

КЛИНИЧЕСКАЯ МЕДИЦИНА

УДК 616-07 – 303.722
© Коллектив авторов, 2022

А.Ш. Загидуллина¹, Б.М. Азнабаев¹,
Р.Х. Зулкарнеев¹, А.И. Арсланова^{1,2}, С.Р. Авхадеева¹
**ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА ПРИ ФОРМИРОВАНИИ
ФЕНОТИПОВ ПЕРВИЧНОЙ ОТКРЫТОУГОЛЬНОЙ ГЛАУКОМЫ**
¹ФГБОУ ВО «Башкирский государственный медицинский университет»
Минздрава России, г. Уфа
²ГБУЗ РБ «Городская клиническая больница №8», г. Уфа

Цель исследования – изучение эффективности применения иерархического кластерного анализа для формирования фенотипов первичной открытоугольной глаукомы.

Материал и методы. Кластерный анализ был применен для обработки стандартизированного набора клинико-генетических показателей у 348 пациентов (696 глаз) в возрасте от 45 до 87 лет (в среднем 61,28±10,13) с различными стадиями первичной открытоугольной глаукомы (ПОУГ).

Результаты. По итогам многофакторного анализа были выявлены параметры, которые играют важную роль в разделении выборки пациентов с ПОУГ на клинические фенотипы. С целью определения возможных фенотипов больных осуществили кластеризацию по методу Варда с формированием древовидной диаграммы и определением четырех кластеров. Методом k-средних Мак-Квина получены структуры кластеров (фенотипов) пациентов ПОУГ с высоким процентом (96,8%) правильности оценки группировки исходных наблюдений.

Выводы. 1. Показана высокая эффективность применения иерархического кластерного анализа при формировании фенотипов ПОУГ. 2. На основе исследования качественных и количественных анамнестических, клинико-функциональных, морфометрических и генетических параметров у пациентов сформированы четыре клинических фенотипа первичной открытоугольной глаукомы.

Ключевые слова: первичная открытоугольная глаукома, кластерный анализ, значимые факторы, кластеры, фенотипы.

A.Sh. Zagidullina, B.M. Aznabaev,
R.Kh. Zulkarneev, A.I. Arslanova, S.R. Avkhadeeva
**APPLICATION OF CLUSTER ANALYSIS IN THE FORMATION
OF PRIMARY OPEN-ANGLE GLAUCOMA PHENOTYPES**

The purpose is to study the effectiveness of using the hierarchical cluster analysis for the formation of primary open-angle glaucoma phenotypes.

Material and methods. Cluster analysis to process a standardized set of clinical and genetic parameters of 348 patients (696 eyes) aged 45 to 87 (mean 61.28±10.13) years with various stages of primary open-angle glaucoma (POAG). was used.

Results. Based on the results of multivariate analysis, parameters were identified that play an important role in dividing the sample of patients with POAG into clinical phenotypes. In order to determine the possible phenotypes of patients, clustering was carried out according to the Ward method with the tree diagram formation and the definition of four clusters. Using the McQueen k-means method the structures of clusters (phenotypes) of POAG patients were obtained with a high percentage (96.8%) of the correctness of the initial observations grouping.

Conclusions. 1. The high efficiency of hierarchical cluster analysis in the formation of POAG phenotypes has been shown. 2. Based on the study of qualitative and quantitative anamnestic, clinical and functional, morphometric and genetic parameters of patients, four clinical phenotypes of primary open-angle glaucoma were formed.

Key words: primary open-angle glaucoma, cluster analysis, significant factors, clusters, phenotypes.

Кластерный анализ, как один из видов многомерного статистического исследования, применяется для разделения множества исследуемых объектов на однородные группы или кластеры не по одной его характеристике, а по совокупности признаков на базе наиболее близкого внутригруппового сходства и значимых различий между сформированными кластерами [1,2]. Такой подход позволяет осуществлять максимально объективное формирование групп исследуемых объектов.

Целью разделения пациентов на кластеры (фенотипы) определенных заболеваний является оптимизация определения их характера и особенностей клинических проявлений, отражающих гетерогенность заболевания в

целом и патофизиологические признаки разных вариантов его течения. Наряду с этим кластерный анализ, в отличие от других методов статистического анализа, позволяет исследовать большие массивы данных с разнообразными характеристиками.

Формирование фенотипов в клинической практике успешно используется при некоторых широко распространенных заболеваниях: бронхиальная астма, хроническая обструктивная болезнь легких, шизофрения, паранойя, цереброваскулярные заболевания и др. [3-5].

Общеизвестно, что первичная открытоугольная глаукома (ПОУГ), как мультифакторное заболевание, является результатом комбинированного взаимодействия генов и

факторов окружающей среды. Глаукома характеризуется многообразием клинических проявлений и вариантов прогрессирования заболевания.

В связи с вышеизложенным представляет интерес применение методов многомерного математического, в том числе кластерного, анализа для формирования клинических фенотипов ПОУГ.

Цель – изучение эффективности применения иерархического кластерного анализа для формирования фенотипов первичной открытоугольной глаукомы.

Материал и методы

В нашем исследовании кластерный анализ был применен для структуризации стандартизированного набора клинико-генетических признаков у пациентов с ПОУГ.

В Центре лазерного восстановления зрения (ЦЛВЗ) «Оптимед» г. Уфа обследованы 348 пациентов (696 глаз) с различными стадиями ПОУГ в возрасте от 45 до 87 лет (в среднем $61,28 \pm 10,13$), из них 257 (73,85%) женщин и 91 (26,15%) мужчин. Всеми участниками было подписано информированное согласие на проведение исследования.

Глаукома I стадии диагностирована на 182 (52,3%) глазах, II – на 122 (35,06%), III – на 36 (10,34%), IV – на 8 (2,3%). В исследование не вошли пациенты с аномалиями рефракции (астигматизм выше 2,0 Д, миопия и гиперметропия средней и высокой степеней), выраженным помутнением оптических сред глаза, в том числе патологическими изменениями роговицы, патологией зрительного нерва неглаукомного генеза и сетчатки, а также острыми воспалительными и дегенеративными заболеваниями органа зрения. Контрольную группу составили 48 человек без глаукомы (96 глаз), сопоставимых с пациентами ПОУГ по полу и возрасту.

Статистическая обработка результатов осуществлена с применением пакета программ IBM SPSS Statistics v.21.

До проведения кластеризации пациентов на фенотипы с целью выявления наиболее значимых из всех анализируемых клинико-генетических признаков осуществлен многофакторный дисперсионный анализ. На основании данного анализа были отобраны 37 демографических, клинических, функциональных, морфометрических и генетических параметров, оказывающих влияние на формирование количества кластеров и клинико-функциональные характеристики каждого из них.

Перед применением методов кластеризации с целью приведения к общему показателю осуществляли стандартизацию (Z-преобразование) количественных переменных, имеющих различный размах, отличающихся на порядки по их значению.

Решение задачи разделения пациентов с ПОУГ на кластеры проводили в два этапа. На первом осуществляли иерархический кластерный анализ по методу Варда (Ward's method) с целью определения возможных кластеров с формированием древовидной диаграммы (дендрограммы – Tree Clustering), описывающей совокупность объектов наблюдения. На втором этапе были получены структуры кластеров (классов) и их члены методом k-средних Мак-Квина (K-means). В качестве меры расстояния как между классами, так и между элементами внутри классов использовали евклидово расстояние.

Результаты

Проведен факторный анализ для определения значимости факторов, имеющих значение при делении выборки пациентов с ПОУГ на кластеры, результаты которого представлены в табл. 1.

Таблица 1
Оценка значимости факторов, влияющих на деление пациентов ПОУГ на кластеры – факторный анализ ANOVA (n=348)

Показатели	Кластер		Ошибка		F	p
	Ср. кв.	Ст. св.	Ср. кв.	Ст. св.		
1	2	3	4	5	6	7
Пол, мужской – 1, женский – 0	0,24	3	0,19	344	1,23	0,298
Возраст, лет	2424,46	3	82,48	344	29,39	0,000
Миопия, 1 – да, 0 – нет	1,96	3	0,2	344	10,04	0,000
Гиперметропия, 1 – да, 0 – нет	0,03	3	0,047	344	0,63	0,595
Катаракта, 1 – да, 0 – нет	6,02	3	0,19	344	30,71	0,000
ССГ, 1 – да, 0 – нет	3,55	3	0,15	344	24,23	0,000
Начало заболевания, лет	805,29	3	69,15	344	11,65	0,000
Длительность заболевания, лет	468,61	3	13,9	344	33,71	0,000
Коморбидность, баллы	59,25	3	2,27	344	26,09	0,000
Наследственность, 1 – да, 0 – нет	0,2	3	0,21	344	0,94	0,420
ВГД, мм рт. ст.	193,85	3	12,44	344	15,58	0,000
ЦТР, мкм	74374,84	3	215,71	344	344,79	0,000
КГ	5,73	3	0,59	344	9,69	0,000
ПЭС, 1 – да, 0 – нет	16,14	3	0,53	344	30,42	0,000

продолжение таблицы 1

1	2	3	4	5	6	7
Ширина УПК, 1 – узкий, 2 – широкий	0,15	3	0,09	344	1,53	0,207
Пигментация УПК (1-4 степени)	5,84	3	0,39	344	14,68	0,000
Цвет радужки	4,61	3	4,11	344	1,12	0,341
Тип радужки	0,47	3	0,58	344	0,81	0,491
Средняя толщина СНВС, μm	14189,78	3	82,53	344	171,94	0,000
Э/Д	1,29	3	0,02	344	68,74	0,000
Площадь НРП, мм^2	5,84	3	0,07	344	81,95	0,000
Площадь ДЗН, мм^2	0,4	3	0,10	344	3,91	0,009
Объем ЭДЗН, мм^3	1,07	3	00,04	344	25,13	0,000
Средняя толщина ГКС, μm	1084,51	3	8,22	344	131,91	0,000
FLV, %	6840,62	3	46,49	344	147,15	0,000
GLV, %	9216,07	3	83,16	344	110,82	0,000
ONH индекс	0,02	3	0,006	344	2,38	0,069
RPC индекс	0,01	3	0,003	344	5,57	0,001
ONH плотность, %	59,24	3	3,56	344	16,66	0,000
RPC плотность, %	6751,72	3	104,29	344	64,74	0,000
MD, dB	1165,542	3	8,78	344	132,75	0,000
PSD, dB	578,05	3	10,24	344	56,46	0,000
Мутации в гене MYOC, 1 – да, 0 – нет	0,07	3	0,03	344	2,47	0,062
Мутации в гене WDR36	0,16	3	0,04	344	4,57	0,004
Мутации в гене NTF4	0,01	3	0,009	344	0,85	0,468
Мутации в гене CYP1B1	0,002	3	0,003	344	0,67	0,574
Двусторонний процесс	0,04	3	0,03	344	1,16	0,325

Примечания. F – критерий Фишера для оценки значимости различий дисперсий выборок; p – уровень достоверности; Ср. кв. – среднее квадратичное значение; Ст. св. – степень свободы; ССГ – синдром сухого глаза; ВГД – внутриглазное давление; ЦТР – центральная толщина роговицы; КГ – корнеальный гистерезис; ПЭС – псевдоэкзофалиативный синдром; УПК – угол передней камеры; СНВС – слой нервных волокон сетчатки; НРП – нейро-ретинальный пояс; ДЗН – диск зрительного нерва; ЭДЗН – экскавация ДЗН; Э/Д – соотношение размера экскавации к размеру ДЗН; ГКС – ганглиозные клетки сетчатки; FLV (focal loss volume) – объем фокальных потерь ГКС; GLV (global loss volume) – объем глобальных потерь ГКС; ONH (index optic nerve head; полнослойный); RPC (radial peripapillary capillaries, слоя сплетения нервных волокон); MD (mean deviation) – среднее отклонение; PSD (pattern standart deviation) – среднее отклонение светочувствительности сетчатки; MYOC – ген миоцилина; OPTN – ген оптиневрина; WDR36 – ген семейства G-белка, содержащего 36 WDR-повторов; NTF4 – ген нейротрофина; CYP1B1 – цитохрома P450.

С целью разделения выборки пациентов с ПОУГ на кластеры был проведен иерархический кластерный анализ с построением древовидной диаграммы (см. рисунок).

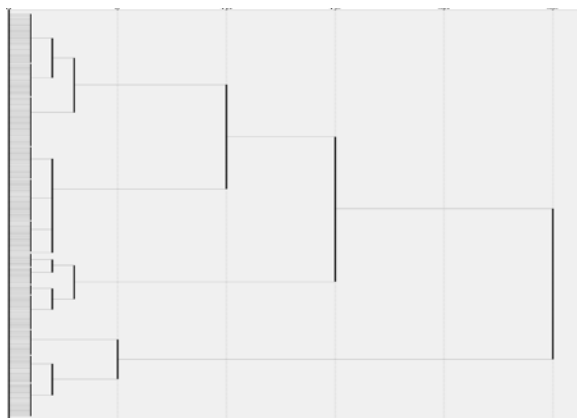


Рис. Древовидная диаграмма разделения пациентов с ПОУГ на кластеры в результате иерархического кластерного анализа. Формирование кластеров по масштабированному расстоянию

На рисунке, отражающем иерархическое дерево, расстояния объединения представлены горизонтальными осями. В результате формирования каждого отдельного кластера была определена величина расстояния, по которому однородные элементы объединены в новый кластер. Визуальный анализ дендрограммы свидетельствует о том, что в результате ветв-

ления «дерева» происходит деление на 2 кластера, каждый из которых в свою очередь делится еще на 2. Таким образом, наиболее оптимальным следует считать выделение четырех кластеров пациентов с ПОУГ.

На следующем этапе кластерного анализа с помощью метода k-средних Мак-Квина (K-means clustering) выборка пациентов была разделена на 4 класса объектов. Использовали те же переменные, что и при проведении иерархического кластерного анализа.

При определении кластеров с использованием пакета программ IBM SPSS Statistics v.21 сходимость была осуществлена по критерию малой величины при стабильном размещении центров кластеров. Для этого понадобилось 22 итерации, после чего расстояние между начальными центрами составило 84,74 (табл. 2). Определены расстояния между конечными центрами кластеров.

Наименее разделенными друг от друга по евклидову расстоянию (42,4 усл. ед.) были кластеры 3 и 4, затем по степени увеличения евклидовых расстояний – кластеры 3 и 2 (евклидово расстояние – 50,02 усл. ед.), кластеры 4 и 1 (50,1 усл. ед.), кластеры 1 и 3 (50,47 усл. ед.), кластеры 1 и 2 (54,07 усл. ед.). Фенотипы 2 и 4 имели наименее близкие характери-

ки, максимальную удаленность друг от друга (евклидово расстояние – 83,42 усл. ед.). В настоящем исследовании кластеры выбира-

лись так, чтобы расхождения между элементами наблюдений определенных кластеров были максимальными.

Таблица 2

Евклидовы расстояния между конечными центрами кластеров				
Кластер	1	2	3	4
1	-	54,07	50,47	50,1
2	54,07	-	50,02	83,42
3	50,47	50,02	-	42,4
4	50,1	83,42	42,4	-

Результаты проведенного анализа принадлежности пациентов к определенным фенотипам (кластерам) показали, что 96,8% исходных сгруппированных наблюдений были классифицированы правильно.

Обсуждение

Клиническая гетерогенность и многофакторность пациентов с ПОУГ обусловлено многообразием ее тяжести и вариантов течения. Следовательно, имеет смысл формирование с помощью кластерного анализа на базе клинико-генетических параметров пациентов отдельных подклассов ПОУГ с целью изучения в дальнейшей перспективе характеристик и особенностей развития динамики заболевания.

Процедура кластеризации исходной выборки на подгруппы, несомненно, обладает рядом преимуществ: не требует предварительного определения целевых значений (функции ошибки или потерь), в связи с чем представляется возможным использование технологии «обучение без учителя»; позволяет анализировать большие массивы данных; при таком методе структуризация объектов более естественна и обоснована, чем при традиционной классификации.

В литературных источниках в основном описываются отдельные клинико-генетические кластеры ПОУГ, сформированные на основе общности метаболических и хромосомных нарушений у пациентов [6,7].

В ряде работ приведен анализ отдельных генетических модификаций, ассоциированных с определенными количественными признаками, такими как ВГД, экскавация ДЗН, соотношение размера экскавации к ДЗН, ЦТР и других, в качестве эндотипов, лежащих в основе заболевания [8,9].

В результате применения кластерного анализа в отношении пациентов с ПОУГ J. Charlesworth et al. [10] были выделены количественные эндотипы ВГД и VCDR, при этом оба основных показателя генети-

чески и фенотипически коррелировали с заболеванием.

За прошедшие годы с помощью иерархического кластерного метода было создано несколько различных кластеров функциональных поражений при ПОУГ на основе показателей полей зрения, предполагающих проекции связей с морфометрической характеристикой оптической когерентной томографии (ОКТ) – RNFL или просто поперечные связи между контрольными точками в периметрических исследованиях [11,12].

При анализе данных опубликованных исследований мы не встретили работ, в которых кластерный анализ применяется с целью разделения пациентов с ПОУГ на клинико-генетические фенотипы.

Формирование кластеров пациентов с ПОУГ с помощью кластерного анализа на основе стандартизированного набора признаков является одним из возможных подходов при выявлении основных клинико-генетических и патогенетических основ развития заболевания и подходов к ведению пациентов.

Выводы

1. Возможность учета больших массивов данных выборки пациентов с различными характеристиками, а также высокий процент (96,8%) правильности оценки группировки исходных наблюдений свидетельствуют о высокой эффективности применения данного вида математического анализа в качестве прикладного инструмента для решения многомерной статистической задачи – формирование фенотипов заболевания.

2. В результате использования иерархического кластерного анализа на основе исследования качественных и количественных анамнестических, клинико-функциональных, морфометрических и генетических параметров у пациентов сформированы четыре клинических фенотипа первичной открытоугольной глаукомы.

Сведения об авторах статьи:

Загидуллина Айгуль Шамилевна – к.м.н., доцент кафедры офтальмологии с курсом ИДПО ФГБОУ ВО БГМУ Минздрава России. Адрес: 450008, г. Уфа, ул. Ленина, 3. Тел. 8(347) 272-41-73. E-mail: aigul.zagidullina@gmail.com.

Азнабаев Булат Маратович – д.м.н., профессор кафедры офтальмологии с курсом ИДПО ФГБОУ ВО БГМУ Минздрава России. Адрес: 450008, г. Уфа, ул. Ленина, 3. Тел. 8(347) 272-41-73.

Зулкарнеев Рустэм Халитович – д.м.н., профессор кафедры пропедевтики внутренних болезней ФГБОУ ВО БГМУ Минздрава России. Адрес: 450008, г. Уфа, ул. Ленина, 3. Тел. 8(347)272-41-73.

Арсланова Айгуль Ирековна – аспирант кафедры офтальмологии с курсом ИДПО ФГБОУ ВО БГМУ Минздрава России, врач-офтальмолог ГБУЗ РБ ГКБ №8 г. Уфа. Адрес: 450008, г. Уфа, ул. Ленина, 3. Тел. 8(347)272-41-73.

Авхадеева Светлана Рудольфовна – к.м.н., доцент кафедры офтальмологии с курсом ИДПО ФГБОУ ВО БГМУ Минздрава России. Адрес: 450008, г. Уфа, ул. Ленина, 3. Тел. 8(347) 272-41-73.

ЛИТЕРАТУРА

1. Берестнева, О.Г. Методы исследования структуры медицинских данных / О.Г. Берестнева, И.А. Осадчая, Е.В. Немеров // Векторы благополучия: экономика и социум. – 2012. – Т. 1, № 2. – С. 333-338.
2. Cluster analysis in Nursing Research: an introduction, historical perspective, and future directions / H. Dunn [et al.] // West J. Nurs. Res. – 2018. – Vol. 40, № 11. – P. 1658-1676.
3. Ненашева, Н.М. Фенотипы бронхиальной астмы и выбор терапии / Н.М. Ненашева // Практическая пульмонология. – 2014. – № 2. – С. 2-11.
4. Радиационно-эпидемиологическая классификация заболеваемости и смертности ликвидаторов от цереброваскулярных заболеваний методами кластерного анализа и главных компонент / В.К. Иванов [и др.] // Радиация и риск. – 2014. – № 4. – С. 8-21.
5. Alshahrani, M. Semantic Disease Gene Embeddings (SmuDGE): phenotype-based disease gene prioritization without phenotypes / M. Alshahrani, R. Hoehndorf // Bioinformatics. – 2018. – Vol. 34, № 17. – P. 901-907.
6. A data mining metabolomics exploration of glaucoma / J. Kouassi Nzoughe [et al.] // Metabolites. – 2020. – Vol. 10, № 2. – P. 49.
7. Differential metabolic markers associated with primary open-angle glaucoma and cataract in human aqueous humor / C.W. Pan [et al.] // BMC Ophthalmol. – 2020. – Vol. 20, № 1. – P. 183.
8. Ocular hypertension treatment study group. The ocular hypertension treatment study: baseline factors that predict the onset of primary open-angle glaucoma / M.O. Gordon [et al.] // Arch. Ophthalmol. – 2002. – Vol. 120, № 6. – P. 714-720.
9. Predictors of long-term progression in the early manifest glaucoma trial / M.C. Leske [et al.] // Ophthalmology. – 2007. – Vol. 114, № 11. – P. 1965-1972.
10. The path to open-angle glaucoma gene discovery: endophenotypic status of intraocular pressure, cup-to-disc ratio, and central corneal thickness / J. Charlesworth [et al.] // Invest. Ophthalmol. Vis. Sci. – 2010. – Vol. 51, № 7. – P. 3509-3514.
11. Gardiner, S.K. Detection of functional deterioration in glaucoma by trend analysis using comprehensive overlapping clusters of locations / S.K. Gardiner, S.L. Mansberger // Sci. Rep. – 2020. – Vol. 10, № 1. – P. 18470.
12. Torres, L.A. Correlating structural and functional damage in glaucoma / L.A. Torres, M. Hatanaka // J. Glaucoma. – 2019. – Vol. 28, № 12. – P. 1079-1085.

REFERENCES

1. Berestneva O.G., Osadchaja I.A., Nemerov E.V. Metody issledovaniya struktury medicinskih dannyh // Vektory blagopoluchija: jekonomika i socium. 2012;1(2): 333-338 (In Russ).
2. Dunn H. et al. Cluster analysis in Nursing Research: an introduction, historical perspective, and future directions // West J. Nurs. Res. 2018; 40(11): 1658-1676. (In English).
3. Nenasheva, N.M. Fenotipy bronhial'noj astmy i vybor terapii // Prakticheskaja pul'monologija. 2014;(2): 2-11 (In Russ).
4. Radiacionno-jepidemiologicheskaja klassifikacija zaboлеваemosti i smernosti likvidatorov ot cerebроваскулярных zaboлеваний metodami klaster'nogo analiza i glavnyh komponent // Radiacija i risk. 2014;(4): 8-21 (In Russ).
5. Alshahrani M., Hoehndorf R. Semantic Disease Gene Embeddings (SmuDGE): phenotype-based disease gene prioritization without phenotypes // Bioinformatics. 2018;34 (17): 901-907. (In English).
6. Kouassi Nzoughe J. et al. A data mining metabolomics exploration of glaucoma // Metabolites. 2020;10(2): 49. (In English).
7. Pan C.W. et al. Differential metabolic markers associated with primary open-angle glaucoma and cataract in human aqueous humor // BMC Ophthalmol. 2020;20(1): 183. (In English).
8. Gordon M.O. et al. Ocular hypertension treatment study group. The ocular hypertension treatment study: baseline factors that predict the onset of primary open-angle glaucoma // Arch. Ophthalmol. 2002;120(6): 714-720. (In English).
9. Leske M. C. et al. Predictors of long-term progression in the early manifest glaucoma trial // Ophthalmology. 2007;114(11): 1965-1972. (In English).
10. Charlesworth J. et al. The path to open-angle glaucoma gene discovery: endophenotypic status of intraocular pressure, cup-to-disc ratio, and central corneal thickness // Invest. Ophthalmol. Vis. Sci. 2010;51(7): 3509-3514. (In English).
11. Gardiner S.K., Mansberger S.L. Detection of functional deterioration in glaucoma by trend analysis using comprehensive overlapping clusters of locations // Sci. Rep. 2020; 10(1): 18470. (In English).
12. Torres L.A., Hatanaka M. Correlating structural and functional damage in glaucoma // J. Glaucoma. 2019;28(12): 1079-1085. (In English).

УДК 616.3-008.1

© Коллектив авторов, 2022

Л.В. Волевач, Н.А. Демидова, А.Ш. Нафикова,
А.А. Камалова, А.С. Салихова, А.Р. Галиуллин,
Л.В. Габбасова, Р.А. Гарипова, Т.С. Загидуллин, И.И. Шакирова
**ФАКТОРЫ РИСКА, ПИЩЕВОЕ ПОВЕДЕНИЕ У ЛИЦ
ПРИ ХРОНИЧЕСКОМ НЕКАЛЬКУЛЕЗНОМ ХОЛЕЦИСТИТЕ
С ИЗБЫТОЧНЫМ ВЕСОМ**

*ФГБОУ ВО «Башкирский государственный медицинский университет»
Минздрава России, г. Уфа*

Целью исследования явилась оценка факторов риска и пищевого поведения у лиц с избыточным весом при патологии желчевыводящих путей.