

УДК 519.237.5:004.43:303.7:616.28-008.1-53.9
<https://doi.org/10.18692/1810-4800-2021-6-8-19>

Пространственный кластерный анализ в моделировании доступности медицинской помощи пожилым пациентам с нарушениями слуха

А. А. Корнеев¹, Е. А. Левина¹, Е. Э. Вяземская¹, С. В. Левин¹, И. Н. Скирпичников²

¹ Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи, Санкт-Петербург, 190013, Россия

² Челябинская областная клиническая больница, Челябинск, 454048, Россия

Spatial cluster modeling of access of elderly patients with hearing loss to medical services

A. A. Korneenkov¹, E. A. Levina¹, E. E. Vyazemskaya¹, S. V. Levin¹, I. N. Skirpichnikov²

¹ Saint Petersburg Research Institute of Ear, Throat, Nose and Speech, Saint Petersburg, 190013, Russia

² Chelyabinsk Regional Clinical Hospital, Chelyabinsk, 454048, Russia

Доступность и качество медицинской помощи обеспечивается организацией оказания медицинской помощи по принципу приближенности к месту жительства. Увеличение количества медицинских учреждений однозначно повысит доступность медицинской помощи, но также повысит и затраты на ее оказание, снизив таким образом в целом эффективность системы здравоохранения. В статье рассматривается решение задачи оптимизации оказания сурдологической помощи пожилым пациентам с нарушением слуха и глухотой путем нахождения оптимального количества медицинских учреждений, оказывающих эту помощь, и их географического размещения в пределах Челябинской области. Задача решалась на основе минимизации расстояния и времени, затрачиваемого пациентами на проезд до места оказания медицинской помощи, с использованием пространственного кластерного анализа. Продемонстрированы возможности пространственной кластеризации к созданию обоснованных предпосылок для оптимального математически обоснованного решения задач повышения доступности медицинской инфраструктуры для населения. Моделирование доступности медицинской помощи на основе показателя времени в пути до медицинской организации требует использования геоинформационных систем, оценивающих в реальном времени дорожный трафик. Подобные задачи необходимо предусмотреть в случае широкого использования показателей доступности для решения задач организации и оптимизации территориальных систем здравоохранения.

Ключевые слова: доступность медицинской помощи, сурдологическая помощь, кластерный анализ, оптимизация сети медицинских учреждений, пространственный анализ.

Для цитирования: Корнеев А. А., Левина Е. А., Вяземская Е. Э., Левин С. В., Скирпичников И. Н. Пространственный кластерный анализ в моделировании доступности медицинской помощи пожилым пациентам с нарушениями слуха. *Российская оториноларингология*. 2021;20(6):8–19. <https://doi.org/10.18692/1810-4800-2021-6-8-19>

The availability and quality of medical care is ensured by the organization of medical care according to the principle of proximity to the place of residence. The increase in the number of medical institutions will definitely increase the availability of medical care, but it will also increase the cost of providing it, thus reducing the overall efficiency of the health care system. The article discusses the solution to the problem of optimizing the provision of audiological care to elderly patients with hearing impairment and deafness by finding the optimal number of medical institutions providing this assistance and their geographical location within the Chelyabinsk region. The problem was solved on the basis of minimizing the distance and time spent by patients to travel to the place of medical care, using spatial cluster analysis. The possibilities of spatial clustering to

create reasonable prerequisites for the optimal solution of the problems of increasing the availability of medical infrastructure for the population are demonstrated.

Keywords: accessibility of medical care, audiological care, cluster analysis, optimization of a network of medical institutions, spatial analysis.

For citation: Korneenkov A. A., Levina E. A., Vyazemskaya E. E., Levin S. V., Skirpichnikov I. N. Spatial cluster modeling of access of elderly patients with hearing loss to medical services. *Rossiiskaya otorinolaringologiya*. 2021;20(6):8-19. <https://doi.org/10.18692/1810-4800-2021-6-8-19>

Используемые сокращения

НСГ – нарушение слуха и глухота

СМП – специализированная медицинская помощь

Used abbreviations

HLD – hearing loss and deafness

SMC – specialized medical care

Введение

Широкая распространенность болезней, связанных с нарушением слуха и глухотой (НСГ), представляет серьезную проблему для здоровья населения [1] и требует значительных затрат на оказание специализированной медицинской помощи (СМП), реабилитацию и абилитацию. Показатели распространенности среди населения инвалидирующих НСГ увеличились с 1% в 1985 г. до 6,1% в 2018 г. [2]. В возникновении НСГ место проживания человека, его пространственные координаты играют определенную роль, обуславливают интенсивность многих факторов возникновения НСГ и могут рассматриваться как самостоятельный пространственный фактор в эпидемиологии НСГ, а также учитываться в организации оказания СМП. Изучением связи различных аспектов здоровья человека и здравоохранения с пространственными координатами давно занимаются в рамках различных прикладных наук на стыке географии, биологии, медицины и т. п., где с успехом применяются методы пространственной статистики [3, 4].

Общей проблемой в пространственной статистике является определение того, группируются ли объекты в группы, и если да, то в какие группы они группируются [5]. Кластерный анализ используется в построении пространственных моделей в так называемом исследовательском пространственном анализе данных (англ. Exploratory Spatial Data Analysis, ESDA) и преследует две основные цели: 1) распознавание пространственных образов (паттернов) с использованием визуализации, пространственной статистики и геостатистики; 2) формирование гипотез для определения реалистичных и проверяемых интерпретаций географических закономерностей, обнаруженных в рамках достижения первой цели. Пространственный кластерный анализ может играть важную роль в решении различных задач здравоохранения. Он обычно используется в медицинской географии, пространственной эпидемиологии, популяционной генетике, медицинской экологии и многих других областях, но его основные принципы остаются теми же.

Introduction

The widespread prevalence of diseases associated with hearing loss and deafness (HLD) is a serious problem for the health of the population [1] and requires significant costs for the provision of specialized medical care (SMC), rehabilitation, and habilitation. The prevalence rates of disabling HLD among the population increased from 1% in 1985 to 6,1% in 2018 [2]. In the emergence of HLD, the place of residence of a person, his spatial coordinates play a certain role, determine the intensity of many factors in the occurrence of HLD, and can be considered as an independent spatial factor in the epidemiology of HLD, as well as considered in the organization of the provision of SMC. The study of the relationship between various aspects of human health and health care with spatial coordinates has long been engaged in various applied sciences at the intersection of geography, biology, medicine, etc., where the methods of spatial statistics are successfully applied [3, 4].

A common problem in spatial statistics is determining whether objects are grouped into groups, and if so, which groups they are grouped into [5]. Cluster analysis is used in the construction of spatial models in the so-called exploratory spatial data analysis (ESDA) and has two main goals: 1) recognition of spatial patterns using visualization, spatial statistics, and geostatistics; 2) forming hypotheses to identify realistic and testable interpretations of geographic patterns found in the first goal. Spatial cluster analysis can play an important role in solving various health problems. It is commonly used in medical geography, spatial epidemiology, population genetics, medical ecology, and many other fields, but its basic principles remain the same.

Как известно, согласно п. 1 статьи 10 Федерального закона от 21.11.2011 № 323-ФЗ (ред. от 02.07.2021) «Об основах охраны здоровья граждан в Российской Федерации» доступность и качество медицинской помощи обеспечиваются организацией оказания медицинской помощи по принципу приближенности к месту жительства, месту работы или обучения. Однако надо учитывать, что географическая близость не всегда означает малое время для проезда до медицинской организации: на это также оказывают влияние развитость транспортной сети, интенсивность дорожного трафика, наличие и достаточность общественного транспорта, расписание движения и т. д. Увеличение количества медицинских организаций однозначно повысит доступность СМП, но также повысит и затраты на ее оказание, снизив таким образом в целом эффективность системы здравоохранения. Для решения подобных оптимизационных задач с успехом применяются современные информационные системы, которые учитывают различные ограничения, факторы и позволяют создавать модели на основе исходных данных о местоположении потребителей СМП, факторах времени и расстояния и возможностях оказания СМП.

Цель исследования

Решение задачи оптимизации оказания сурдологической помощи пациентам с нарушением слуха и глухотой (НСГ) путем нахождения оптимального количества медицинских центров, оказывающих эту помощь, и их географического размещения в пределах Челябинской области. Задача решалась на основе минимизации расстояния и времени, затрачиваемого пациентами на проезд до места оказания медицинской помощи, с использованием пространственного кластерного анализа.

Материалы и методы исследования

Ретроспективное поперечное обсервационное исследование было проведено с использованием обезличенных данных пациентов, обратившихся в период с 2016 по 2019 г. в областной сурдологопедический центр на базе Челябинской областной клинической больницы (далее – сурдоцентр).

Для обработки и визуализации пространственных данных были использованы различные библиотеки (ggplot2, rdal и другие) программного пакета R, который хорошо себя зарекомендовал для решения широкого круга статистических задач [3, 4, 6]. Пространственные данные об административно-территориальном делении Челябинской области (полигоны муниципальных округов, районов, дорог и т. п.) получены из OpenStreetMap (сокращенно OSM) – некоммерче-

As is commonly known, following clause 1 of article 10 of the Federal Law of November 21, 2011, No. 323-FZ (as amended on July 2, 2021) „On fundamental healthcare principles in the Russian Federation“, the availability and quality of medical care are ensured by the organization of medical care according to the principle of proximity to the place of residence, place of work or study. However, it should be borne in mind that geographical proximity does not always mean a short travel time to a medical institution: this is also influenced by the development of the transport network, the intensity of road traffic, the availability, and the sufficiency of public transport, the timetable, etc. An increase in the number of medical organizations will increase the accessibility of SMC, but will also increase the cost of its provision, thus reducing the overall efficiency of the health care system. To solve such optimization problems, modern information systems are successfully used, which consider various constraints, factors and make it possible to create models based on the initial data on the location of SMC consumers, time and distance factors, and the possibilities of SMC provision.

Objectives

Solving the problem of optimizing the provision of audiological care to patients with HLD by finding the optimal number of medical centers providing this assistance and their geographical location within Chelyabinsk region. The problem was solved based on minimizing the distance and time spent by patients on travel to the place of medical care, using spatial cluster analysis.

Materials and research methods

A retrospective cross-sectional observational study was carried out using anonymized data of patients who applied in the period from 2016 to 2019 at the regional audiological center based on Chelyabinsk regional clinical hospital (hereinafter, the audiology center).

For processing and visualization of spatial data, various libraries (ggplot2, rdal, and others) of the R software package were used, which has proven itself well for solving a wide range of statistical problems [3, 4, 6]. Spatial data on the administrative-territorial division of Chelyabinsk region (polygons of municipal and rural districts, roads, etc.) were obtained from OpenStreetMap (abbreviated OSM) – a non-commercial web mapping project using the Overpass Turbo service (<http://>

ского веб-картографического проекта с помощью сервиса Overpass Turbo (<http://overpass-turbo.eu/>) и специализированных программных библиотек R.

Результаты исследования

Челябинская область является субъектом Российской Федерации, входит в состав Уральского федерального округа, административный центр – город Челябинск. Площадь Челябинской области равна 88,5 тысячи квадратных километров, протяженность области с севера на юг – 490 км, с запада на восток – 400 км. Географический центр области располагается в более ста километрах к юго-западу от Челябинска. Челябинская область по территории занимает 39-е место по России. Согласно Закону «Об административно-территориальном устройстве Челябинской области» [7] и Уставу Челябинской области, субъект РФ включает следующие административно-территориальные единицы: 15 городов областного значения, 15 городов районного значения, 27 районов; 13 поселков городского типа, 242 сельсовета. Общая протяженность автодорог – 18 766 км.

Система здравоохранения Челябинской области представлена различными медицинскими организациями, относящимися к государственным, муниципальным учреждениям здравоохранения, учреждениям федерального подчинения (медицинские организации, находящиеся в ведении Минздрава России, органов исполнительной власти субъектов Российской Федерации в сфере охраны здоровья и муниципальных образований), оказывающим помощь в амбулаторных и стационарных условиях.

В настоящее время пациенты с НСГ, проживающие на территории Челябинской области, получают СМП в сурдоцентре областной клинической больницы Челябинска, при котором создан специализированный регистр пациентов с НСГ. Однако, учитывая большую протяженность Челябинской области, максимальное время в пути пациента до сурдоцентра, даже при использовании личного автотранспорта, может достигать трех и более часов, а при использовании общественного транспорта это время может значительно увеличиться. В исследовании использованы данные пациентов с НСГ (точнее, сенсоневральной тугоухостью) в возрастной группе пожилых старше 60 лет. Как известно, мобильность пациентов этой возрастной группы значительно снижена по самым разным причинам, что уменьшает доступность медицинских организаций для получения СМП. В период с 2016 по 2019 г. в сурдоцентр обратилось 3446 человек с сенсоневральной тугоухостью и глухотой в возрасте старше 60 лет, т. е. в среднем более 800 первичных обращений в год. Из общего числа пациентов случайным образом (точнее, псевдо-

overpass-turbo.eu/) and specialized software libraries R.

Research results

Chelyabinsk region is a constituent entity of the Russian Federation, it is part of the Ural Federal District, the administrative center is the city of Chelyabinsk. The area of Chelyabinsk region is 88.5 thousand square kilometers, the length of the region from north to south is 490 km, from west to east – 400 km. The geographic center of the region is located more than one hundred kilometers southwest of Chelyabinsk. Chelyabinsk region occupies 39th place in Russia in terms of territory. According to the Law „On the administrative-territorial structure of Chelyabinsk region“ [7] and the Charter of Chelyabinsk region, the subject of the Russian Federation includes the following administrative-territorial units: 15 cities of regional significance, 15 cities of district significance, 27 districts; 13 urban-type settlements, 242 village councils. The total length of highways is 18,766 km.

The health care system of Chelyabinsk region is represented by various medical organizations related to state, municipal health care institutions, federal institutions (medical organizations under the jurisdiction of the Ministry of Health of Russia, executive bodies of the constituent entities of the Russian Federation in the field of health care and municipalities), aiding in outpatient and inpatient conditions.

Currently, patients with HLD living on the territory of Chelyabinsk region receive SMC in the audiology center of the regional clinical hospital in Chelyabinsk, where a specialized register of patients with HLD has been created. However, given the large length of Chelyabinsk region, the maximum travel time of a patient to the audiology center, even when using personal vehicles, can reach three or more hours, and when using public transport, this time can increase significantly. The study used data from patients with HLD (more precisely, sensorineural hearing loss) in the age group of the elderly, over 60 years old. As you know, the mobility of patients in this age group is significantly reduced for a variety of reasons, which reduces the availability of medical organizations for receiving SMC. In the period from 2016 to 2019, 3446 people with HLD over the age of 60 applied to the audiology center, i.e. on average, more than 800 initial calls per year. From the total number of patients, a sample of 100 people was selected at random (more precisely, pseudo-randomly, using the sample () function of the R language), for which descriptive and test statistics were calculated. Such a

случайным, с помощью функции `sample()` языка R) была отобрана выборка из 100 человек, для которых и были рассчитаны описательные и тестовые статистики. Такая выборка позволила сократить машинное время выполнения расчетов среднего времени в пути без ощутимой потери точности.

Общая схема исследования включала несколько этапов:

1) статистическое оценивание времени в пути пациентов с НСГ до сурдоцентра Челябинской области на основе данных OpenStreetMap (OSM), включая вычисление описательных статистик и оценку распределения времени пути;

2) поиск оптимального числа возможных значений кластеров и расчетных центров (центроидов) кластеризации пациентов, используя иерархический кластерный анализ;

3) иерархическая кластеризация пожилых пациентов с НСГ в пределах Челябинской области; поиск реальных центров (городских больниц и т. п.), ближайших к рассчитанным центроидами;

4) визуализация оптимального числа кластеров, нахождение времени в пути между местоположением пациентов с НСГ и возможными реальными центрами, представление изохронов и гистограмм времени в пути.

Статистическое оценивание времени в пути пациентов с НСГ до сурдоцентра Челябинской области

Время проезда пациента до сурдологического центра моделировалось с помощью бесплатного сетевого сервиса Open Source Routing Machine, или OSRM, сочетающего в себе сложные алгоритмы маршрутизации в дорожных сетях с открытыми и бесплатными данными дорожной сети проекта OpenStreetMap (OSM). Определялось время проезда на автомобиле (по сути – минимальное, базовое время) как время проезда на автотранспорте без учета пробок и других факторов, увеличивающих время в пути. Среднее время в пути, по данным OSRM, при существующей системе оказания СМП на базе одного сурдоцентра составило $97,3 \pm 12,8$ мин (95%-ный доверительный интервал) (медиана $M_n = 82,2$ мин, $\min = 3,4$ мин, $\max = 334,7$ мин, среднее квадратическое отклонение $SD = 64,8$ мин).

Поиск оптимального числа возможных значений кластеров и центров (центроидов) кластеризации пациентов

Ключевой проблемой в кластерном анализе является поиск наилучшего, оптимального, числа кластеров в наборе данных. Для оптимизации числа кластеров могут быть использованы разные методы, чаще всего связанные с перебором возможных значений показателя эффективности кластеризации (например, внутрикластерных расстояний) при разном числе заданных

sample made it possible to reduce the machine time for performing calculations of the average travel time without a noticeable loss of accuracy.

The general research scheme included several stages:

1) statistical estimation of the travel time of patients with HLD to the audiology center of Chelyabinsk region based on OpenStreetMap (OSM) data, including the calculation of descriptive statistics and an estimate of the travel time distribution;

2) search for the optimal number of possible values of clusters and calculated centers (centroids) for clustering patients using hierarchical cluster analysis;

3) hierarchical clustering of elderly patients with HLD within Chelyabinsk region. Search for real centers (city hospitals, etc.) closest to the calculated centroids;

4) visualization of the optimal number of clusters, finding the travel time between the location of patients with HLD and possible real centers, presentation of isochrones, and histograms of travel time.

A statistical estimation of the travel time of patients with HLD to the audiology center of Chelyabinsk region

The patient's travel times to the audiology center were modeled using the free Open Source Routing Machine, or OSRM, which combines sophisticated road routing algorithms with open and free road network data from the OpenStreetMap (OSM) project. The travel time by car (in fact, the minimum, basic time) was determined as the travel time by car without regard to traffic jams and other factors that increase the travel time. The average travel time according to OSRM data with the existing SMC delivery system based on one audiology center was 97.3 ± 12.8 minutes (95% confidence interval) (median $M_n = 82.2$ minutes, $\min = 3.4$ minutes, $\max = 334.7$ minutes, standard deviation $SD = 64.8$ minutes).

Search for the optimal number of possible values of clusters and centers (centroids) of clustering patients

The key problem in cluster analysis is finding the „best”, optimal number of clusters in a dataset. To optimize the number of clusters, different methods can be used, most often associated with enumerating possible values of the clustering efficiency indicator (for example, intra-cluster distances) for a different number of specified clusters. Considering that with an increase in the number of clusters, the differences

кластеров. Учитывая, что при увеличении числа кластеров различия между ними размываются, затрудняя их интерпретацию в рамках задач исследования, их количество редко превышает 5–7.

Методы иерархической кластеризации как совокупность алгоритмов упорядочивания данных, направленных на создание иерархии (дерева) вложенных кластеров, включают два класса методов иерархической кластеризации: агломеративные методы (англ. *agglomerative*), когда новые кластеры создаются путем объединения более мелких кластеров, и дивизивные, или дивизионные, методы (англ. *divisive*), когда новые кластеры создаются путем деления более крупных кластеров на более мелкие. В исследовании использовался агломеративный алгоритм иерархической кластеризации на основе метода полной связи (*complete linkage*), который определяет кластерное расстояние между двумя кластерами как максимальное расстояние между их отдельными компонентами. На каждом этапе итеративного процесса кластеризации два ближайших кластера объединяются в новый кластер, пока весь набор данных не будет агломератирован в один кластер.

Визуальным представлением иерархии или кластерного дерева является дендрограмма, которая определяет уровень сходства между парами кластеров. Дендрограмма строится по матрице заданных заранее мер сходства (или различия). В пространственной кластеризации в качестве мер сходства используется матрица расстояний, полученная разными алгоритмами, в зависимости от того, какая форма земли закладывается в алгоритм: эллипсоид, сфера и т. п.

Определенную сложность представляет кластеризация в тех случаях, когда заданы границы области, в которой производится кластеризация, и расстояние между точками определяется расстоянием не по прямой, а по дорожной сети. Например, между двумя близкими расположенными населенными пунктами одной административной области может находиться участок другой области. Алгоритм кластеризации на основе матрицы расстояний без учета этих особенностей отнесет эти два населенных пункта, скорее всего, к одному кластеру, однако в реальности, исходя из принципа обеспечения доступности СМП, они будут принадлежать разным зонам медицинского обслуживания и прикреплены к разным медицинским организациям.

Время в пути до медицинской организации – один из параметров доступности СМП. В исследовании именно этот показатель используется для оценки эффективности кластеризации локаций пожилых пациентов с НСГ. Если расчеты матрицы расстояний между локациями-объектами (в нашем случае пациентами) по эллипсоиду не представляет сложности, то вычисление матрицы

between them become blurred, making it difficult to interpret them within the framework of the research objectives, their number rarely exceeds 5–7.

Hierarchical clustering methods, as a set of data ordering algorithms aimed at creating a hierarchy (tree) of nested clusters, include two classes of hierarchical clustering methods: agglomerative methods, when new clusters are created by combining smaller clusters and divisional methods when new clusters are created by dividing larger clusters into smaller ones. The study used an agglomerative hierarchical clustering algorithm based on the complete linkage method, which defines the cluster distance between two clusters as the maximum distance between their components. At each stage of the iterative clustering process, the two nearest clusters are combined into a new cluster until the entire dataset is agglomerated into a single cluster.

A visual representation of a hierarchy or cluster tree is a dendrogram that determines the level of similarity between pairs of clusters. The dendrogram is constructed from a matrix of predetermined measures of similarity (or difference). In spatial clustering, a matrix of distances obtained by different algorithms is used as measures of similarity, depending on which shape of the earth is put into the algorithm: an ellipsoid, a sphere, etc.

A certain difficulty is presented by clustering in those cases when the boundaries of the area in which the clustering is performed are set, and the distance between the points is determined by the distance not along a straight line, but the road network. For example, between two closely located settlements of one administrative region, there may be a plot of another region. The clustering algorithm based on the distance matrix, without regard to these features, will assign these two settlements, most likely to one cluster, but in reality, based on the principle of ensuring the accessibility of SMC, they will belong to different areas of medical care and are attached to different medical organizations.

Travel time to a medical organization is one of the parameters of the availability of SMC. In the study, it is this indicator that is used to assess the effectiveness of clustering the locations of elderly patients with HLD. If calculating the matrix of distances between locations of objects (in our case, patients) along an ellipsoid is not difficult, then calculating the matrix of travel time between locations requires queries to constantly

времени в пути между локациями требует запросов к постоянно обновляемым ресурсам геоинформационных систем Google, Yandex, OSM и пр. Так как распространенность болезни, как в случае с оптимистичной оценкой инвалидизирующих НСГ в 6,1%, может составлять примерно 210 тысяч локаций только для Челябинской области, выполнение такого запроса для двумерной матрицы (2100002) через обычный доступ к геоинформационной системе может быть достаточно длительным и затратным. В связи с этим в исследовании использовалось компромиссное решение, когда начальный поиск центроидов кластеров проводился по матрице расстояний между местоположением пациентов по эллипсоиду, а оценка эффективности кластеризации проводилась по времени пациента в пути и расстоянию внутри уже определенных кластеров по запросам к ГИС. Чтобы еще больше сократить количество запросов к ГИС без ощутимой потери точности, из локаций пациентов была отобрана случайная выборка, для элементов которой и были запрошены расстояния и время по дорожной сети до определенных центроидов.

Для проверки гипотезы о равенстве времени в пути пациентов до сурдоцентра для разного числа кластеров применялся статистический критерий Краскела – Уоллиса (Kruskal – Wallis). По результатам расчетов нулевая гипотеза была опровергнута (уровень значимости $p\text{-value} = 0,00079$). Поэтому можно сделать вывод, что кластеризация оказывает статистически значимое ($p < 0,05$) влияние на время проезда пациентов с НСГ до медицинской организации. Для уточнения того, какое количество кластеров (сурдоцентров) статистически значимо изменит время в пути, был проведен статистический тест на основе критерия Вилкоксона (Wilcoxon). На рис. 1 представлены диаграмма размаха или «ящик с усами» (box plot) времени в пути до сурдоцентра и значения $p\text{-value}$ для попарных сравнений между сетями с разным количеством сурдоцентров. В верхней части диаграммы горизонтальные линии соединяют попарно сравниваемые группы. Число на горизонтальных линиях – уровень значимости p по парному критерию Вилкоксона (Wilcoxon). Как видно на диаграмме (рис. 1), увеличение числа кластеров до трех включительно приводит к статистически значимому уменьшению времени пути до сурдоцентра, однако начиная с трех кластеров статистически значимых различий не выявляется ($p > 0,05$).

Иерархическая кластеризация пожилых пациентов с НСГ в пределах Челябинской области

Как видно на рис. 1, оптимальным числом кластера в этом контексте можно считать три кластера. При такой кластеризации все пациенты могут быть отнесены к разным кластерам таким обра-

updated resources of geographic information systems Google, Yandex, OSM, etc. Since the prevalence of the disease, as in the case of an optimistic estimate of disabling HLD of 6.1%, it can be about 210 thousand locations only for Chelyabinsk region, fulfilling such a request for a two-dimensional matrix (2100002) through ordinary access to a geographic information system can be quite lengthy and costly. In this regard, the study used a compromise solution, when the initial search for cluster centroids was carried out using a matrix of distances between the locations of patients along the ellipsoid, and the assessment of the effectiveness of clustering was carried out according to the patient's travel time and the distance within already defined clusters by GIS queries. To further reduce the number of requests to GIS without a noticeable loss of accuracy, a random sample was selected from the patient locations, for the elements of which the distances and time along with the road network to certain centroids were requested.

The Kruskal – Wallis statistical test was used to test the hypothesis about the equality of the travel time of patients to the audiology center for different numbers of clusters. Based on the calculation results, the null hypothesis was refuted (significance level $p\text{-value} = 0.00079$). Therefore, we can conclude that clustering has a statistically significant ($p < 0.05$) effect on the travel time of patients with HLD to a medical institution. To clarify how many clusters (audiology centers) will statistically significantly change travel time, a statistical test was performed based on the Wilcoxon test. The figure shows a box plot of travel time to the audiology center and $p\text{-value}$ for pairwise comparisons between networks with different numbers of audiology centers. At the top of the diagram, horizontal lines connect the compared groups in pairs. The number on the horizontal lines is the p significance level according to the paired Wilcoxon test. As can be seen in the diagram (Figure 1), an increase in the number of clusters up to three inclusive leads to a statistically significant decrease in the travel time to the audiology center, but starting from three clusters, no statistically significant differences are detected ($p > 0.05$).

Hierarchical clustering of elderly patients with HLD within Chelyabinsk region

As can be seen in the previous Figure 1, the optimal number of a cluster in this context can be considered three clusters. With this clustering, all patients can be assigned to different clusters in the

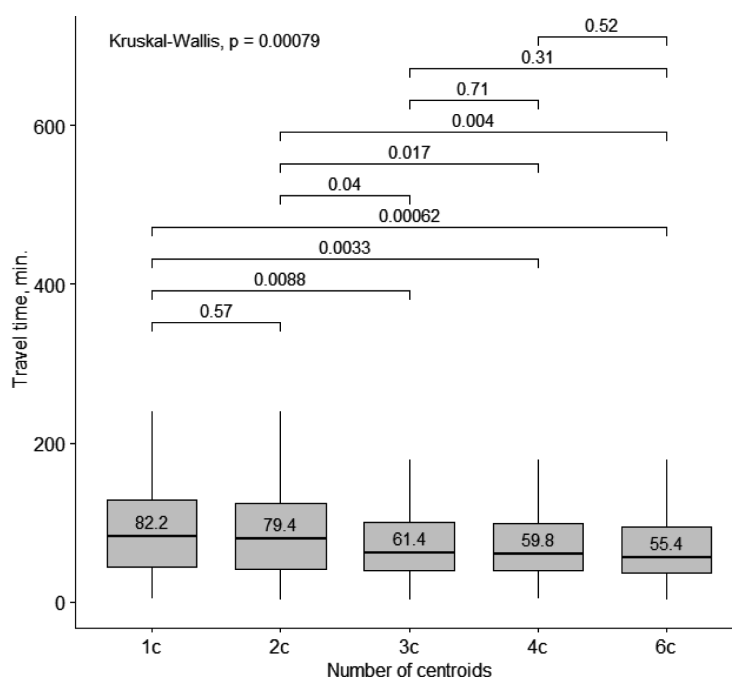


Рис. 1. Диаграмма размаха показателя времени в пути при различном числе сурдоцентров в специализированной сети (границы ящика – первый и третий квартили (25-й и 75-й процентиля соответственно); линия в середине ящика – медиана (50-й процентиль))

Fig. 1. Diagram of the range of travel time for a different number of medical facilities in a specialized network. (Box borders are the first and third quartiles (25th and 75th percentiles, respectively); the line in the middle of the box is the median (50th percentile))

зом, как показано на рис. 2. Наибольшее (67,9%) число пациентов относится к кластеру № 1, к кластеру № 2 – 19,2% пациентов, 12,9% пациентов отнесено к кластеру № 3.

Визуализация оптимального числа кластеров, нахождение времени в пути между местоположением пациентов с НСГ и возможными реальными центрами, представление изохронов и гистограмм времени в пути

Одним из наиболее удобных и наглядных способов отображения зон доступности по времени может служить использование изохрон (от гр. ἴσο – похожий и χρόνος – время) – линий одновременности, ограничивающих зоны с одинаковым временем в пути до интересующих географических точек, объектов, медицинских организаций [8]. Изохроны используются для анализа пространственных данных, например, для расчета охвата – сколько людей проживает или работает в доступности от заданного объекта (например, сурдоцентра).

На рис. 3 показаны изохроны с шагом в 20 мин, показывающие зоны временной доступности СМП в расчетных местоположениях сурдоцентров, соответствующих трем оптимальным кластерам – от фиолетового цвета (наибольшая временная доступность до желтого (наименьшая – от 100 до 120 мин)). Все остальные пациенты проживают в местах, где время проезда на машине занимает более 2 ч (120 мин).

way shown in the following Figure 2. The largest (67,9%) number of patients belongs to cluster No. 1, to cluster No. 2 – 19,2% of patients and 12.9% of patients belong to cluster No. 3

Visualization of the optimal number of clusters, finding the travel time between the location of patients with HLD and possible real centers, presentation of isochrones, and histograms of travel time

One of the most convenient and visual ways of displaying accessibility zones in time can be the use of isochrones (from the Greek ἴσο – similar and χρόνος – time) – lines of simultaneity, limiting zones with the same travel time to interesting geographic points, objects, medical organizations [8]. Isochrones are used to analyze spatial data, for example, to calculate the coverage – how many people live or work within reach of a given object (for example, an audiology center).

Figure 3 shows isochrones with a step of 20 min., these are the zones of temporary accessibility of the SMC in the calculated locations of the auditory centers corresponding to three optimal clusters (from purple (the highest temporal availability) to yellow (the least is from 100 to 120 minutes)). All other patients live in places where travel time by car takes more than 2 hours (120 minutes).

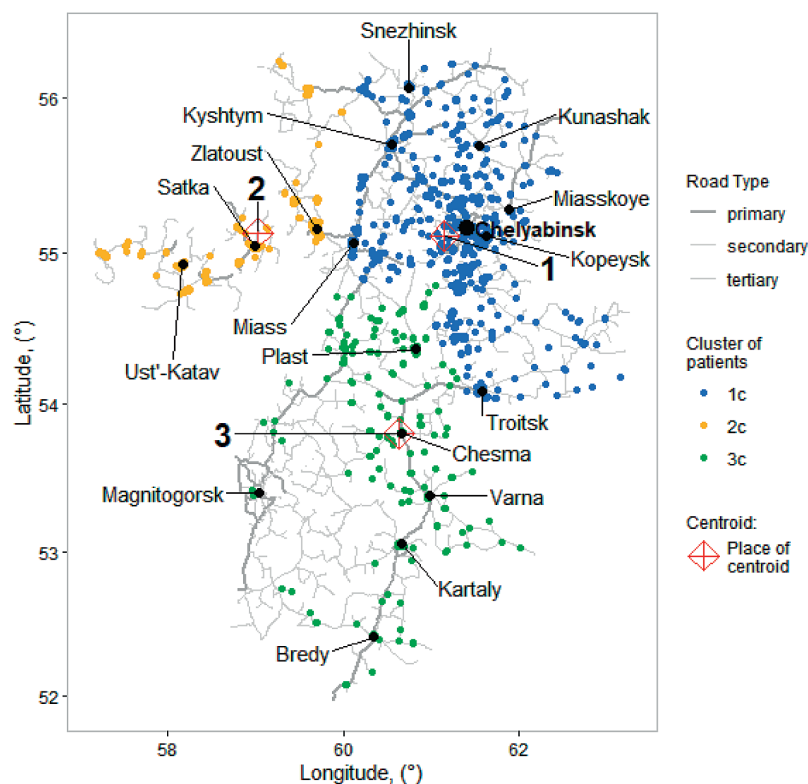


Рис. 2. Распределение локаций пожилых пациентов с НСГ по трем кластерам (обозначены: 1с, 2с, 3с), указаны вычисленные центроиды кластеров и дорожная сеть Челябинской области

Fig. 2. Distribution of locations of elderly patients with hearing loss in three clusters (designated: 1c, 2c, 3c). The calculated centroids in each cluster are indicated; road network of the Chelyabinsk region

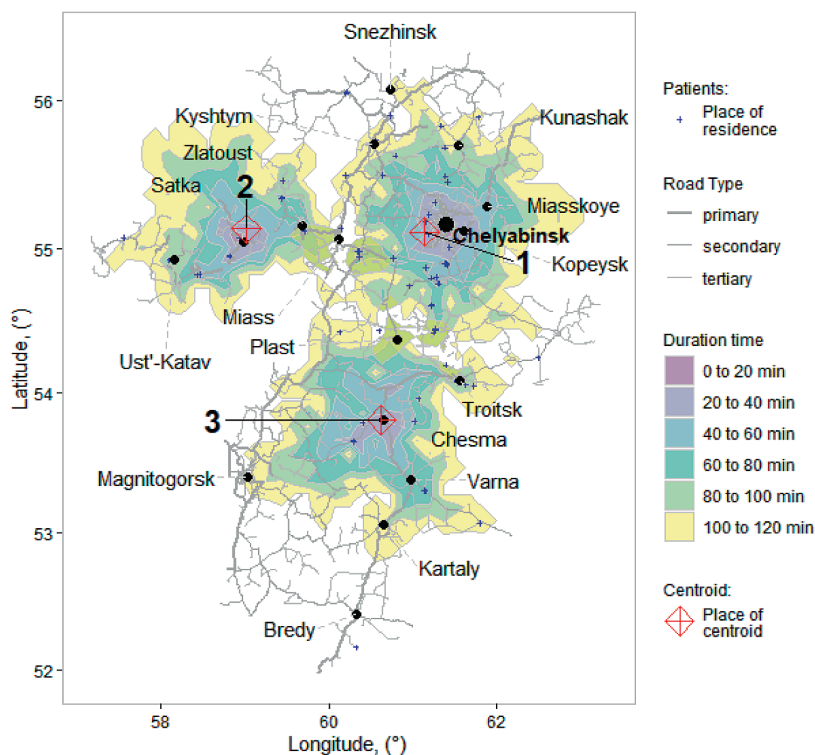


Рис. 3. Изохроны временной доступности сурдоцентров, соответствующих трем кластерам

Fig. 3. Isochrones of temporary availability of medical centers corresponding to three clusters

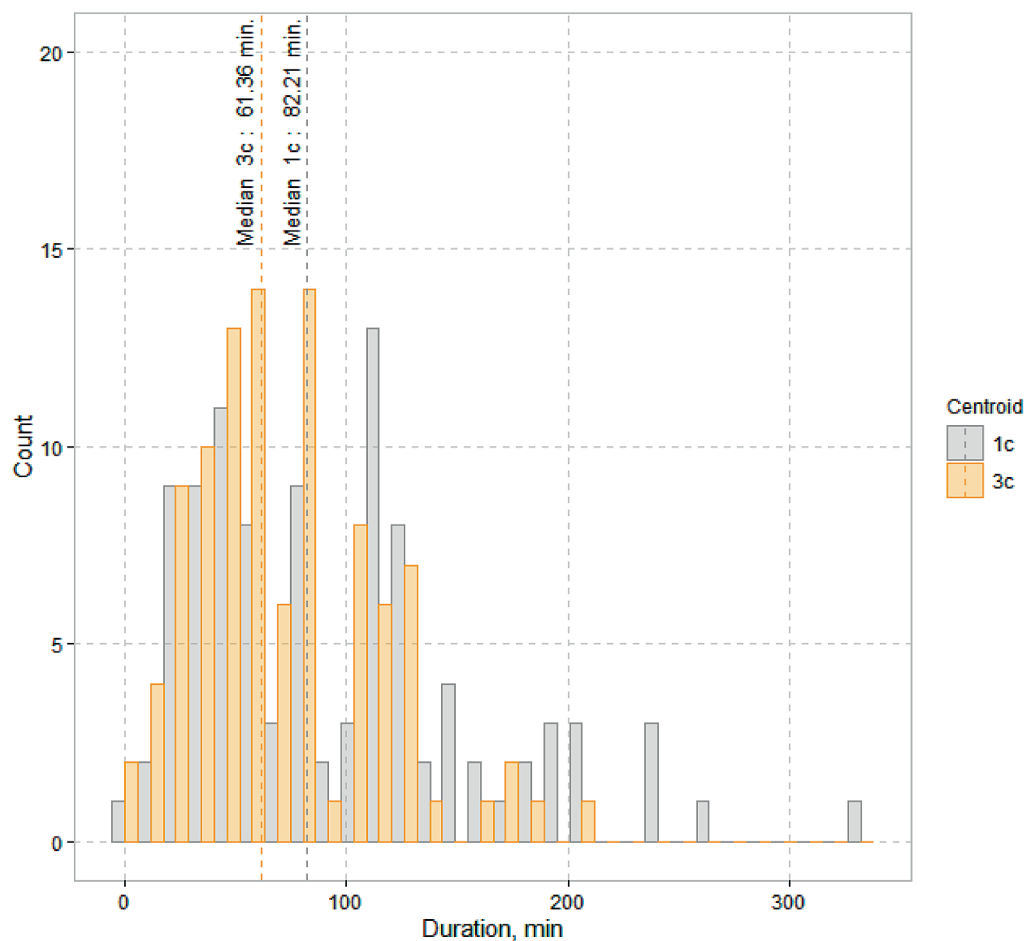


Рис. 4. Гистограмма распределения времени в пути при организации сурдологической сети из одного и из трех сурдоцентров (т. е. один и три кластера)

Fig. 4. Histogram of travel time distribution when organizing a network of medical centers: from one and three medical centers (i.e. one 1 c and three clusters 3 c)

Использование кластеризации уменьшает оценку среднего времени в пути с 97,3 мин (медиана – 82,21 мин) до 70,7 мин (медиана – 61,4 мин). На рис. 4 представлены гистограммы распределения времени в пути при организации сурдологической сети из одного и из трех сурдоцентров. В качестве эффекта [6] снижения времени в пути до сурдоцентра можно рассматривать разность средних (MD) – 26,6 мин (т. е. снижение времени в пути почти на треть от исходного среднего) и стандартизованную разность средних (SMD) – 0,5, значение которой можно оценить как «средний» размер эффекта.

Однако необходимо обратить внимание на еще одну важную характеристику – уменьшение колеблемости времени в пути: показатель размаха времени в пути уменьшился с 331,3 до 200,3 мин, а максимальное значение времени в пути уменьшилось с 334,7 мин при одноцентровой сети до 202,3 мин при организации сети из трех сурдоцентров.

Using clustering reduces the average travel time estimate from 97.3 minutes. (median – 82.21 minutes), up to 70.7 minutes. (median 61.4 minutes). Figure 4 shows histograms of the travel time distribution when organizing an audiological network from one and three audiology centers. As an effect [6] of reducing the travel time to the audiology center, the mean difference (MD) – 26.6 minutes can be considered (i.e. a reduction in travel time by almost a third of the baseline mean), and a standardized mean difference (SMD) of 0.5, which can be estimated as the “mean” effect size.

However, it is necessary to pay attention to another important characteristic – a decrease in the variability of travel time: the indicator of the travel time range has decreased from 331.3 minutes. to 200.3 minutes, and the maximum travel time decreased from 334.7 minutes. with a single-center network up to 202.3 min. when organizing a network of three audiology centers.

Заключение

Использование пространственной кластеризации позволяет разработать математически обоснованные решения при организации медицинской помощи на определенной территории с учетом показателя доступности медицинской помощи – времени в пути до медицинской организации.

Моделирование доступности медицинской помощи на основе показателя времени в пути до медицинской организации требует использования геоинформационных систем, оценивающих в реальном времени дорожный трафик. Подобные задачи необходимо предусмотреть в случае широкого использования показателей доступности для решения задач организации и оптимизации территориальных систем здравоохранения.

Кластеризация – это скорее инструмент, который помогает исследовать набор данных и не всегда должен использоваться в качестве автоматического метода классификации данных. Таким образом, не всегда можно применить алгоритмы кластеризации для реального сценария организации медицинской помощи. Они часто слишком ненадежны, и одна кластеризация сама по себе не может дать всю информацию, которую можно извлечь из набора данных. Однако этот инструмент позволяет создать обоснованные предпосылки для оптимального решения задач организации сети специализированных центров и не только.

Conclusion

The use of spatial clustering makes it possible to develop mathematically sound solutions for the organization of medical care in a certain area regard to the indicator of the availability of medical care – the travel time to a medical organization.

Modeling the availability of medical care based on the travel time to a medical organization requires the use of geographic information systems that assess road traffic in real-time. Such tasks must be envisaged in the case of widespread use of availability indicators for solving the problems of organizing and optimizing territorial health care systems.

Clustering is more of a tool that helps explore a dataset and should not always be used as an automatic method for classifying data. Thus, it is not always possible to apply clustering algorithms for a real scenario of the organization of medical care. They are often too unreliable and clustering alone cannot provide all the information that can be extracted from a dataset. However, this tool allows you to create reasonable preconditions for the optimal solution to the problems of organizing a network of specialized centers and not only.

Авторы заявляют об отсутствии конфликта интересов.

The authors declare no conflicts of interest.

ЛИТЕРАТУРА/REFERENCES

1. Davis A. C., Hoffman H. J. Hearing loss: rising prevalence and impact. Bull World Health Organ 2019;97:646–646A. Accessed January 26, 2021. <https://www.who.int/bulletin/volumes/97/10/19-224683.pdf>
2. Addressing the rising prevalence of hearing loss. Geneva: World Health Organization; 2018. Accessed January 26, 2021. <http://www.who.int/pbd/deafness/estimates/en/>
3. Корнеев А. А., Рязанцев С. В., Фанта И. В., Вяземская Е. Э., Левин С. В., Левина Е. А. Пространственный анализ данных в эпидемиологии сенсорной тугоухости. *Российская оториноларингология*. 2020;19(4):13–20. [Korneenkov A. A., Ryazantsev S. V., Fanta I. V., Vyazemskaya E. E., Levin S. V., Levina E. A. Spatial Data Analysis in the epidemiology of sensorineural hearing loss. *Rossiiskaya otorinolaringologiya*. 2020;19(4):13–20. (In Russ.)] <https://doi.org/10.18692/1810-4800-2020-4-13-20>
4. Корнеев А. А., Рязанцев С. В., Левин С. В., Храмов А. В., Вяземская Е. Э., Скирпичников И. Н., Левина Е. А., Павлова С. С. Пространственно-статистический анализ данных о нарушениях слуха у жителей Челябинской области. *Российская оториноларингология*. 2021;20(3):39–50. <https://doi.org/10.18692/1810-4800-2021-3-39-50> [Korneenkov A. A., Ryazantsev S. V., Levin S. V. et al. Spatial analysis of data on hearing impairment in residents of the Chelyabinsk region. *Rossiiskaya otorinolaringologiya*. 2021;20(3):39–50. (In Russ.)] <https://doi.org/10.18692/1810-4800-2021-3-39-50>
5. Jacquez G. M. 2008. Spatial Cluster Analysis. Chapter 22 In «The Handbook of Geographic Information Science», S. Fotheringham and J. Wilson (Eds.). Blackwell Publishing, pages 395–416 https://www.biomedware.com/files/jacquez_ch22_preprint.pdf
6. Корнеев А. А., Фанта И. В. Оценка размера эффекта клинического воздействия. *Российская оториноларингология*. 2020;19(2):42–50. <https://doi.org/10.18692/1810-4800-2020-2-42-50> [Korneenkov A. A., Fanta I. V. Estimation of the effect size of clinical intervention in otorhinolaryngology. *Rossiiskaya otorinolaringologiya*. 2020;19(2):42–50. (In Russ.)] <https://doi.org/10.18692/1810-4800-2020-2-42-50>
7. Закон Челябинской области от 8 ноября 2006 года № 66-ЗО «Об административно-территориальном устройстве Челябинской области». Электронный ресурс. Электронный фонд нормативно-технической и нормативно-правовой информации Консорциума «Кодекс» – <https://docs.cntd.ru/document/802080108> [Law of the Chelyabinsk region of November 8, 2006 N 66-ZO «On the administrative-territorial structure of the Chelyabinsk

region» Electronic resource. electronic fund of normative-technical and normative-legal information of the Consortium «Code» (In Russ.) <https://docs.cntd.ru/document/802080108>

8. Allen J. Using Network Segments in the Visualization of Urban Isochrones. *Cartographica. The International Journal for Geographic Information and Geovisualization*. 2018;53(4):262-270. <https://doi.org/10.3138/cart.53.4.2018-0013>.

Информация об авторах

✉ **Корнеев Алексей Александрович** – доктор медицинских наук, профессор, заведующий лабораторией информатики и статистики, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); e-mail: korneyenkov@gmail.com

ORCID: <https://orcid.org/0000-0001-5870-8042>

Левина Елена Алексеевна – кандидат медицинских наук, старший научный сотрудник, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); e-mail: xramoval@gmail.com

ORCID: <https://orcid.org/0000-0003-0285-6526>

Вяземская Елена Эмильевна – инженер лаборатории информатики и статистики, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); e-mail: vyazemskaya.elena@gmail.com

ORCID: <https://orcid.org/0000-0002-4141-2226>

Скирпичников Илья Николаевич – заведующий ЛОР-отделением, Челябинская областная клиническая больница (454048, Россия, Челябинск, ул. Воровского, д. 70); e-mail: 070173@mail.ru

ORCID: <https://orcid.org/0000-0003-0285-6526>

Левин Сергей Владимирович – кандидат медицинских наук, старший научный сотрудник, Санкт-Петербургский научно-исследовательский институт уха, горла, носа и речи (190013, Россия, Санкт-Петербург, Бронницкая ул., д. 9); e-mail: sergeyln@mail.ru

ORCID: <https://orcid.org/0000-0001-9770-7739>

Information about authors

✉ **Alexei A. Korneenkov** – MD, Professor, Head of the Laboratory of Informatics and Statistics, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya Str., Saint Petersburg, Russia, 190013); e-mail: korneyenkov@gmail.com

ORCID: <https://orcid.org/0000-0001-5870-8042>

Levina Elena Alekseevna – MD Candidate, Researcher, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya Str., Saint Petersburg, Russia, 190013); e-mail: xramoval@gmail.com

ORCID: <https://orcid.org/0000-0003-0285-6526>

Elena E. Vyazemskaya – engineer of the Laboratory of Informatics and Statistics, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya Str., Saint Petersburg, Russia, 190013); e-mail: vyazemskaya.elena@gmail.com

ORCID: <https://orcid.org/0000-0002-4141-2226>

Il'ya N. Skirpichnikov – Head of ENT department, Chelyabinsk Regional Clinical Hospital (70, Vorovsky str., Chelyabinsk, Russia, 454048); e-mail: 070173@mail.ru

ORCID: <https://orcid.org/0000-0003-0285-6526>

Sergei V. Levin – MD Candidate, Researcher, Saint-Petersburg Research Institute of Ear, Throat, Nose and Speech (9, Bronnitskaya Str., Saint Petersburg, Russia, 190013); e-mail: sergeyln@mail.ru

ORCID: <https://orcid.org/0000-0001-9770-7739>