In the main analysis, in order to identify the clusters, two-step cluster analysis (TSCA) was performed. This method was used due to the large size of the sample and the presence of continuous and discrete variables. TSCA determined the significance rank of the classification variables playing a role in predicting the model and identified the number of clusters automatically. The fitting of the model was achieved based on Schwarz’s Bayesian information criterion (BIC) using the average silhouette coefficient. The silhouette coefficient is a measure of internal validity, which ranges from 0 to 1. A score closer to 1 indicates a better model. Cluster analysis was performed using cerebrovascular history, cardiovascular history, and LOS. In the next stage, the clusters were compared with regard to cerebrovascular and cardiovascular histories using the Fisher exact test (because of the absence of the main assumption of the chi-square test). Due to the skewed distribution in the LOS, the clusters were compared using Kruskal-Wallis test.

In the next step, multinomial logistic regression analysis was performed for identifying the associated factors of the derived clusters. Gender and age along with all the other medical variables including the histories related to hypertension, diabetes, hyperlipidemia, smoking, addiction, and CeVDs in the family were entered into the model simultaneously. Since five clusters were identified, Cluster 1 (healthier subjects) was considered as the reference cluster and analyses were adjusted for gender and age. The results of the analysis were presented in the form of adjusted odds ratios (OR) with 95% confidence intervals (CIs).

In the final stage, the type of patient discharge, readmission related to CeVDs and the type of discharge, the waiting period between initial hospitalization and readmission, and LOS after the readmission were explored in the total population and based on individual clusters. In order to evaluate the significance of the difference between individual clusters and the reference cluster, Mann-Whitney U test and one-variable chi-square test were used. In order to do this, the frequency proportion of each cluster compared to the reference cluster was calculated, and after assigning weights to the frequencies, the significance of the difference between the two clusters was computed. The discrete variables (waiting time for readmission and LOS in readmission) were compared using Mann-Whitney U test. All the statistical analyses were carried out using...
Results
Identified Clusters
Table 1 depicts the results of the TSCA as well as a summary of the model. As can be seen, the silhouette measure of cohesion and separation is completely acceptable. All three factors have played a significant role in identifying the clusters. Based on the results in this table, there is a significant difference ($P<0.001$) between the clusters with regard to cerebrovascular and cardiovascular histories (since in three cells, frequencies are less than 25% of the total frequency, the Fisher exact test was used). In addition, there is a significant difference between the clusters with regard to the LOS in the hospital (Kruskal-Wallis test = 11.919; $P=0.018$). This model has suggested five clusters, as follows: the patients (i) without any cardio-cerebrovascular history and LOS = 5 days (36.2%); (ii) without any cardio-cerebrovascular history and LOS = 6 days (21.6%); (iii) with cerebrovascular history and LOS = 6 days (18.6%); (iv) with cardiovascular history and LOS = 6 days (16.1%); and (v) with cardio-cerebrovascular history and LOS = 6 days (7.5%).

Demographic and Medical Factors Related to the Clusters
Table 2 presents the characteristics of the participants based on individual clusters. Moreover, this table depicts the results of multinomial logistic regression after adjustment on individual clusters. Compared to the reference cluster (Cluster 1), Clusters 2 ($P<0.001$), 3 ($P=0.011$), and 4 ($P<0.001$) contain fewer women and more men. Moreover, Cluster 2 includes younger patients ($P=0.002$), and Cluster 4 ($P=0.012$) includes older patients. The results of this table also indicate that the CeVDs in the family ($P=0.013$) and smoking ($P<0.001$) in Cluster 2 are significantly higher compared to the patients in the reference cluster. Moreover, history of hypertension and diabetes in Clusters 3 ($P=0.008; P=0.018$), 4 ($P=0.001; P=0.007$), and 5 ($P=0.002; P=0.024$) are significantly higher compared to the reference cluster.

Comparing Clusters with Regard to 12-Month Health Outcomes
Table 3 compares clusters based on 12-month health outcomes. As can be seen, compared to the reference cluster, the frequency of readmission and mortality caused by it is higher in Clusters 3 ($P=0.021; P=0.047$) and 5 ($P=0.002; P=0.040$). In Clusters 2, waiting time for readmission is shorter than samples in the reference cluster ($P=0.049$). There is no significant difference with regard to other variables between the other clusters and the reference cluster ($P>0.05$).

Discussion
Main Findings
- Five clusters were identified based on cerebrovascular history, cardiovascular history, and LOS: two clusters without any cardio-cerebrovascular history and LOS less than or equal to six days, and three clusters with cardiovascular history, cerebrovascular history, or both and LOS equal to 6 days.
- The results show that 42.2% of the patients in the sample (Clusters 3, 4, and 5) have cardio-cerebrovascular history and LOS equal to 6 days. Only 36.2% of the participants (Cluster 1) do not have a history of the disease and have a LOS less than six days.
- The history of hypertension and diabetes, sex, smoking, history of CeVDs in the family, and age are

Table 1. Cardio-cerebrovascular History and LOS Profile Derived from Two-Stage Cluster Analysis (n = 2291)

<table>
<thead>
<tr>
<th>Factors</th>
<th>Total (n = 2293)</th>
<th>Cluster 1 (n = 838; 36.2%)</th>
<th>Cluster 2 (n = 495; 21.6%)</th>
<th>Cluster 3 (n = 427; 18.6%)</th>
<th>Cluster 4 (n = 376; 16.1%)</th>
<th>Cluster 5 (n = 171; 7.5%)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Patients without Cardio-cerebrovascular History and LOS = 5 Day</td>
<td>Patients without Cardio-cerebrovascular History and LOS = 6 Day</td>
<td>Patients with Cardio-cerebrovascular History and LOS = 6 Day</td>
<td>Patients with Cardio-cerebrovascular History and LOS = 6 Day</td>
<td>Patients with Cardio-cerebrovascular History and LOS = 6 Day</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerebrovascular history (%)</td>
<td>598 (26.1)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>427 (100)</td>
<td>0 (0)</td>
<td>171 (100)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Cardiovascular history (%)</td>
<td>541 (23.6)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>370 (100)</td>
<td>171 (100)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Length of stay, day (median &amp; interquartile range)</td>
<td>6 (4-10)</td>
<td>5 (3-9)</td>
<td>6 (4-10)</td>
<td>6 (4-11)</td>
<td>6 (4-10)</td>
<td>6 (4-10)</td>
<td>0.018</td>
</tr>
</tbody>
</table>

Summary of model: Silhouette measure of cohesion and separation is 0.7; Ratio of sizes for largest to the smallest cluster is 4.85; the most important predictors are: cerebrovascular and cardiovascular history = 1.0, Length of stay = 0.9.

NA: Not applicable.

* Considering that in three cells, frequencies are less than 25% of the total frequency (=zero), the Fisher’s exact test was used.

* Kruskal-Wallis test.
Table 2. Results of Multinomial Logistic Regression for Identifying Correlates

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Total (n = 2293)</th>
<th>Cluster 1 (n = 830; 36.2%) Reference</th>
<th>Cluster 2 (n = 495; 21.6%)</th>
<th>Cluster 3 (n = 427; 18.6%)</th>
<th>Cluster 4 (n = 370; 16.1%)</th>
<th>Cluster 5 (n = 171; 7.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Patients without Cardio-Cerebrovascular History and LOS = 5 Day</td>
<td>Patients without Cardio-Cerebrovascular History and LOS = 6 Day</td>
<td>Patients with Cerebrovascular History and LOS = 6 Day</td>
<td>Patients with Cardiovascular History and LOS = 6 Day</td>
<td>Patients with Cardio-Cerebrovascular History and LOS = 6 Day</td>
</tr>
<tr>
<td>Sex, female (%)</td>
<td>1070 (46.7)</td>
<td>409 (49.3)</td>
<td>225 (45.5)</td>
<td>178 (43.8)</td>
<td>168 (45.4)</td>
<td>81 (47.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.13 [0.04 – 0.36]</td>
<td>0.20 [0.06 – 0.69]</td>
<td>0.10 [0.03 – 0.38]</td>
<td>0.38 [0.05 – 2.55]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P &lt; 0.001</td>
<td>P = 0.011</td>
<td>P &lt; 0.001</td>
<td>P = 0.316</td>
</tr>
<tr>
<td>Age, y (Mean SD)</td>
<td>67.3 ± 14.3</td>
<td>65.4 ± 15.2</td>
<td>65.1 ± 15.5</td>
<td>69.4 ± 12.5</td>
<td>70.0 ± 12.2</td>
<td>71.3 ± 11.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.98 [0.97 – 0.99]</td>
<td>1.0 [1.0 – 1.2]</td>
<td>1.0 [0.99 – 1.02]</td>
<td>1.02 [1.0 – 1.04]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P = 0.002</td>
<td>P = 0.163</td>
<td>P = 0.323</td>
<td>P = 0.012</td>
</tr>
<tr>
<td>Medical history (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HTN</td>
<td>1358 (59.2)</td>
<td>446 (53.7)</td>
<td>263 (53.1)</td>
<td>276 (64.6)</td>
<td>249 (67.3)</td>
<td>124 (72.5)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.0 [0.79 – 1.27]</td>
<td>1.41 [1.10 – 1.82]</td>
<td>1.56 [1.19 – 2.05]</td>
<td>1.83 [1.25 – 2.67]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P = 0.996</td>
<td>P = 0.008</td>
<td>P = 0.001</td>
<td>P = 0.002</td>
</tr>
<tr>
<td>DM</td>
<td>537 (23.4)</td>
<td>162 (19.5)</td>
<td>107 (21.6)</td>
<td>113 (26.5)</td>
<td>102 (27.6)</td>
<td>53 (31.0)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1.19 [0.89 – 1.59]</td>
<td>1.42 [1.06 – 1.89]</td>
<td>1.51 [1.12 – 2.04]</td>
<td>1.57 [1.06 – 2.31]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P = 0.235</td>
<td>P = 0.018</td>
<td>P = 0.007</td>
<td>P = 0.024</td>
</tr>
<tr>
<td>HLP</td>
<td>228 (9.9)</td>
<td>83 (10.0)</td>
<td>41 (8.3)</td>
<td>44 (10.3)</td>
<td>32 (8.6)</td>
<td>28 (16.4)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.81 [0.54 – 1.21]</td>
<td>0.93 [0.62 – 1.39]</td>
<td>0.74 [0.48 – 1.16]</td>
<td>1.47 [0.91 – 2.39]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P = 0.296</td>
<td>P = 0.722</td>
<td>P = 0.189</td>
<td>P = 0.117</td>
</tr>
<tr>
<td>Cerebrovascular disease in family</td>
<td>30 (1.3)</td>
<td>19 (2.3)</td>
<td>2 (0.4)</td>
<td>4 (0.9)</td>
<td>3 (0.8)</td>
<td>2 (1.2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.16 [0.4 – 0.68]</td>
<td>0.40 [0.13 – 1.20]</td>
<td>0.37 [0.11 – 1.27]</td>
<td>0.50 [0.11 – 2.23]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P = 0.013</td>
<td>P = 0.102</td>
<td>P = 0.114</td>
<td>P = 0.366</td>
</tr>
<tr>
<td>Smoking</td>
<td>210 (9.2)</td>
<td>59 (7.1)</td>
<td>68 (13.7)</td>
<td>41 (9.6)</td>
<td>27 (7.3)</td>
<td>15 (8.8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.09 [1.38 – 3.17]</td>
<td>1.19 [0.75 – 1.89]</td>
<td>1.01 [0.60 – 1.70]</td>
<td>1.10 [0.57 – 2.12]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P &lt; 0.001</td>
<td>P = 0.467</td>
<td>P = 0.958</td>
<td>P = 0.781</td>
</tr>
<tr>
<td>Drug addiction</td>
<td>197 (8.6)</td>
<td>60 (7.2)</td>
<td>48 (9.7)</td>
<td>44 (10.3)</td>
<td>26 (7.0)</td>
<td>19 (11.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.93 [0.59 – 1.46]</td>
<td>1.37 [0.87 – 2.17]</td>
<td>0.97 [0.57 – 1.63]</td>
<td>1.69 [0.92 – 3.11]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>P = 0.757</td>
<td>P = 0.173</td>
<td>P = 0.898</td>
<td>P = 0.093</td>
</tr>
</tbody>
</table>

HTN, hypertension; DM, diabetes mellitus; HLP, hyperlipidemia.

The socio-demographics and other factors in this table were all included as covariates in the generation of the multinomial logistic regression model. The results were adjusted for age and sex (using interaction effects between the two variables on the model). Boldface indicates statistical significance (P < 0.05).

Summary of model: The model fitting information is: Chi-square = 170.805, P < 0.0005; Pseudo R-square based on McFadden and Nagelkerke = 0.025 to 0.076.
<table>
<thead>
<tr>
<th>Components</th>
<th>Total (n = 2293)</th>
<th>Cluster 1 (n = 830; 36.2%)</th>
<th>Cluster 2 (n = 495; 21.6%)</th>
<th>Cluster 3 (n = 427; 18.6%)</th>
<th>Cluster 4 (n = 370; 16.1%)</th>
<th>Cluster 5 (n = 171; 7.5%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital discharge (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved</td>
<td>1936 (84.4)</td>
<td>705 (84.9)</td>
<td>367 (85.9)</td>
<td>311 (82.2)</td>
<td>146 (85.9)</td>
<td>0.816</td>
</tr>
<tr>
<td>Dead</td>
<td>357 (15.6)</td>
<td>125 (15.1)</td>
<td>60 (14.1)</td>
<td>59 (15.9)</td>
<td>25 (14.6)</td>
<td>0.999</td>
</tr>
<tr>
<td>Hospital readmission (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Improved</td>
<td>104 (4.5)</td>
<td>17 (2.0)</td>
<td>10 (2.0)</td>
<td>12 (2.4)</td>
<td>8 (2.2)</td>
<td>0.999</td>
</tr>
<tr>
<td>Dead</td>
<td>22 (1.0)</td>
<td>12 (1.4)</td>
<td>2 (0.5)</td>
<td>2 (0.4)</td>
<td>3 (1.8)</td>
<td>0.999</td>
</tr>
<tr>
<td>LOS in readmission, day</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median &amp; interquartile range</td>
<td>6 (5-10)</td>
<td>6 (5-10)</td>
<td>6 (5-10)</td>
<td>5 (4-10)</td>
<td>5 (4-10)</td>
<td>0.856</td>
</tr>
</tbody>
</table>

LOS, length of stay. P value of chi-square test; P value of Mann-Whitney U test for comparing all the groups with the reference group; Boldface indicates statistical significance (P < 0.05) compared to the reference group.
During a one-year follow-up, a larger number of members of Clusters 3 and 5, compared to the reference cluster, were readmitted and died after readmission. Moreover, samples of Cluster 2 had a shorter waiting time for readmission.

In the current study, five clusters were identified and suggested based on cardio-cerebrovascular histories and LOS. In line with previous studies, our study was able to successfully categorize the patients into different groups, providing valuable consistent information. Van Rheenen et al. were able to successfully classify stroke patients based on the type of stroke, mortality, and risk factors. Partitioning the stroke risk factors and its related family history can be effective in identifying individuals vulnerable to the risk of stroke.

The results of the current study show that hypertension, diabetes, and smoking are the most significant modifiable predictors for risky clusters and LOS equal to 6 days. Recent original and review studies have comprehensively highlighted the importance of these modifiable factors and their related mechanisms for preventing stroke. Hypertension can be mitigated through appropriate medication, and diabetes can be controlled through correct diet and nutrition, regular exercise, and stress control. Controlling these risk factors can, in turn, reduce the risk of stroke or its fatal outcomes. While the reports related to the effects of educational and behavioral interventions on controlling the risk factors of stroke are contradictory, these interventions may be useful if they are accurately designed and cost-effective. Moreover, smoking is one of the most significant risk factors for stroke and quitting smoking can significantly reduce the likelihood of this disease.

Furthermore, a history of cerebrovascular diseases in the family, aging, and being male are the most important non-modifiable predictors for high-risk clusters. Recently, a review study provided a comprehensive exploration of the genetic risk of stroke and the mechanisms related to it. A family history of stroke is an independent risk factor for development of stroke in healthy populations. Studies on large populations have shown that early stroke in the family and a history stroke in parents or siblings can significantly increase the likelihood of this event for an individual. In addition, a family history of stroke is related to poor levels of physiological factors such as hypertension.

With regard to age, previous studies clearly confirm the role of advanced age in increasing the risk of stroke and the mortality caused by it. However, the findings related to the risk of stroke in men and women are contradictory; i.e. the findings of some studies show that stroke is more prevalent and severe among women, while other studies have argued that men are more vulnerable to the risk of stroke. These differences are most likely related to sex hormones and their changes during various life stages, as well as the complex interaction between these hormones and the immune system. Reinforcing this argument, the findings of another study show that while the prevalence of stroke, in general, is higher among women, young men will experience stroke more than women.

In the one-year follow-up, clusters containing patients with a history of cerebrovascular diseases and LOS equal to six days (26.1% in total) were more prone to readmission and mortality compared to the participants in the reference cluster. A history of cerebrovascular diseases is one of the most important risk factors for repeated strokes; indeed, roughly 14%–45% of patients with this sort of history will develop repeated strokes. On the other hand, increased LOS, which is generally due to complications such as infection, pneumonia, and constipation, is directly related to readmission and mortality.

Limitations
The retrospective nature of the study and lack of direct interviews with the patients and their family members were the most important limitations of this study. Moreover, the current study does not exploit various types and intensity of strokes. Finally, the small number of cases in some of the subgroups in terms of events such as cerebrovascular disease in family, HLP, smoking, and drug addiction may have resulted in sparse-data bias. This makes it difficult to interpret logistic regression results. Therefore, one of the limitations of the study is the likelihood of sparse-data bias because of the small number of events in the subgroups. Mitigating these challenges in future studies can facilitate the generalization of the findings to other populations.

Conclusion
In conclusions, partitioning patients based on cardio-cerebrovascular history and LOS will provide useful information about the potential predictors and outcomes of stroke. Hypertension, diabetes, and smoking were respectively the most significant modifiable risk factors, while sex, cerebrovascular diseases in the family, and age were respectively the most significant non-modifiable risk factors in high-risk clusters and LOS equal to six days. Considering the modifiable risk factors identified in the current study, providing programs for preventing readmission and potential death for Clusters 3 and 5 seems essential.

Authors’ Contribution
AS, PS, NB, NSA, and SK prepared study proposal. NB and NSA are involved in the data gathering process. SK generated study hypotheses and drafted the manuscript. SK analyzed data. Critical revisions were done by AS and PS. All authors read and approved the final manuscript.
Conflict of Interest Disclosures
There are no conflicts of interest.

Ethical Statement
Kermanshah University of Medical Sciences (ID: KUMS. REC.1395.325).

Financial Support
Kermanshah University of Medical Sciences (ID: 95282).

Acknowledgments
We extend our appreciation to the Kermanshah University of Medical Sciences for funding support for this project. Ali Soroush received a financial support from the Kermanshah University of Medical Sciences (grant number: 95282).

References
Correlates and Outcomes of Stroke Profiles

CD009103.pub3.


© 2019 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.