

УДК 615.035

DOI 10.24412/2312-2935-2025-1-65-80

ИССЛЕДОВАНИЕ ПОТЕНЦИАЛА ИСПОЛЬЗОВАНИЯ РЕГРЕССИОННОГО АНАЛИЗА БОЛЬШИХ ДАННЫХ ДЛЯ ОЦЕНКИ ФАКТОРОВ, ВЛИЯЮЩИХ НА СТОИМОСТЬ ФАРМАЦЕВТИЧЕСКОЙ ПОМОЩИ ДЕТЯМ

А.А. Кондрашов¹, М.М. Курашов.², Е.Е. Лоскутова²

¹ ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова», г. Москва

² ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы», г. Москва

Введение. Оптимизация затрат на фармацевтическую помощь детям является особо важной задачей для системы здравоохранения. Дети являются одной из самых уязвимых групп населения, и их здоровье требует особого внимания и заботы. Высокие затраты на лекарственную терапию заболеваний населения детского возраста становятся значительным финансовым бременем, в первую очередь для их семей, а также для системы обязательного и добровольного медицинского страхования. Современные подходы к экономическому анализу детерминант затрат в сфере здравоохранения включают методы описательной статистики, регрессионного анализа и кластеризации.

Цель. Рассмотреть потенциал применения регрессионного анализа для изучения факторов, влияющих на стоимость фармацевтической помощи детям.

Материалы и методы. База исследования включала данные о назначениях лекарственных препаратов врачами из медицинской информационной системы сети клиник Москвы за 2023 год. Анализ данных выполнялся с использованием языка программирования «Python» и «Jupyter Notebook», включая библиотеки «statsmodels», «scipy», «pandas». Были проведены расчеты размера выборки, проверка нормальности распределения, детальный регрессионный анализ.

Результаты. Исследована база данных назначений лекарственных препаратов и их стоимости в аптеках Москвы. Выявлены факторы, влияющие на стоимость фармацевтической помощи: количество диагнозов, сезон, количество посещений, средняя стоимость на пациента, тип приема, количество препаратов, а также пол, возраст и диагноз пациента. Проведен робастный регрессионный анализ, показавший значимые взаимосвязи между стоимостью фармацевтической помощи и рядом факторов.

Обсуждения. Такие факторы, как диагноз, количество посещений и средняя стоимость на пациента оказались статистически незначимыми или мало влияющими на стоимость. Остальные факторы показали значимую взаимосвязь со стоимостью фармацевтической помощи. Помимо всего прочего был выявлен факт сезонных колебаний и влияния времени года на среднюю стоимость фармацевтической помощи.

Выводы. Робастный регрессионный анализ показал, что выбранная модель объясняет около 70,2% случаев. Результаты исследования могут быть использованы для оптимизации затрат на фармацевтическую помощь, разработки эффективных программ лекарственного страхования и улучшения доступности лекарственных препаратов для детей.

Ключевые слова: фармацевтическая помощь, лекарственные препараты, регрессионный анализ, затраты, дети

INVESTIGATION OF THE POTENTIAL OF USING BIG DATA REGRESSION ANALYSIS TO ASSESS THE FACTORS INFLUENCING THE COST OF PHARMACEUTICAL CARE FOR CHILDREN

A.A. Kondrashov¹, M.M. Kurashov², E.E. Loskutova²

¹ *Lomonosov Moscow State University, Moscow*

² *Peoples' Friendship University of Russia named after Patrice Lumumba, Moscow*

Introduction. Optimizing the costs of pharmaceutical care for children is a particularly important task for the health care system. Children are one of the most vulnerable population groups and their health requires special attention and care. High costs of drug therapy of children's diseases become a significant financial burden, primarily for their families, as well as for the system of compulsory and voluntary health insurance. Modern approaches to the economic analysis of cost determinants in health care include methods of descriptive statistics, regression analysis and clustering.

Objective. To examine the potential of using regression analysis to study the factors affecting the cost of pharmaceutical care for children.

Materials and Methods. The study base included data on prescriptions of pharmaceuticals by physicians from the medical information system of the Moscow clinic network for 2023. Data analysis was performed using the «Python» programming language and «Jupyter Notebook», including the libraries «statsmodels», «scipy», and «pandas». Sample size calculations, normality test of the distribution, and detailed regression analysis were performed.

Results. The database of drug prescriptions and their cost in Moscow pharmacies was studied. Factors affecting the cost of pharmaceutical care were identified: number of diagnoses, season, number of visits, average cost per patient, type of prescription, number of drugs, as well as sex, age and diagnosis of the patient. A robust regression analysis was performed, showing significant relationships between the cost of pharmaceutical care and a number of factors.

Discussion. Factors such as diagnosis, number of visits and average cost per patient were found to have statistically insignificant or little effect on cost. The remaining factors showed a significant relationship with the cost of pharmaceutical care. Among other things, the fact of seasonal variation and the effect of time of year on the average cost of pharmaceutical care was found to be significant.

Conclusions. Robust regression analysis showed that the selected model explains about 70,2% of cases. The results of the study can be used to optimize the costs of pharmaceutical care, develop effective drug insurance programs and improve the availability of medicines for children.

Key words: pharmaceutical care, medicines, regression analysis, costs, children

Введение. Оптимизация затрат на фармацевтическую помощь детям является особо важной задачей для системы здравоохранения. Дети являются одной из самых уязвимых групп населения, и их здоровье требует особого внимания и заботы. Высокие затраты на лекарственную терапию заболеваний населения детского возраста становятся значительным финансовым бременем, в первую очередь, для их семей, а также для системы обязательного и

добровольного медицинского страхования в контексте существующих программ страхования и льготного лекарственного обеспечения. Kim и коллеги в своей статье, посвященной основным процессам и ключевым концепциям экономического анализа в сфере здравоохранения, подчеркивают его важность как инструмента для принятия обоснованных решений о распределении ограниченных ресурсов, таких как денежные затраты на медицинскую и фармацевтическую помощь [1]. Современные подходы к экономическому анализу детерминант затрат в сфере здравоохранения включают в себя методы описательной статистики, а также регрессионного анализа. Различные типы регрессионных моделей, такие как линейные регрессии, логистические регрессии и модели выживания, используемые в экономической оценке разных показателей и областей здравоохранения, используются для идентификации факторов, влияющих на затраты, таких как демографические характеристики, клинические параметры и организационные факторы. Согласно исследованию Gregori и коллег, примерами подобных идентифицируемых факторов, влияющих на различные аспекты затрат, являются стоимость госпитализации, амбулаторного лечения и лекарственной терапии [2]. Lo и коллеги в контексте исследования затрат на терапию пациентов, страдающих артритом, использовали метод квантильной регрессии и определили, что факторы, такие как возраст, пол, наличие сопутствующих заболеваний и уровень инвалидности, были тесно связаны с увеличением затрат [3]. Преимущества использования регрессионных моделей в исследовании подобных факторов включают в себя возможность учета индивидуальных характеристик пациентов и стратегий лечения для получения персонализированных оценок экономической эффективности, а также гибкость обработки сложных взаимосвязей между переменными, учитывая нелинейный характер их взаимодействия [4]. Использование современных статистических методов также соответствует актуальным направлениям развития научных исследований в области медицины, таким как data-driven decision-making (от англ. «принятие решений на основе больших данных») [5,6].

Цель исследования – рассмотреть потенциал применения регрессионного анализа в качестве метода изучения факторов, влияющих на стоимость фармацевтической помощи детям.

Материалы и методы. База исследования включала данные о назначениях лекарственных препаратов врачами из медицинской информационной системы сети клиник города Москвы, оказывающих амбулаторную помощь детям, за период с января по декабрь 2023 года включительно. Предварительная обработка, агрегирование, парсинг (от англ.

«parsing» – автоматизированный сбор и структурирование информации при помощи программы или сервиса) данных о ценообразовании с сайтов открытых источников, статистическая обработка, выделение признаков и визуализация данных были выполнены с использованием инструментов языка программирования «Python» и среды разработки «Jupyter Notebook». Анализ включает расчет размера выборки, проверку нормальности распределения и выполнение детального регрессионного анализа. Для выполнения данного анализа использовались следующие библиотеки: «statsmodels» версии 0.14.2, «scipy» версии 1.14.1, «pandas» версии 2.2.2.

Результаты. Конечной точкой первого этапа исследования являлось получение сведенной воедино базы данных назначений лекарственных препаратов (ЛП) вместе со стоимостью их приобретения в аптеках Москвы. Для сбора данных о стоимости ЛП была написана программа, осуществляющая сбор данных из открытых источников. Программа была написана при помощи инструментов языка программирования «Python»: 1) библиотека «beautifulsoup4» использовалась для получения контента веб-страниц в удобном для выделения нужных ключевых компонентов языка разметки «html» виде, что необходимо для идентификации и поиска нужной информации о ЛП; 2) библиотека «Selenium» использовалась для имитации действий пользователя по вводу данных в текстовые поля, раскрытию блоков с необходимой информацией и переключению между страницами и разделами сайта. По итогу, была получена база, содержащая информацию о ЛП и ценах на них, которая включала в себя следующие поля: «Действующее вещество», «Нозологическая классификация (МКБ-10)», «Фармакологическая группа», «Показания», «Данные по ценам в аптеках», «Название», «Дата обращения», «Лекарственная форма», «Форма выпуска», «Условия отпуска из аптек». Все эти данные необходимы для сопоставления цены с назначенным пациенту ЛП. Сопоставление ЛП из полученной базы данных о ценах с наименованием назначенного ребенку препарата происходило по принципу работы поисковых систем: 1) вычисление коэффициента схожести по степени соответствия названий ЛП, которая вычислялась по методу поиска подстрок в поле «название» и по соответствию действительности информации из других полей; 2) вычисление релевантности происходило по частоте выписывания ЛП по данным базы данных о назначениях. Цена препарата, имеющего наиболее высокие коэффициенты в контексте конкретного случая его назначения, заносилась в отдельную колонку получаемой сводной таблицы и суммировалась с итоговой стоимостью фармацевтической помощи в контексте одного приема конкретного пациента у конкретного врача. Полученная сводная таблица

содержала следующие поля с описанной в них информацией о пациенте, приеме и стоимости фармацевтической помощи: «ID пациента», «Код МКБ», «Диагноз», «Лекарственный препарат», «Схема лечения», «Дата», «Номер мед карты», «Пол», «Возраст», «Специализация врача», «Тип заключения», «ФИО врача», «Стоимость препаратов», «Стоимость фармацевтической помощи». В полученной таблице также была проведена проверка значений на дубликаты по совокупности значений и их содержащих полей.

Таблица 1

Исследуемые показатели и подходы к их предварительной обработке

№	Фактор	Тип данных до предобработки	Процесс предобработки	Тип данных после предобработки
1	Диагноз	Список строк	LabelEncoder* по релевантности	Число
2	Пол	Символьный	OneHotEncoder**	Число
3	Тип заключения	Строка	LabelEncoder* по релевантности	Число
4	Количество препаратов	Число	Предобработка не требуется	Число
5	Возраст	Строка	Преобразование в дробь	Число с точкой
6	Тип приема	Строка	OneHotEncoder**	Число
7	Сезон	Строка	LabelEncoder*	Число
8	Количество диагнозов	Число	Предобработка не требуется	Число
9	Количество посещений	Число	Предобработка не требуется	Число
10	Средняя стоимость на пациента	Число с плавающей точкой	Предобработка не требуется	Число с точкой
* LabelEncoder – это инструмент в Python библиотеке scikit-learn, который преобразует категориальные метки в числовой формат, назначая каждой категории уникальное целое число. ** OneHotEncoder – это инструмент в Python библиотеке scikit-learn, который преобразует категориальные переменные в бинарные векторы, создавая отдельную колонку для каждой категории и указывая 1 при её наличии и 0 при отсутствии.				

Следующий этап работы заключался в предварительной обработке данных, или предобработке, – трансформации данных в вид, пригодный для анализа, и выделении ключевых и потенциально значимых признаков – тех самых факторов, влияние которых на

итоговую стоимость планировалось исследовать. Часть факторов, влияние которых на величину затрат на фармацевтическую помощь следовало оценить, уже имелась в полученной сводной таблице, но требовала предобработки в удобный формат для интерпретации используемыми инструментами аналитики и статистической обработки. В результате дальнейшего исследования имеющегося набора данных были выведены следующие факторы: количество диагнозов, количество посещений, количество препаратов, тип приема, средняя стоимость на пациента, сезон.

Каждый из полученных факторов потребовал свой особый способ предобработки для построения модели регрессионного анализа. В таблице 1 содержится информация о факторах, выделенных в качестве исследуемых показателей, типах содержащихся в них данных, а также подходах к их предобработке и трансформации в необходимый для анализа формат.

Таким образом, после всех мероприятий по предобработке, была получена таблица, содержащая 134951 уникальное ненулевое значение, подразумевающее под собой аналогичное число приемов у врача детей до 18 лет, которое, по результатам расчета размера выборки на основе статистической мощности, в 16 раз больше необходимого размера выборки для нашего исследования (8406). Необходимый размер выборки вычислялся при следующих значениях параметров: $\alpha = 0,05$, ρ (мощность) = 0,9, MDE = 0,05. Тесты Шапиро-Уилка и Колмогоров-Смирнова показывают, что распределение данных по стоимости фармацевтической помощи значительно отличается от нормального. Это важно учитывать при выборе методов анализа данных. Например, для регрессионного анализа необходимо рассмотреть методы, устойчивые к ненормальности данных, или использовать преобразования данных (например, логарифмическое преобразование) для приведения их к нормальному виду. Полученные результаты тестов Шапиро-Уилка и Колмогорова-Смирнова представлены в таблице 2.

Таблица 2

Определение нормальности распределения по Колмогорову-Смирнову и Шапиро-Уилку

Тест на нормальность	Статистика	р-значение	Заключение
Шапиро-Уилка	0,7804	$4,3268 \times e^{-142}$	Распределение не нормальное
Колмогорова-Смирнова	0,1391	0	Локация статистики: 34, Знак статистики: -1, Распределение не нормальное

Третьим этапом данного исследования являлось определение детерминант и ковариант при помощи модели регрессионного анализа и расчета мультиколлинеарности. В данной ситуации стоит отойти от традиционных методов регрессионного анализа, таких как метод наименьших квадратов (МНК), и выбрать робастный анализ. Результаты тестов Шапиро-Уилка и Колмогорова-Смирнова указывают на значительные отклонения от нормальности распределения данных, что может привести к неточным и ненадежным выводам при использовании традиционных методов регрессионного анализа.

Для построения моделей робастной регрессии используется метод итеративно взвешенных наименьших квадратов (англ. IRLS). Вместо того чтобы минимизировать сумму квадратов всех ошибок (как в обычной линейной регрессии), он делает это итеративно, взвешивая данные так, чтобы выбросы и аномальные значения оказывали меньшее влияние на итоговые коэффициенты. Норма Хьюбера (англ. HuberT) используется для управления чувствительностью модели к выбросам. При небольших ошибках она действует как обычная линейная регрессия (квадратичные ошибки), но при больших ошибках она перестает учитывать квадраты, превращаясь в линейную функцию, уменьшая влияние выбросов, тем самым, модель продолжает учитывать важные наблюдения, но не позволяет выбросам существенно искажать результаты. Оценка масштаба с помощью медианного абсолютного отклонения (англ. MAD) используется как робастный способ измерения разброса данных. В отличие от стандартного отклонения, которое чувствительно к выбросам, медианное абсолютное отклонение лучше описывает «нормальные» данные, оставляя выбросы в стороне, что дает более точное представление о том, насколько данные отклоняются от центральной тенденции, без чрезмерного влияния выбросов. Робастная оценка ковариации (тип H1) учитывает возможные гетероскедастичность и выбросы, делая модель менее чувствительной к вариациям данных, что дает более надежные оценки стандартных ошибок и, как следствие, более точные р-значения и доверительные интервалы для коэффициентов.

В качестве зависимой переменной выступают показатели из поля «Стоимость фармацевтической помощи»; независимые переменные соответствуют факторам, описанным в таблице 1. Построение модели регрессионного анализа происходило при помощи интерфейса библиотеки «statsmodels» языка программирования «Python». Итоговые результаты робастного регрессионного анализа представлены в таблице 3.

Таблица 3

Параметры и результаты робастного регрессионного анализа

Параметр	Значение					
Переменная:	Стоимость фармацевтической помощи					
Количество наблюдений:	134951					
Остатки DF (Степени свободы остатков):	134940					
DF модели (Степени свободы модели):	10					
Метод:	IRLS (Итеративно взвешенные наименьшие квадраты)					
Норма:	HuberT (Норма Хьюбера)					
Оценка масштаба:	MAD (Медианное абсолютное отклонение)					
Тип ковариации:	H1 - тип ковариационной матрицы (робастная оценка ковариации)					
Дата:	Ср, 07 Авг 2024					
Время:	00:00:00					
Количество итераций:	20					
Псевдо-R ² :	0,7022					
Независимая переменная	Коэф.	Стд. ошибка	t	P > t	Доверительный интервал	
					0,025	0,975
const	-688,9927	6,7800	-101,6170	0	-702,2820	-675,7030
Диагноз	-0,0013	0,0010	-1,3650	0,1720	-0,0030	0,0010
Тип заключения	32,0837	0,5130	62,6000	0	31,0790	33,0880
Количество препаратов	348,2295	0,7550	461,4870	0	346,7510	349,7080
Возраст дробь	3,4309	0,4990	6,8800	0	2,4530	4,4080
Сезон	-5,694	1,665	-3,3450	0,0001	-8,8330	-2,3060
Количество диагнозов	48,7428	3,0420	16,0250	0	42,7810	54,7050
Количество посещений	-0,3074	0,0770	-3,9810	0	-0,4590	-0,1560
Средняя стоимость на пациента	0,4590	0,0020	226,1290	0	0,4550	0,4630
Пол_М	13,7878	3,8470	3,5840	0	6,2470	21,3280
Тип приема_Повторный	-28,4623	5,0870	-5,5950	0	-38,4320	-18,4920

Процесс оценки коэффициентов происходил итеративно (20 итераций), что означает, что модель постепенно корректирует веса наблюдений, улучшая точность коэффициентов.

Псевдо- R^2 для робастной регрессии показывает, как хорошо модель объясняет вариацию зависимой переменной. Значение 0,7022 (70,22%) указывает на хорошую объясняющую силу модели, что подтверждает, что модель адекватно справляется с описанием данных.

Мультиколлинеарность возникает, когда две или более независимые переменные в регрессионной модели сильно коррелируют между собой. Это может привести к нестабильным оценкам коэффициентов регрессии, что затрудняет интерпретацию модели и снижает ее предсказательную силу. Для выявления мультиколлинеарности используется показатель VIF (от англ. «Variance Inflation Factor») - фактор инфляции дисперсии. Высокое значение VIF (обычно выше 10) указывает на сильную мультиколлинеарность и необходимость корректировки модели, например, путем удаления или преобразования переменных. Таблица 4 описывает результаты проведения анализа на мультиколлинеарность.

Таблица 4

Результаты анализа факторов на мультиколлинеарность

Фактор	VIF
const	12,4715
Диагноз	1,5668
Тип заключения	1,0729
Количество препаратов	1,3360
Возраст дробь	1,0625
Сезон	1,0021
Количество диагнозов	1,5410
Количество посещений	1,0715
Средняя стоимость на пациента	1,3517
Пол_M	1,0026
Тип приема_Повторный	1,0454

Результаты анализа VIF показывают, что ни один из коэффициентов VIF не превышает пороговое значение 10, что указывает на отсутствие значительной мультиколлинеарности среди независимых переменных. Самое высокое значение VIF наблюдается у константы (12,4715), что может быть связано с масштабом модели, но остальные показатели находятся в пределах допустимых значений. Это говорит о том, что мультиколлинеарность не является проблемой для данной модели и независимые переменные не влияют друг на друга в значительной степени.

Обсуждения. В результате проведенного анализа было выявлено, что такие факторы, как «Диагноз» (релевантность диагноза), «Количество посещений» и «Средняя стоимость на

пациента» являются статистически незначимыми, или же влияют на стоимость фармацевтической помощи очень незначительно. Остальные же факторы, как показала модель робастного регрессионного анализа, в той или иной степени имеют взаимосвязь со стоимостью фармацевтической помощи. Для валидации полученных результатов использовали методы описательной статистики и визуализации.

«Тип заключения». Этот фактор подразумевает релевантность типа заключения врача-специалиста, где минимальное значение имеет самый релевантный показатель. Полученные результаты регрессионного анализа этого фактора можно трактовать следующим образом: при уменьшении релевантности типа заключения на единицу, стоимость оказания фармацевтической помощи увеличивается примерно на 32 руб. На рисунке 1 показан график зависимости средней стоимости фармацевтической помощи от типа заключения врача-специалиста, где значения на оси абсцисс отсортированы по убыванию релевантности.

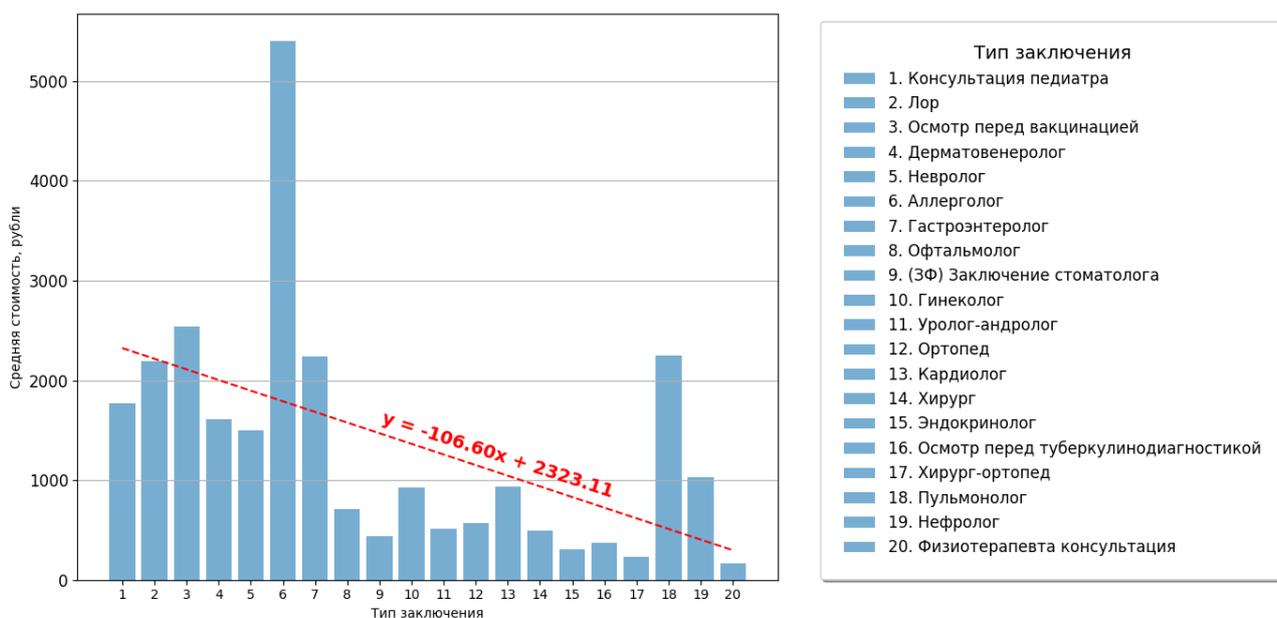


Рисунок 1. Зависимость средней стоимости фармацевтической помощи от типа заключения

В модели регрессионного анализа веса признаков определяются путем минимизации функции потерь, где распространенность признаков в наборе данных влияет на их статистическую значимость и вклад в объяснение вариации зависимой переменной, что отражается в присвоенных им весах. На долю первых 7 типов заключения приходится почти 90% всех заключений врачей, поэтому модель могла определить их суммарный вес как больший, что может служить потенциальной причиной того, почему модель посчитала, что

фактор уменьшения релевантности типа заключения врача увеличивает стоимости фармацевтической помощи.

«Количество препаратов». Положительный вклад этого фактора в стоимость фармацевтической помощи очевиден. По нашим данным стоимость препаратов, назначаемых детям может варьироваться от 10,00 руб. (стоимость самого дешевого препарата под названием «Бриллиантовый зеленый 1% раствор для наружного применения спиртовой 10 мл флакон») до 12770,00 руб. (стоимость самого дорогого препарата под названием «Церебрум композитум, раствор для инъекций 2,2 мл амп. 50 шт.»). Средняя стоимость ЛП составляет $455,89 \pm 482,47$ руб. Согласно результатам регрессионного анализа, при повышении количества назначаемых препаратов стоимость фармацевтической помощи повышается примерно на 348,00 руб., что входит в доверительный интервал средней стоимости ЛП.

«Возраст». Регрессионный анализ фактора возраста показал, что при увеличении возраста пациента на 1 месяц, стоимость фармацевтической помощи ребенку увеличивается примерно на 3,43 руб. На рисунке 2 представлена зависимость средней стоимости фармацевтической помощи от возраста пациента. Полученный график и линия тренда также подтверждают положительный вклад этого фактора в стоимость.

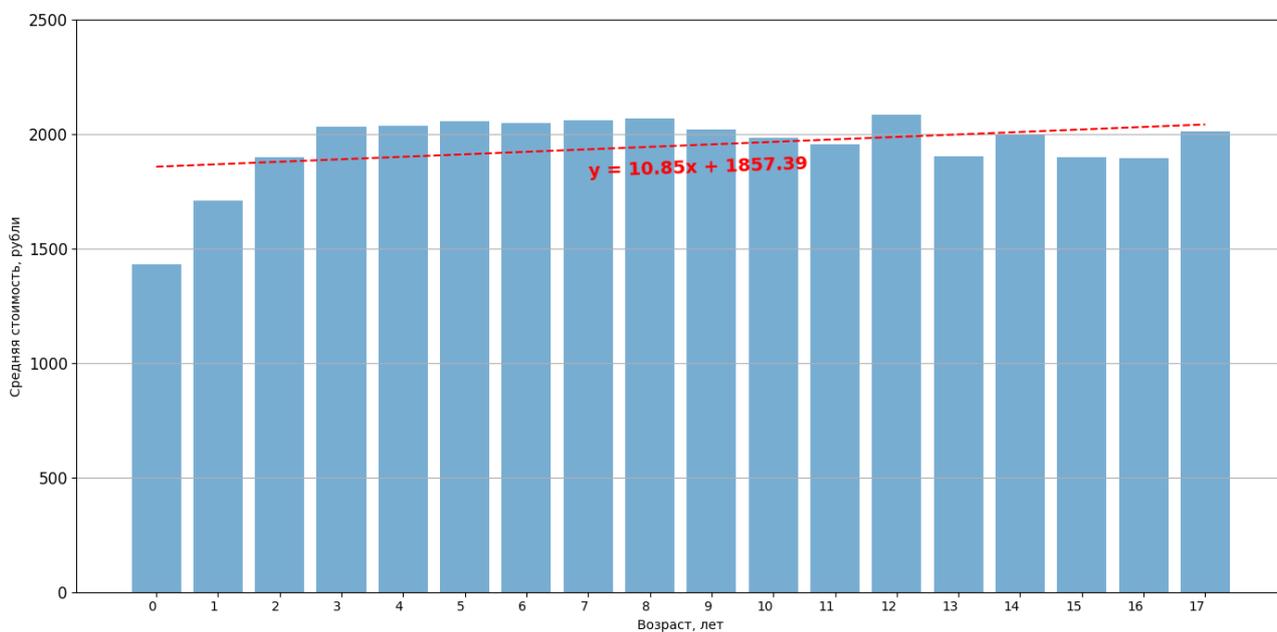


Рисунок 2. Зависимость средней стоимости фармацевтической помощи от возраста пациента

«Количество диагнозов». Анализ данного фактора показал, что, при увеличении количества поставленных врачом диагнозов, стоимость фармацевтической помощи

увеличивается примерно на 48,7 руб. Большое количество поставленных диагнозов может указывать на сложность заболевания и необходимость консультации с профильными врачами-специалистами для уточнения диагнозов. Как правило, в таких ситуациях назначается большее число препаратов, что увеличивает стоимость фармацевтической помощи. Зависимость средней стоимости фармацевтической помощи от количества поставленных врачом-специалистом диагнозов, описанная на рисунке 3, подтверждает влияние данного фактора на стоимость фармацевтической помощи детям.

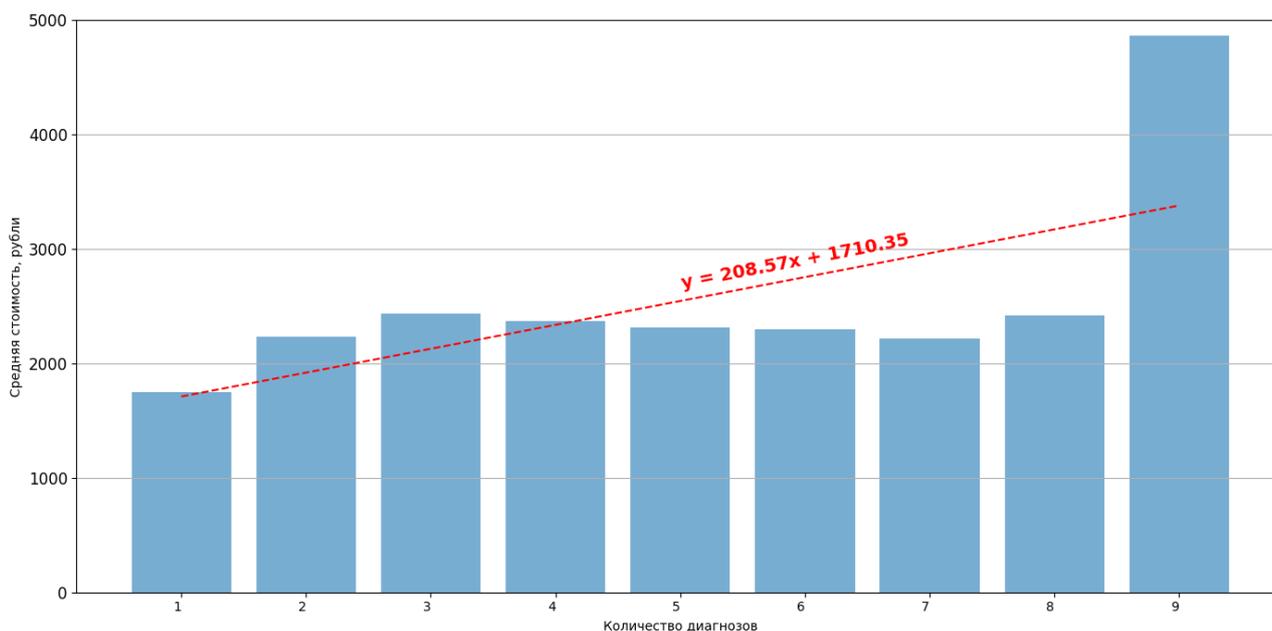


Рисунок 3. Зависимость средней стоимости фармацевтической помощи от количества поставленных врачом диагнозов

«Пол». В таблице образовалось 2 столбца: «Пол_Ж» и «Пол_М». Столбец «Пол_Ж» был удален во избежание мультиколлинеарности, в результате чего полученная модель робастного регрессионного анализа оценивала влияние на стоимость фармацевтической помощи факта, является ли ребенок мальчиком. Согласно полученным результатам, факт наличия у ребенка мужского пола увеличивает стоимость назначаемой лекарственной терапии (добавляет примерно 13,8 руб.). Таблица 6 содержит описание зависимости средней стоимости фармацевтической помощи от пола ребенка.

Таблица 6

Зависимость средней стоимости фармацевтической помощи от пола ребенка

Пол	Количество случаев	Средняя стоимость, руб.	Стандартное отклонение, руб.
Мужской	69833	1963,97	1776,95
Женский	65118	1874,66	1711,47

Средняя стоимость фармацевтической помощи мальчикам примерно на 90,00 руб. \pm 9,50 руб. больше, чем девочкам, что также является фактом, подтверждающим влияние пола ребенка на стоимость.

«Тип приема». Приемы можно разделить на первичные и повторные. Предобработка этого фактора происходила так же, как и пола ребенка – модель регрессионного анализа учитывала влияние на стоимость того, является ли прием повторным. Согласно результатам регрессионного анализа, факт повторного приема снижает стоимость фармацевтической помощи детям. Отрицательное влияние этого фактора также подтверждает таблица 7, где средняя стоимость фармацевтической помощи в рамках первичного приема больше повторного примерно на 69 руб.

Таблица 7

Зависимость средней стоимости фармацевтической помощи от типа приема

Тип приема	Количество случаев	Средняя стоимость, руб.	Стандартное отклонение, руб.
Первичный	110379	1933,50	1737,09
Повторный	24572	1864,17	1785,62

«Сезон». Согласно проведенному робастному регрессионному анализу, увеличение номера сезона (где минимальный номер — это зима, а максимальный – осень), снижает стоимость фармацевтической помощи: с увеличением сезона на единицу стоимость падает примерно на 5,57 руб. Данный факт находит свое подтверждение на диаграмме (рисунок 4), где угловой коэффициент линии тренда является отрицательным. Таким образом, показатель общей стоимости фармацевтической помощи детям подвержен сезонным колебаниям.

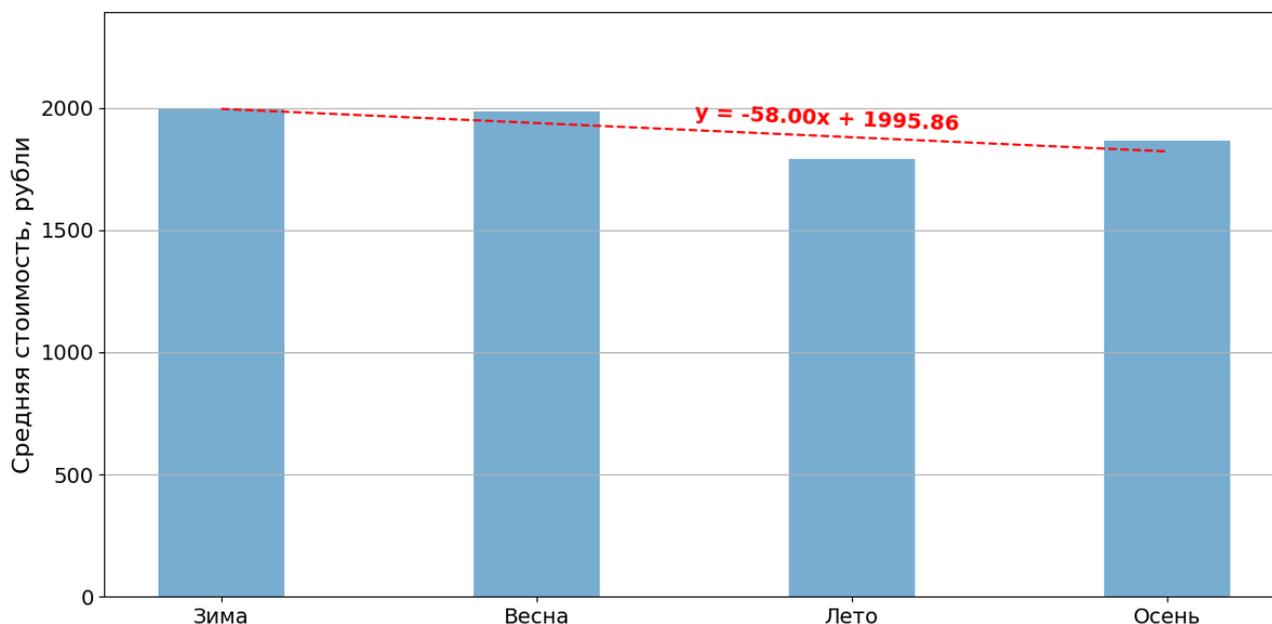


Рисунок 4. Зависимость средней стоимости фармацевтической помощи от времени года

Выводы. На основе данных более чем 134 тысяч медицинских назначений в детских клиниках Москвы за 2023 год был проведен робастный регрессионный анализ и выявлены ключевые детерминанты стоимости фармацевтической помощи детям: количество назначенных препаратов, возраст, количество диагнозов, пол пациента, тип приема и сезон. Робастный метод регрессионного анализа позволил минимизировать влияние выбросов и обеспечить устойчивость результатов в условиях ненормального распределения данных. Лимитирующими факторами подобных методов являются неполнота и плохое качество медицинских данных, а также ограниченные мощности современных персональных компьютеров. Для создания более продвинутых регрессионных моделей требуются дополнительные затраты на вычислительную инфраструктуру, а также исследование данных из разных источников путем создания систем сквозной аналитики и генерации новых признаков и факторов для оценки их влияния на зависимые переменные. Тем не менее, полученные результаты демонстрируют потенциал использования регрессионных методов в решении задач анализа затрат на фармацевтическую помощь детям и открывают широкие возможности для оптимизации расходов и улучшения доступности ЛП, путем доработки и создания выгодных для потребителей программ лекарственного страхования.

Список литературы

1. Kim Y, Kim Y, Lee H-J, et al. The Primary Process and Key Concepts of Economic Evaluation in Healthcare. *Journal of Preventive Medicine and Public Health* 2022;55(5):415–423; doi: 10.3961/jpmph.22.195.
2. Gregori D, Petrinco M, Bo S, et al. Regression models for analyzing costs and their determinants in health care: an introductory review. *International Journal for Quality in Health Care* 2011;23(3):331–341; doi: 10.1093/intqhc/mzr010.
3. Lo T, Parkinson L, Cunich M, et al. Factors associated with higher healthcare costs in individuals living with arthritis: evidence from the quantile regression approach. *Expert Rev Pharmacoecon Outcomes Res* 2015;15(5):833–841; doi: 10.1586/14737167.2015.1037833.
4. Goto D, Shih Y-CT, Lecomte P, et al. Regression-Based Approaches to Patient-Centered Cost-Effectiveness Analysis. *Pharmacoeconomics* 2017;35(7):685–695; doi: 10.1007/s40273-017-0505-5.
5. Basile LJ, Carbonara N, Pellegrino R, et al. Business intelligence in the healthcare industry: The utilization of a data-driven approach to support clinical decision making. *Technovation* 2023;120:102482; doi: 10.1016/j.technovation.2022.102482.
6. Freitas AT. Data-Driven Approaches in Healthcare: Challenges and Emerging Trends. 2024; pp. 65–80; doi: 10.1007/978-3-031-41264-6_4.

References

1. Kim Y, Kim Y, Lee H-J, et al. The Primary Process and Key Concepts of Economic Evaluation in Healthcare. *Journal of Preventive Medicine and Public Health* 2022;55(5):415–423; doi: 10.3961/jpmph.22.195.
2. Gregori D, Petrinco M, Bo S, et al. Regression models for analyzing costs and their determinants in health care: an introductory review. *International Journal for Quality in Health Care* 2011;23(3):331–341; doi: 10.1093/intqhc/mzr010.
3. Lo T, Parkinson L, Cunich M, et al. Factors associated with higher healthcare costs in individuals living with arthritis: evidence from the quantile regression approach. *Expert Rev Pharmacoecon Outcomes Res* 2015;15(5):833–841; doi: 10.1586/14737167.2015.1037833.
4. Goto D, Shih Y-CT, Lecomte P, et al. Regression-Based Approaches to Patient-Centered Cost-Effectiveness Analysis. *Pharmacoeconomics* 2017;35(7):685–695; doi: 10.1007/s40273-017-0505-5.

5. Basile LJ, Carbonara N, Pellegrino R, et al. Business intelligence in the healthcare industry: The utilization of a data-driven approach to support clinical decision making. *Technovation* 2023;120:102482; doi: 10.1016/j.technovation.2022.102482.

6. Freitas AT. Data-Driven Approaches in Healthcare: Challenges and Emerging Trends. 2024; pp. 65–80; doi: 10.1007/978-3-031-41264-6_4.

Финансирование. Исследование не имело спонсорской поддержки.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Acknowledgments. The study did not have sponsorship.

Conflict of interests. The authors declare no conflict of interest.

Информация об авторах

Кондрашов Александр Андреевич – аспирант кафедры фармацевтической химии и организации фармацевтического дела, ФГБОУ ВО «Московский государственный университет имени М.В. Ломоносова», 119234, г. Москва, тер Ленинские Горы, д. 1., e-mail: kondrasasha@mail.ru; ORCID: 0009-0004-0601-3205; SPIN: 1222-7345

Курашов Максим Михайлович – кандидат фармацевтических наук, доцент кафедры управления и экономики фармации ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы», 17198, Россия, г. Москва, ул. Миклухо-Маклая, д. 6., e-mail: kurashov-mm@rudn.ru; ORCID: 0000-0002-0349-905X; SPIN: 8741-3469

Лоскутова Екатерина Ефимовна - доктор фармацевтических наук, профессор, заведующая кафедрой управления и экономики фармации ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы», 17198, Россия, г. Москва, ул. Миклухо-Маклая, д. 6., email: loskutova-ee@rudn.ru; ORCID: 0000-0002-1514-0941; SPIN: 1133-0394

About the authors

Alexander A. Kondrashov – Postgraduate student of the Department of Pharmaceutical Chemistry and Organization of Pharmaceutical Business, Lomonosov Moscow State University, 19234, Russia, Moscow, Leninskie Gory 1, email: kondrasasha@mail.ru; ORCID: 0009-0004-0601-3205; SPIN: 1222-7345

Maxim M. Kurashov – PhD of Pharmaceutical Sciences, Associate Professor of Department Management and economics of pharmacy, Peoples' Friendship University of Russia named after Patrice Lumumba, 17198, Russia, Moscow, Miklouho-Maklaya 6., email: kurashov-mm@rudn.ru; ORCID: 0000-0002-0349-905X; SPIN: 8741-3469

Ekaterina E. Loskutova - Doctor of Pharmaceutical Sciences, Professor, Head of Department Management and economics of pharmacy, Peoples' Friendship University of Russia named after Patrice Lumumba, 17198, Russia, Moscow, Miklouho-Maklaya 6., email: loskutova-ee@rudn.ru; ORCID: 0000-0002-1514-0941; SPIN: 1133-0394

Статья получена: 19.09.2024 г.

Принята к публикации: 25.03.2025 г.