Population Level Analysis of Acute Stroke Care Patterns in Hungary

István Vassányi, Zsolt Vassy
Medical Informatics Research and Development Centre
University of Pannonia, Veszprém, Hungary
e-mail: vassanyi@almos.vein.hu

Tamás Kováts, György Surján
National Health Development Institute
Budapest, Hungary
e-mail: kovats.tamas@aeek.hu

Zoltán Nagy
National Institute of Clinical Neurosciences
Budapest, Hungary
e-mail: profnagyzoltan@gmail.com

Abstract—**Stroke is a global health challenge and it represents a significant economic and social burden. The treatment of stroke requires urgent and coordinated procedures. In this clinical data analysis study, our aim is to identify anomalies in the stroke care system. We analyze the stroke care patterns using case records of all 281,948 publicly financed cases between 2010 and 2017 in Hungary. The essence of the method is creating care events of some basic types, organizing event series into episodes, classifying episodes with respect to relevant care patterns, computing 'spectra' of episode type frequencies for the providers and forming clusters of the providers based on the correlations among their spectra. A similar method is applied for postal code areas. The results show that two clusters can be defined that divide the 61 clinics into a smaller and a larger cluster with significantly different care practices. The spatial analysis also revealed that the clusters of the postal code areas form geographically co-located patches marking anomalies in the stroke care system. The novelty of the paper is the proposed method and its application to stroke care. The findings may be used for quality management of the national stroke network.**

Keywords-stroke care; patient pathways; clustering; clinical data analysis.

I. INTRODUCTION

Stroke is major cause of death in developed countries and the decreased quality of life of the survivors also means a heavy social and economic burden, making this disease the focus of several epidemiological studies [1].

The currently recommended best practices for the treatment of acute stroke involves in several cases thrombolysis and thrombectomia procedures, for which specialized *stroke units* are required at the care providers. Time is a crucial factor of the treatment as these procedures can only be applied during the first ca. 4.5 hours after onset according to the relevant medical protocols [2]. After the acute stroke has been diagnosed, the patient should be transferred to a stroke center as soon as possible. For the precise diagnosis, a Computer Tomography (CT) scan of the skull is normally considered a prerequisite. Due to the proven success of the specialized stroke units [3][4], a country-wide network of stroke centers with stroke units has been developed in most European countries over the past two decades.

Due to the economic and social importance of stroke, the efficiency of the care services is vital and should be monitored, yet there are no widely accepted methods for this purpose. Our study focuses on Hungary, a country with a population of ca. 10 million, where the stroke center network was set up between 2010 and 2012. We do not analyze the clinical outcome (e.g., 30-day survival) of the stroke episodes, rather, we try to identify the typical care patterns of the stroke centers and the other, non-specialized clinics as well. The results presented here build upon our earlier results in this field [5]. Since the different patterns are associated with different costs and procedural risks, such results may be used for the planning of the national stroke care network. The most important original contribution of the paper is the proposed analysis method. To our best knowledge, no similar methodology has been used to date to characterize the stroke healthcare network at the population level. In the most closely related, recent analysis, the authors analyzed acute stroke-care quality for the cases of 74,000 patients in Great Britain, according to the connection between time of the day and the day of the week of the start of the care and the 30-day survival [6]. In our study, the main objective is to identify anomalies in the care system.

This paper is organized as follows. In Section II and III, we present the input data and the analysis methods. The results are stated and visualized in Section IV, and discussed in Section V. Finally, conclusions are drawn.

II. INPUT DATA

Hungary has a single, centralized health care insurance system which makes it possible to gain access to the data of *all* publicly financed cases of the past 20 years in an anonymized Data Store (TEA) of the National Healthcare Services Center. The private domain care volume in this filed is negligible. The case data include patient demography, the start and end date of an inpatient case, the associated International Classification of Diseases (ICD) codes and the codes and time stamps of the performed procedures. This is an administrative data base, so it has some definite shortcomings for clinical analysis: the start time of the case and the time of the CT scans are stored only at date precision and also the ICD and WHO coding practices of the clinics must be understood properly in order to successfully reconstruct the real sequence of events in a case.

For this study, we queried acute ischemic stroke cases from the TEA between 2010 and 2017 (eight full years) which

 ISBN: 978-1-61208-742-9

had main ICD codes of I63 and I66. Since such main codes are often used too liberally for uncertain cases, we discarded cases which did not contain a skull CT scan in the time frame of -1 to 7 days of the inpatient case start date. Thus, we had a total of 281,948 cases belonging to 228,751 patients over the 8-year period.

III. Methods

The basic methodology of care pattern analysis is based on forming and classifying 'episodes' containing care events and computing a 'care spectrum' for each provider, and it was originally proposed for another care domain [7]. The 'spectrum' contains the relative occurrence ratios of the various episode types and our basic assumption is that if these ratios are similar for two providers, then their care practice is also similar. Finally, the overall stroke care system can be assessed by finding groups of 'similar' providers at the national, i.e., population level.

A. *Data cleaning*

Since the TEA data was collected for financial reimbursement purposes, first we had to go through an elaborate data cleaning process:

1) Patients for whom the gender, age, or residence data was missing or unclear, were excluded from further analysis.

2) Since we wanted to assess the care profile of the care providers, we excluded the cases and the individual procedures if the provider could not be identified from the TEA records.

3) Based on the case data, we created a list of care events for each patient. We had only four possible event types, the CT event, the thrombolysis event (TL), the thrombectomy event (TE) and the simple care event (C), i.e., an event in which the status of the patient is assessed by a medical professional without imaging. This may happen for example when the doctor decides to refer the patient further to another clinic, possibly a stroke center.

For more details on data cleaning, please see the study by Vassányi et al. [5].

B. *Creating and classifying episodes*

An 'episode' means as all care activities related to a new acute stroke occurrence. Technically, we prescribed that an episode must be preceded by at least one event-free day and it may not last longer than two days (five days only in case of repeated TL/TE procedures). Since the transfer of an acute patient from a provider to another involves a time delay and a medical risk, we distinguished episodes involving a transfer from those that do not. We defined five episode types as shown in Table I. The table also shows the number of episodes of the respective types in bold face.

C. *Clustering the providers*

We created clusters from the providers as follows.

1) We assigned each episode to a single clinic, which in the case of the type 4 and 5 episodes, was the first clinic, i.e., the clinic where no procedure was performed.

2) We created a profile for each clinic. The profile is a template that consists of five numbers which are the frequencies of the clinic's episodes of the above five episode types.

TABLE I. EPISODE TYPES

Type Code	Event sequence
1	An episode without CT, with no further referral to another provider and no procedure performed. This may be due to a light stroke that requires no further clinical care or a late delivery to the clinic (over the 6-hour time window). Number of Type 1 episodes: **21,983**
2	A CT was performed, but there was no TE or TL and the patient was not transferred to another clinic. The reason for this may be a not so severe stroke or a late delivery (over the 6-hour time window). Number of Type 2 episodes: **220,061**
3	A CT and a TE or TL was performed, and the patient was not transferred to another clinic. The typical case for this type is a severe stroke patient that was delivered straight to a stroke clinic, within 6 hours from onset. Number of Type 3 episodes: **10,794**
4	The patient was delivered to a clinic, where no CT or TL/TE was performed, then the patient was transferred to another clinic (in most cases, a stroke center) where a CT and TL/TE was performed (the TL/TE is optional). This is a less favorable scenario, because the patient should probably have been transferred straight to the second provider. Number of Type 4 episodes: **15,678**
5	Same as Type 4 with the difference that a CT was taken at the first clinic. Such an episode type may emerge from a case when the CT suggests a TL/TE procedure, but first clinic has no such facilities and they are still within 6 hours. Number of Type 5 episodes: **2,182**

3) We used Pearson's correlation as a measure of similarity between any two pairs of clinics, to build up a symmetrical correlation matrix in which the $element_{i,j}$ is the correlation between the *i*-th and *j*-th clinic's profile.

4) We created a network using this correlation matrix in which the nodes are the clinics and the edge weights are the linearly transformed matrix elements. Thus, two clinics following a 'similar' clinical care practice were connected with an edge of a heavy weight.

5) We used the Louvain network clustering algorithm based on modularity functions to identify significantly strongly connected sub-networks, i.e., clusters of clinics, of the above network [8].

6) In order to characterize the spatial distribution patterns of the episode types, we also used the clustering process described in the steps 3-5 above to cluster the postal code (ZIP) areas. Since the patient demography data contained the postal area code of the patient at the time of onset, we could compute the profile for each area from the episode types of the area's patients. For that, we used the number of episode types per 1000 inhabitants of the area.

For more algorithmic details on steps 3, 4 and 5, please refer to the study of Vassy et al. [7].

 ISBN: 978-1-61208-742-9

D. *Software tools used for the analysis*

The data cleaning and provider profiling was performed using MS SQL Server 2014 database server. For the Louvain clustering we used the Modularity Optimizer tool, version 1.2.0. Statistical pre- and post-processing was implemented with the R 3.1.1 tool. Data visualization for heat maps was created with the SeaBorn package in Python.

IV. RESULTS

At the end of the data cleaning process we had 904,089 events, out of which there were 617,276 CT, 274,292 C, 11,813 TL and 708 TE type events.

The total number of episodes is 270,698 with an average of 2.09 events per episode. Table I shows the number of episodes of the respective types.

There were 37 stroke centers in operation in Hungary in the analyzed period. Considering only the 37 dedicated stroke centers, the clustering procedure did not find significantly different clusters. When we considered all 61 clinics that have a total episode count over 200, two distinct clusters were identified:

1) A small cluster of 9 clinics, of which there was only one specialized stroke clinic;

2) A large cluster of the rest 52 clinics.

The frequencies of the five episode types, as well as the average number of episodes for the two clusters are shown in Table II.

In order to visualize the homogeneity of the two clusters, Figure 2 shows a graphical visualization of the correlation matrix in the form of a 'heat map' in which the correlations are color-coded according to the color ramp key at the left side, light shades meaning strong correlation. The 9 clinics of the small cluster are located in the first 9 rows of the symmetrical matrix.

We also performed the clustering with slightly different algorithmic parameters of the Modularity tool and with another clustering methods, but there was no significant change in the cluster assignments, showing that the clustering method is quite robust.

The clustering of the 2,637 postal code areas also resulted in two clusters. The cluster parameters are shown in the columns 'Cl. 1 (all)' and 'Cl. 2 (all)' of Table III. Cluster 1 had a very high intra-cluster average correlation of 0.97.

TABLE II. CLINIC CLUSTER FEATURES

Cluster feature	Cluster 1	Cluster 2
Number of clinics	9	52
Avg. ratio of Type 1 episodes	0.13	0.08
Avg. ratio of Type 2 episodes	0.37	0.83
Avg. ratio of Type 3 episodes	0.003	0.043
Avg. ratio of Type 4 episodes	0.50	0.04
Avg. ratio of Type 5 episodes	0.004	0.013
Avg. No. of episodes per clinic	1,118.44	3,284.19
St. dev. of the No. of episodes	831.74	1,939.32

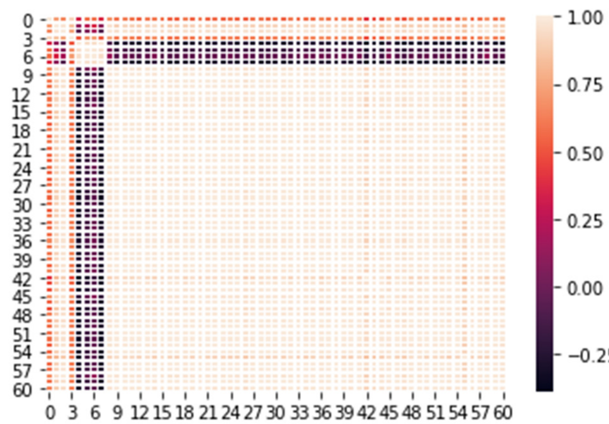

Figure 1. Heat map of the two provider clusters

When we excluded the regions with less than 1000 inhabitants, we had very similar results that are shown in the in the columns 'Cl. 1 <1000' and 'Cl. 2 <1000' of the same table. The table shows the number of episodes / 1000 inhabitants in the areas belonging to the cluster.

TABLE III. POSTAL CODE AREA CLUSTER FEATURES

Cluster feature	Cl. 1 (all)	Cl. 2 (all)	Cl. 1 >1000	Cl. 2 >1000
Number of areas	592	2,045	296	1,173
Avg. # of Type 1/1000 inh.	1.77	1.26	1.60	1.30
Avg. # of Type 2/1000 inh.	13.42	17.91	12.22	16.85
Avg. # of Type 3/1000 inh.	0.72	0.96	0.68	0.89
Avg. # of Type 4/1000 inh.	5.02	0.30	4.01	0.34
Avg. # of Type 5/1000 inh.	0.28	0.22	0.26	0.17

We visualized the spatial location of the areas belonging to the same cluster to see whether they appear randomly at any part of the country or they form homogeneous patches. The resulting two maps are shown in the Figure 2.

V. DISCUSSION

The features of the two clusters of the clinics show quite characteristic differences.

- In general, Cluster 1 is characterized by less valuable clinical services (c.f. the higher ratio of CT-less episodes) and a relative preference of 'forwarding' stroke care to other clinics, proven by an order-of-magnitude difference in the type 4, and also a huge difference in type 5 episode frequencies.
- Cluster 2, containing five sixths of the clinics, is characterized by a relative preference of on-site treatment, with or without a TL/TE procedure. This pattern is more in line with the expected practice according to the protocol.

The very high correlation coefficients within the clusters and the low inter-cluster values (dark shades) in Figure 1 show a marked difference between the two groups.

Since the cause of treating a stroke case without taking an emergency skull CT may be either a too late delivery of the

 ISBN: 978-1-61208-742-9

patient to the clinic or the inability to perform a CT a possible reason for a clinic to belong to Cluster 1 could be a lack of proper imaging infrastructure or skilled staff. A possible interpretation of this cluster is that it is an outlier group as contrasted to the 'normal' majority of the other clinics, but this requires further investigation. On the other side, a high frequency of too late deliveries in the vicinity of a clinic may mean anomalies in logistics or emergency services.

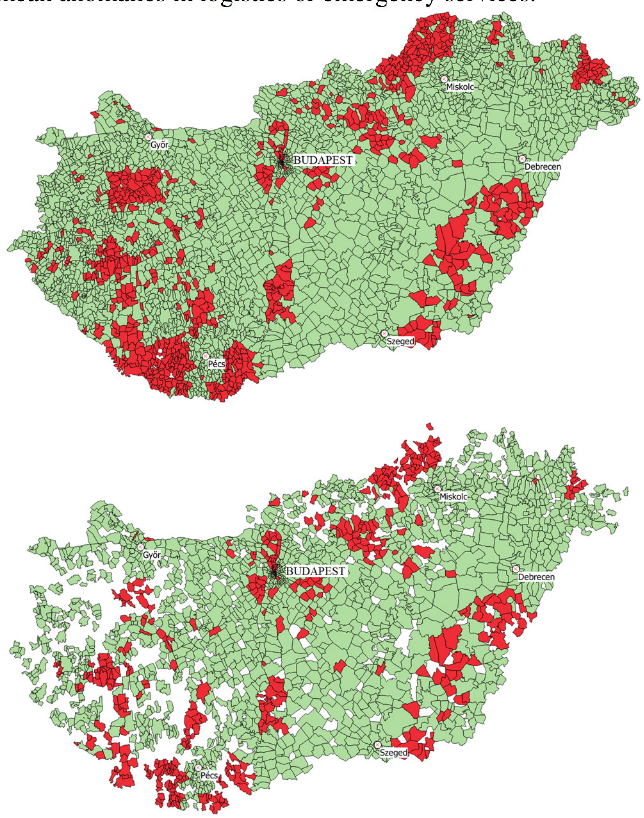

Figure 2. Postal code area clusters for all areas (upper) and without areas with a ppopulation of less than 1000 (lower). Cluster 1: red patches, Cluster 2: light green patches, areas with less than 1000 are white

As Table III shows, the two postal code area clusters, with huge differences in the number of Type 4 episodes, are very similar in nature to the two clusters of the clinics. The visually apparent, quite large red patches in Figure 2 show the regions where stroke services should be reviewed. Since we used no geographical information for the formation of the clusters, any contiguous patches of a cluster mean an anomaly in the care network. In our earlier work, we observed a similar spatial effect in the care patterns of ischemic heart diseases [7].

It should be noted that the ICD coding practices may vary slightly from clinic to clinic, which may add a bias to our survey. In other words, whether or not a case is coded as an acute stroke may depend on the subjective opinion of the medical professional in charge. Such variations, however, cannot be assumed for the CT, TL, and TE procedures, because a procedure has either been performed (and paid for) or not. The clear identification of the *de facto* heterogeneous care practices can be used when planning the care services at the national level.

VI. Conclusion and Future Work

We proposed a comprehensive methodology for the assessment of the stroke care network at the national level. The analysis identified significantly different clusters of both clinics and geographical areas. The anomalies cannot be explained by different coding practices. Hungary, and Central and Eastern Europe in general has yet to reach the Western European quality parameters of stroke treatment [9]. The results presented in the paper provide a clear, fact based starting point to start in the right direction.

Future work includes the analysis of the connection between the episode types and clinical outcomes as well as the associated costs of the treatment.

Acknowledgment

We acknowledge the financial support of Széchenyi 2020 programme under the project No EFOP-3.6.1-16-2016-00015.

References

[1] A, Wutzler et al. "Stroke prevention in patients with acute ischemic stroke and atrial fibrillation in Germany - a cross sectional survey" BMC Neurology, vol. 19(25), https://doi.org/10.1186/s12883-019-1249-y

[2] E, Jauch et al. "Guidelines for the early management of patients with acute ischemic stroke: a guideline for healthcare professionals from the American Heart Association" Stroke, 2013 March, vol. 44(3) pp. 870-947, https://www.ncbi.nlm.nih.gov/pubmed/23370205

[3] S, Morris et al.,"Impact and sustainability of centralising acute stroke services in English metropolitan areas: retrospective analysis of hospital episode statistics and stroke national audit data." BMJ, 2019 Jan; vol. 364:(l1), https://www.bmj.com/content/364/bmj.l1

[4] V, Demarin and H, Budinčević. "Ischemic stroke – the challenge continues" Croat Med J., 2016 Jun; vol. 57(3), pp. 217–218, https://dx.doi.org/10.3325%2Fcmj.2016.57.217

[5] I, Vassányi, T, Kováts, G, Surján, and Z, Nagy. "Analysis of patient pathways in acute stroke care episodes" Proc. Pannonian Conference on Advances in Information Technology (PCIT'2019), Veszprém, Hungary, 31 May - 1 June 2019, ISBN 978-963-396-127-8, pp. 71-76

[6] B, Bray et al. "Weekly variation in health-care quality by day and time of admission: a nationwide, registry-based, prospective cohort study of acute stroke care" The Lancet, 2016 July, vol. 388,(10040), pp. 170-177, https://doi.org/10.1016/S0140-6736(16)30443-3

[7] Z, Vassy, I, Kósa, and I, Vassányi. "Correlation Clustering of Stable Angina Clinical Care Patterns for 506 Thousand Patients" Journal of Healthcare Engineering, vol 2017, Article ID 6937194, https://doi.org/10.1155/2017/6937194

[8] N, Bansal, A, Blum, and S, Chawla. "Correlation Clustering",. Machine Learning (2004) 56: 89, pp.89-113 https://doi.org/10.1023/B:MACH.0000033116.57574.95

[9] The CEESS Working Group. "Management of ischemic stroke in Central and Eastern Europe", Int. J of Stroke, 2015, vol. 10(SA100), pp. 125–127., https://doi.org/10.1111/ijs.12575