## Анализ влияния социально-экономических факторов на эпидемическую ситуацию по туберкулезу в регионах Российской Федерации

Шахгельдян К.И.12, Гельцер Б.И.1, Емцева Е.Д.2, Середа В. Г. 1, Мазелис Л.С. 2, Сахарова О.Б.1

1ГАОУ ВО Дальневосточный федеральный университет, Школа биомедицины, МОН России, 690950, г. Владивосток, Россия;

2ФГБОУ ВО Владивостокский государственный университет экономики и сервиса, МОН России, 690014, Владивосток, Россия;

Для корреспонденции: Шахгельдян Карина Иосифовна, доктор технических наук, доцент, директор института Информационных технологий, Владивостокский государственный университет экономики и сервиса; заведующий лабораторией Анализа больших данных в здравоохранении и биомедицине Школы биомедицины ДВФУ e-mail: carinash@vvsu.ru

For correspondence: Karina J. Shakhgeldyan, Dr. Sci.of Computer, chief of information technology department, Vladivostok State University of Economics, e-mail: carinash@vvsu.ru

**Information about authors:**

Shakhgeldyan K.J. <http://orcid.org/0000-0002-4539-685X>

Geltser B.I. <http://orcid.org/0000-0002-9250-557X>

Emtseva E.D. https://orcid.org/0000-0002-9760-5481

Sereda V.G <https://orcid.org/0000-0001-9432-4847>

Mazelis L.S. <http://orcid.org/0000-0001-7346-3960>

Sakharova O.B. https://orcid.org/0000-0001-9569-5340

Цель исследования состояла в идентификации социально-экономических факторов, влияющих на эпидемическую ситуацию по туберкулезу (ТБ) в различных регионах РФ, и оценке степени этого влияния на основе использования современных методов машинного обучения. Авторами были разработаны многофакторные линейные регрессионные модели на основе панельной структуры данных. Разделение регионов РФ по бремени ТБ выполняли методами кластерного анализа. Оценку степени влияния изучаемых факторов на эпидемический процесс выполняли с помощью метрики Пратта. Всего было обработано около 100 тыс. записей по 78 регионам РФ. Было выделено 4 региональных кластера, различающихся между собой по уровню заболеваемости, распространенности и смертности от ТБ. Разнообразие социально-экономических факторов, влияющих на основные индикаторы эпидемического процесса, сокращается по мере роста его напряженности. Если в первом кластере на заболеваемость влияли 7 факторов, а на уровень смертности - 4, то в четвертом кластере только 1 фактор влиял на заболеваемость и 2 - на смертность. При этом интенсивность их воздействия на эпидемический процесс существенно возрастала. Так, например, роль фактора “площадь жилья на душу населения” в снижении заболеваемости ТБ среди регионов 4-го кластера оценивалась в 79%, а среди регионов первого кластера - в 17%. Результаты исследования показали, что современные методы машинного обучения позволяют разрабатывать модели, с помощью которых можно оценить в количественном выражении влияние социально-экономических условий конкретного региона на активность эпидемического процесса ТБ.

### Ключевые слова: туберкулез; социально-экономические факторы; регрессионные модели; кластерный анализ.

Analysis of the impact of socio-economic factors on the epidemiological process of tuberculosis in the regions of the Russian Federation

The purpose of the study was to identify the socio-economic factors affecting the epidemiological situation of tuberculosis (TB) in various regions of the Russian Federation (RF), and to assess the extent of this influence using modern methods of machine learning. To achieve it, the authors developed multifactorial linear regression models based on a panel data structure. The division of the regions of the RF into the TB-burden was carried out by methods of cluster analysis. Evaluation of the degree of influence of the studied factors on the epidemiological process was carried out using the Pratt’s metric. Total processed about 100 thousand records for 78 regions of RF. Four regional clusters were distinguished, differing in terms of morbidity, prevalence and mortality from TB. The variety of socio-economic factors that affect the main indicators of the epidemiological process is decreasing as its tension grows. If in the first cluster, the morbidity was influenced by 7 factors, and by the death rate - 4, then in the fourth cluster only 1 factor influenced morbidity and 2 - on mortality. At the same time, the intensity of their impact on the epidemiological process increased significantly. For example, the role of the factor "per capita housing space" in reducing the morbidity of TB among the regions of the 4th cluster was estimated at 79%, and among the regions of the first cluster - in 17%. The results of the research showed that modern methods of machine learning allow us to develop models that can be used to quantify the impact of the socio-economic conditions of a particular region on the TB epidemiological process.

Keywords: tuberculosis; socio-economic factors; regression models; cluster analysis.

Туберкулез (ТБ) является одной из самых актуальных проблем современного общества и индикатором его социально-экономического благополучия. Именно поэтому в научных исследованиях последних лет особое внимание уделяется изучению взаимосвязей различных показателей, определяющих уровень социально-экономического развития отдельных стран и регионов, с индикаторами, характеризующими активность эпидемического процесса ТБ. В ряде работ было показано, что к основным факторам, ограничивающими распространение ТБ относятся улучшение санитарно-гигиенических условий проживания и качества питания населения [1], высокий уровень его образования, повышение доходов [2] и доступности медицинской помощи [3]. Среди причин, усиливающих активность эпидемического процесса, выделяют алкоголизм [4], табакокурение, высокий уровень бездомности, бедности и безработицы [5], растущую концентрацию доходов и др. [6] Важным фактором, оказывающим влияние на развитие ТБ, является социально-экономический статус индивидуума, в характеристику которого помимо некоторых вышеуказанных показателей, входит генетически обусловленная восприимчивость к ТБ, возраст, пол, а также этническая и расовая принадлежность [1, 7]. Дополнительными факторами, влияющими на развитие ТБ у инфицированных лиц, являются ВИЧ-инфекция, сахарный диабет, различные варианты химиотерапии, подавляющие иммунитет, а также низкий индекс массы тела [1].

Выделение приоритетных факторов, ухудшающих эпидемическую ситуацию по ТБ, и реализация мероприятий, ограничивающих их влияние, обеспечили в ряде стран снижение заболеваемости. Основные исследования по оценке взаимосвязей индикаторов социально-экономического развития и активности эпидемического процесса относятся к государствам с развитой экономикой [7, 8]. В последние годы появились аналогичные работы по некоторым странам БРИК и развивающимся странам [5]. Россия имеет самую большую в мире площадь территории, состоящую из регионов, которые существенно различаются между собой по уровню экономического развития, плотности населения, его благосостояния и условий проживания, степени урбанизации, климатическим условиям и др. Уникальность и вариативность социально-экономических факторов и их комбинаций в отдельных субъектах РФ подчеркивает необходимость дифференцированного подхода к оценке их влияния на эпидемический процесс ТБ.

Цель исследования состояла в идентификации социально-экономических факторов, влияющих на активность эпидемического процесса ТБ в различных регионах РФ, и оценке степени этого влияния на основе использования современных методов машинного обучения.

**Материалы и методы**

Показатели активности эпидемического процесса ТБ в 78 регионах РФ за 2002-2016 гг. получены из открытых материалов сайта Центрального НИИ организации и информатизации здравоохранения (<http://mednet.ru/en.html>). Показатели социально-экономического развития субъектов РФ получены из материалов официального сайта Российской статистической службы ([http://www.gks.ru](http://www.gks.ru/wps/wcm/connect/rosstat_main/rosstat/ru/statistics/publications/catalog/doc_1138623506156)). Общий объем используемых данных составил около 100 тыс. записей. Для их хранения использовался сервер с двумя Intel Xeon процессорами E5-2267 v2, 768Gb DDR3 RAM, с Windows Server 2012 Standart и системой управления базой данных MS SQL Server Enterprise Edition 2012.

Активность эпидемического процесса ТБ оценивали по показателям заболеваемости, распространенности и смертности от ТБ [9]. В исследование было включено 21 социально-экономических фактор, сгруппированные по 5 категориям: уровень экономического развития региона; условия проживания местного населения; социальные риски; качество питания; доступность фтизиатрической помощи. Экономическое развитие региона характеризовали валовым региональным продуктом на душу населения в ценах 2001 г. (далее - ВРП), среднедушевыми денежными доходами в ценах 2001 г. (далее - Доходы), уровнем бедности (число лиц на 1 тыс. населения с доходами ниже прожиточного минимума) и безработицы (далее - Безработица). К категории “условия проживания” относили удельный вес жилой площади, оборудованной канализацией и водопроводом (далее - Водопровод), удельный вес ветхого и аварийного жилищного фонда (далее - Ветхое жилье), площадь жилищ, приходящаяся в среднем на одного жителя (далее - Площадь жилья), удельный вес городского населения в его общей численности (далее - Городское население). Кроме того, к этой категории отнесли плотность населения, проживающего в регионе, и численность студентов вузов на 10 тыс. населения (далее - Студенты), которая характеризует скученность проживания. Анализировали также количество выбросов в атмосферу загрязняющих веществ, отходящих от стационарных источников (далее - Выбросы), использование свежей воды на душу населения. К социальным рискам относили объем потребления алкоголя на душу населения (далее - Потребление алкоголя), который вычисляли через розничную продажу водки, ликеро-водочных изделий, виноградных и плодовых вин и пива, а также распространенность наркомании (число наркоманов на 10 тыс. населения). Исследование качества питания включало анализ потребления мяса и мясопродуктов (далее - Потребление мяса), молока и молочных продуктов, яиц, овощей и продовольственных бахчевых культур, приходящихся на душу населения. Доступность фтизиатрической помощи оценивали через показатель обеспеченности фтизиатрами.

Пространственно-временной массив данных по регионам РФ за 14 лет обрабатывался с помощью специализированных методов анализа панельных данных [10]. Для оценки влияния социально-экономических факторов на эпидемический процесс ТБ авторами были разработаны многофакторные линейные регрессионные модели, объединяющие только те данные, которые имели достоверность p-value < 0,01, при множественном коэффициенте детерминации R2 > 0,35. При этом общая точность моделей составила p-value <0,01. Для всех факторов, участвующих в их построении, индекс инфляции дисперсии был меньше 3, что свидетельствовало об отсутствии мультиколлинеарности [11].

Разделение регионов РФ по бремени ТБ выполняли методами иерархической кластеризации и К-средних, используя показатели заболеваемости, распространенности и смертности от ТБ [12]. Оценку валидности кластеризации проводили с помощью индекса Дэвиса-Болдина, уровень которого составил 0,57, что свидетельствовало о высоком уровне ее достоверности [13].

Для каждого кластера были разработаны отдельные линейные регрессионные модели, определяющие особенности взаимосвязей показателей, характеризующих социально-экономический статус регионов и активность эпидемического процесса. Оценку степени влияния изучаемых факторов на эпидемический процесс выполняли с помощью метрики Пратта [14]. Обработку данных проводили в программе R-Studio.

**Результаты исследования**

Пространственный анализ активности эпидемического процесса ТБ позволил определить вектор его изменений с запада на восток России, что согласуется с результатами других исследований [15]. Минимальные уровни заболеваемости и смертности от ТБ наблюдались в Центральном (ЦФО) и Северо-Западном (СЗФО) Федеральных округах, а их максимальный уровень фиксировался в Сибирском (СФО) и Дальневосточном (ДФО) Федеральных округах.

На предварительном этапе исследования были построены 2 линейные регрессионные модели, описывающие зависимость показателей заболеваемости и смертности от 21 фактора социально-экономического развития регионов. Однако, в итоговые модели вошли только 12 из них, прогностическая значимость которых была подтверждена малым значением статистической ошибки первого рода (p-value<0,01) (табл. 1). Величины множественных коэффициентов детерминации 0,51 и 0,43 свидетельствовали о том, что построенные модели способны объяснить только 51% дисперсии по заболеваемости и 43% дисперсии по смертности от ТБ. Для моделирования оставшейся, “необъяснённой”, части дисперсии необходимо использовать другие факторы, не вошедшие в данное исследование (генетические, иммунологические, микробиологические, климато-географические и др.).

Таблица 1.

Коэффициенты линейной регрессии и степень влияния социально-экономических факторов на заболеваемость и смертность от ТБ в РФ

|  |  |  |
| --- | --- | --- |
| Факторы\* | Заболеваемость ТБ | Смертность от ТБ |
| Коэффициент | Степень влияния, % | Коэффициент | Степень влияния, % |
| Безработица | - | - | 0,06 | 2,4 |
| Городское население | -0,54 | 4,2 | - | - |
| Студенты | 0,24 | 11,8 | - | - |
| Площадь жилья | -0,23 | 17,9 | -0,29 | 16,2 |
| Выбросы  | 0,26 | 1,7 | - | - |
| Водопровод | -0,2 | 5,6 | - | - |
| Ветхое жилье | -0,17 | 2,6 | - | - |
| Потребление алкоголя | 0,06 | 0,8 | - | - |
| Потребление крепкого алкоголя | - |  | 0,18 | 10,9 |
| Потребление яиц | -0,1 | 3,4 | - | - |
| Потребление мяса | -0,08 | 6,9 | -0,19 | 13,5 |
| p-value модели | <2,2e-16 |  | <2,2e-16 |  |
| p-value всех факторов | <0,003 |  | <0,008 |  |
| R2 множественный коэффициент детерминации | 0,51 |  | 0,43 |  |

\*Примечание: В таблицу вошли только те факторы, оценка влияние которых имела значение статистической ошибки первого рода p-value < 0,01.

Результаты исследования показали, что площадь жилья имеет обратную зависимость от заболеваемости ТБ и вносит наибольший вклад в изменение ее уровня. Количественная характеристика этого влияния составила 17,9%. Число студентов вузов, как индикатор скученности проживания, указывает на высокий риск тесного и/или длительного контакта с возможным источником инфицирования. Вклад этого фактора в заболеваемость составляет 11,8%. В то же время влияние на заболеваемость условий проживания в ветхом и аварийном жилье оценивалось в 2,6%. Росту заболеваемости ТБ способствовало также и ухудшение качества атмосферного воздуха, доля которого в общем пуле анализируемых факторов была 1,7%. Заметное влияние на заболеваемость ТБ в регионах РФ оказывал удельный вес городского населения, с увеличением которого активность эпидемического процесса снижалась. Вклад этого фактора в заболеваемость составил 4,2%. Эти данные отличаются от исследования, в котором установлена прямая связь заболеваемости с увеличением городского населения в некоторых развитых странах мира [16].

Сбалансированное питание с необходимым количеством белков животного происхождения предотвращает, а употребление алкоголя увеличивает риск возникновения заболевания. Роль этих факторов в заболеваемости ТБ составила 10,3% и 0,8% соответственно.

Площадь жилья оказывает существенное влияние не только на заболеваемость, но и на смертность от ТБ с долей участия в 16,2%. Социальная дезадаптация населения регионов также относилась к факторам, влияющим на уровень смертности от ТБ. При этом, роль безработицы оценивалась в 2,4%, а потребление крепких алкогольных напитков - в 10,9%. Снижение потребления мяса и мясопродуктов, как одно из проявлений ухудшения качества жизни, способствовало увеличению смертности с долей участия в 13,5%**.**

В зависимости от активности эпидемического процесса ТБ было выделено 4 региональных кластера, различающихся между собой по значениям анализируемых показателей (табл. 2).

Таблица 2.

Средние значения показателей активности эпидемического процесса
в различных региональных кластерах

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Кластер | Кол-во регионов | Заболеваемость, на 100 тыс. населения | Распространенность, на 100 тыс. населения | Смертность, на 100 тыс. населения |
| Средние значения | 95% CI\* | Средние значения | 95% CI\* | Средние значения | 95% CI\* |
| 1 | 28 | 41,3 | 37-45 | 76 | 66,5-83 | 4,8 | 4-5,5 |
| 2 | 29 | 57,8 | 52-63 | 136,7 | 129-144 | 9,1 | 7,6-10,3 |
| 3 | 17 | 90,6 | 81-100 | 225,8 | 208-244 | 15,7 | 13-18,4 |
| 4 | 4 | 145 | 118-172 | 392 | 220-566 | 31 | 11,7-50,3 |

\*Примечание: 95% CI - доверительный интервал

Первый, наиболее благополучный по эпидемической ситуации кластер, состоит из большинства регионов ЦФО, почти всех субъектов СЗФО, части Приволжского (ПФО) Федерального округа, а также Республики Дагестан и Томской области. Во второй кластер вошли регионы ЦФО и СЗФО, Северо-Кавказского Федерального округа (СКФО), не вошедшие в первый кластер, а также часть регионов Южного (ЮФО) и Уральского (УФО) Федеральных округов, ПФО, СФО, и ДФО. Третий кластер состоит из части регионов ДФО, СФО, УФО, а также 2-х регионов ПФО и 3-х регионов ЮФО. Четвертый кластер состоит из 4-х регионов РФ, расположенных на территориях Дальнего Востока и Восточной Сибири.

Оценка результатов исследований, характеризующих зависимость основных показателей активности эпидемического процесса ТБ от социально-экономических факторов в отдельных региональных кластерах, показала наличие определенных взаимосвязей между ними (табл. 3).

Таблица 3

Коэффициенты линейной регрессии и степень влияния социально-экономических факторов на заболеваемость и смертность от ТБ в региональных кластерах

|  |  |  |
| --- | --- | --- |
| Факторы | Заболеваемость в кластерах | Смертность в кластерах |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| ВРП | - | -0,287,1% | - | - | - | - | - | - |
| Доход | - | - | - | - | -0,48 25,6% | -0,5126,7% | - | -0,4115,6% |
| Безработица | - | - | - | - | - | - | - | 0,5262,4% |
| Городское население | -1,047,5% | - | - | - | - | - | - | - |
| Плотность населения | 1,830,6% | - | - | - | - | - | - | - |
| Студенты | 0,4010,1% | 0,2210,7% | 0,35 16,0% | - | - | - | - | - |
| Площадь жилья  | -0,2317% | -0,2518,9% | -0,4328,4% | -0,8879% | -0,46 30,1% | - | - | - |
| Водопровод  | -0,7418,5% | -0,4117,1% |  |  | - | - | - | - |
| Ветхое жилье |  |  |  |  | 0,34 3,6% | 0,40,8% | 0,454,7% | - |
| Потребление алкоголя | 0,150,5% | 0,140,8% | 0,222,6% |  | - | - | - | - |
| Потребление крепкого алкоголя | - | - | - | - | 0,18 1% | 0,2611,6% | 0,313,7% | - |
| Потребление мяса | -0,3322,8% | - | - | - | - | -0,212,2% | -0,4219,7% | - |
| Потребление яиц | - | -0,256,9% | - | - | - | - | - | - |
| p-value | <2•e-16 | <2•e-16 | <2•e-16 | <1,5•e-15 | <2•e-16 | <2•e-16 | <2•e-16 | <2•e-16 |
| p-value всех факторов | <0,0002 | <0,002 | <0,0006 | <1,5•e-15 | <0,0003 | <0,008 | <0,002 | <0,0006 |
| Множественный коэффициент детерминации R2 | 0,76 | 0,53 | 0,47 | 0,79 | 0,69 | 0,51 | 0,38 | 0,78 |
| Количество факторов | 7 | 6 | 3 | 2 | 4 | 4 | 3 | 2 |

Примечание: Здесь и в других табл.: абсолютные значения - коэффициенты регрессии, % обозначена степень влияния факторов на заболеваемость и смертность

Для регионов первого кластера была построена модель, которая позволила выделить 11 социально-экономических факторов и описать их влияние на заболеваемость и смертность от ТБ. Наибольший вклад в заболеваемость вносили факторы, характеризующие условия проживания населения (площадь жилья и наличие централизованного водоснабжения), совокупный вклад которых составил 35,5%. Аналогичное, но менее заметное влияние на этот показатель оказывало потребление мяса и мясопродуктов с долей участия в 22,8%. Определенное влияние на заболеваемость оказывали демографические факторы (плотность населения и удельный вес городского населения), роль которых в совокупности оценивалась на уровне 8,1%. При этом воздействие данных факторов на эпидемический процесс было разнонаправленным. Так, удельный вес городского населения в этих регионах был в обратной зависимости от заболеваемости, а вклад указанного фактора в ее снижение составил 7,5%, что объясняется, по-видимому, большей доступностью фтизиатрической помощи для жителей городов. Было установлено, что уровень благосостояния населения и условия его проживания доминируют в своем влиянии на смертность в этих регионах, роль которых в совокупности оценивалась почти в 60%. В этом кластере потребление алкоголя не оказывало существенного воздействия на уровень смертности и составляло всего 1%.

Во втором кластере, где эпидемическая ситуация была более напряженной, количество социально-экономических факторов, оказывающих влияние на нее снижалось до 10, но интенсивность их воздействия нарастала. Так, например, в этом кластере отмечалась тесная отрицательная взаимосвязь между показателями экономического развития региона и условиями проживания населения - с одной стороны и показателями заболеваемости и смертности от ТБ - с другой. Совокупное влияние этих факторов на заболеваемость и смертность составило 43,1% и 27,5% соответственно. В этом кластере увеличивалось также потребление крепкого алкоголя, а его вклад в смертность от ТБ возрастал до 11,6%.

В третьем кластере было выделено 6 факторов, тесно связанных с активностью эпидемического процесса. Так, уровень заболеваемости был взаимосвязан с индикаторами скученности проживания: площадью жилья и количеством студентов, влияние которых составило 10,7% и 18,9% соответственно. Росту смертности от ТБ в этих регионах способствовало увеличение потребления крепкого алкоголя, роль которого оценивалась в 13,7%. Обратная зависимость показателя смертности от потребления мяса и мясопродуктов демонстрировала значение сбалансированного питания в ее сокращении, роль которого оценивалась в 19,7%. Определенный вклад в увеличение смертности вносил фактор ветхого жилья: его удельный вес составлял 4,7%.

В четвертый кластер, с наиболее тяжелой эпидемической ситуацией по ТБ, вошли Еврейская автономная область, Приморский край, Республика Тыва, Чукотский автономный округ. В этом кластере было выделено всего 3 фактора с доминирующим влиянием на эпидемический процесс. К ним относились: недостаточная площадь жилья, низкий доход на душу населения и безработица. Вклад первого из них в уровень заболеваемости составил 79%, а совокупный вклад двух других в смертность - 78%.

С целью детальной оценки влияния социально-экономических факторов на эпидемический процесс отдельно построили модели линейной регрессии для самых неблагоприятных по ТБ территорий: Приморского края и Республики Тыва (табл.4).

Таблица 4.

Коэффициенты линейной регрессии и степень влияния социально-экономических факторов на заболеваемость и смертность от ТБ в Республике Тыва и Приморском крае

|  |  |  |
| --- | --- | --- |
| Факторы | Заболеваемость | Смертность |
| Тыва | Приморский край | Тыва | Приморский край |
| ВРП  | -0,55 (48,1%) | - | - | - |
| Потребление алкоголя |  | 1,12 (55,6%) | - | - |
| Потребление крепкого алкоголя | 0,45 (35,7%) |  |  |  |
| Обеспеченность фтизиатрами | -0,22 (12,1%) | - | - | - |
| Площадь жилья | - | -0,91 (11,1%) | - |  |
| Потребление мяса | - | - | - | -0,82 (67%) |
| p-value | <2,8e-7 | <0,003 | **-** | <0,0004 |
| p-value всех коэффициентов | <0,01 | <0,004 | **-** | <0,0004 |
| Множественный коэффициент детерминации R2 | 0,96 | 0,67 | **-** | 0,67 |

 Результаты исследования показали, что в Республике Тыва на заболеваемость ТБ оказывают влияние низкий уровень ВРП (48,1%), интенсивное потребление крепкого алкоголя (35,7%) и обеспеченность фтизиатрами (12,1%). В Приморском крае на уровень заболеваемости ТБ наибольшее влияние оказывают только два фактора: дефицит площади жилья и возрастающее потребление алкоголя, а на показатель смертности - снижение доли мяса и мясопродуктов в рационе питания, влияние которого составило 67%.

### **Обсуждение**

В исследованиях последних лет получены убедительные данные о все большем влиянии социально-экономических факторов на формирование неблагополучной ситуации по ТБ. Ряд авторов отмечает, что диспропорции эпидемических показателей в различных регионах РФ связаны с уровнем их экономического развития, качеством отраслевого и государственного управления, а также с эффективностью использования ресурсов здравоохранения для реализации противотуберкулезных мероприятий [17].

Проведенное исследование показало, что разработка математических моделей, описывающих зависимости эпидемического процесса ТБ от уровня социально-экономического развития региона, может иметь важное значение для повышение результативности фтизиатрической службы. Преимуществом разработанных моделей является возможность количественной оценки влияния отдельных факторов или их комбинаций на ключевые показатели активности эпидемического процесса ТБ: заболеваемость и смертность.

Предварительная оценка информативности 21 социально-экономического фактора, включенного в модели исследования, показала, что только 12 из них обладали наиболее высокой прогностической ценностью. В дальнейшем именно эти факторы использовались в построении итоговых моделей. Результаты исследования показали, что наибольшее влияние на заболеваемость ТБ оказывают факторы, характеризующие условия проживания населения, совокупный вклад которых в ее динамику составляет около 38%. Ведущими социально-экономическими факторами, влияющими на смертность от ТБ, были дефицит площади жилья, высокий уровень потребления крепкого алкоголя, недостаточное потребление мясопродуктов и безработица. Наиболее существенным из них был дефицит площади жилья с долей влияния на смертность в 16,2%. Взаимосвязь дефицита площади жилья и заболеваемости возрастает по мере усиления напряженности эпидемической ситуации в регионах: от первого кластера к четвертому, а роль этого фактора варьирует от 17% до 79% соответственно. Скученность проживания студентов в общежитиях оказывает значительное влияние на заболеваемость, которое оценивается от 10% до 16% в различных кластерах. В то же время улучшение условий проживания за счет обеспечения жилья водопроводом позволяет значительно снизить заболеваемость ТБ. В тех регионах, где такие мероприятия проводились, их вклад в снижение заболеваемости достигал 18,5%. Влияние потребления алкоголя на заболеваемость и смертность от ТБ было выявлено в 3-х кластерах. При этом его роль в эпидемическом процессе в различных кластерах варьировала от 0,5 до 2,6% по показателю заболеваемости и от 1% до 13,7% - по показателю смертности. Потребление в достаточном количестве белков животного происхождения ограничивает рост заболеваемости и смертности от ТБ. Влияние этого показателя на заболеваемость имело место только в первом кластере, а на смертность - во втором и третьем.

Разнообразие социально-экономических факторов, влияющих на основные индикаторы эпидемического процесса, сокращается по мере роста его напряженности. Если в первом кластере на заболеваемость влияли 7 факторов, а на уровень смертности – 4, то в четвертом кластере только 1 фактор влиял на заболеваемость и 2 - на смертность. При этом интенсивность их воздействия на эпидемический процесс существенно возрастает. Полученные данные указывают на то, что в регионах с относительно не тяжелым бременем ТБ, наибольшее влияние на эпидемический процесс оказывают социально-экономические факторы, относящиеся к категориям экономического развития, условиям проживания и качеству питания. В регионах с тяжелым бременем ТБ спектр социально значимых факторов расширяется за счет включения в их состав социальных рисков.

Важно отметить, что почти все показатели экономического развития (ВРП, доход на душу населения и безработица) не оказывали существенного влияния на заболеваемость ТБ, но доминировали в своем воздействии на смертность. Так, значение безработицы в увеличении уровня смертности от ТБ среди регионов четвертого кластера оценивалась в 62,4%. Единственный регион, где заболеваемость ТБ тесно ассоциировалась с низким уровнем ВРП, была Республика Тыва. Роль этого фактора составляла 62%, а обеспеченность фтизиатрами - 12%. Необходимо также отметить “диагностическое” значение коэффициентов детерминации, позволяющих оценивать суммарные эффекты воздействия анализируемых факторов на заболеваемость и смертность в каждом кластере. Так, максимальные величины этих коэффициентов фиксировались в четвертом кластере, что свидетельствует о превалирующем влиянии социально-экономических условий на эпидемический процесс в этих регионах.

Таким образом, современные методы машинного обучения позволяют разрабатывать модели, с помощью которых возможно оценить в количественном выражении значение социально-экономических условий для эпидемического процесса ТБ в конкретном регионе. Дифференцированный подход к анализу этих взаимосвязей позволяет выявить наиболее важные факторы, влияющие на эпидемическую ситуацию в регионах, и конкретизировать мероприятия по ее улучшению.

**Список литературы**

1. Ladefoged K., Rendal T., Skifte T., Andersson M., Soborg B., Koch A. Risk factors for tuberculosis in Greenland: case-control study. *Int J Tuberc Lung Dis*. 2011; 15( 1 ): 44-53.
2. Lönnroth K., [Jaramillo E.](https://elibrary.ru/author_items.asp?refid=377436532&fam=Jaramillo&init=E), [Williams B. G.](https://elibrary.ru/author_items.asp?refid=377436532&fam=Williams&init=B+G), et al. Drivers of tuberculosis epidemics: The role of risk factors and social determinants*. Soc Sci Med*. 2009; 68: 2240-2246.
3. Гельцер Б.И., Шахгельдян К.И., Гмарь Д.В., Аршинов А.Ю., Медведев В.И., Транковская Л.В. Доступность фтизиатрической помощи в Приморском крае. *Проблемы социальной гигиены, здравоохранения и истории медицины*. 2017; 4: 229-233.
4. Przybylski G., Dabrowska A., Pilaczynska-Cemel M., Krawiecka D. Unemployment in TB patients - ten-year observation at Regional Center of Pulmonology in Bydgoszcz, Poland. *Med Sci Mon*. 2014; 20: 2125-2131.
5. [Lönnroth K.](https://elibrary.ru/author_items.asp?refid=377436538&fam=L%C3%B6nnroth&init=K), [Williams B. G.](https://elibrary.ru/author_items.asp?refid=377436538&fam=Williams&init=B+G), [Jaramillo E.](https://elibrary.ru/author_items.asp?refid=377436538&fam=Jaramillo&init=E) et al. Alcohol use as a risk factor for tuberculosis -a systematic review. [*BMC Public Health*](https://elibrary.ru/contents.asp?titleid=13747). 2008; 8: 289.
6. de Castro BD, Costa Pinto R, Cláudio de Albuquerque B, Sadahiro M, Ueleres Braga J. The Socioeconomic factors and the indigenous component of tuberculosis in Amazonas. *PLOS ONE*. 2016; 11( 6).
7. Nahid P, Horne DJ, Jarlsberg LG, Reiner AP, et al. Racial Differences in Tuberculosis Infection in United States Communities: The Coronary Artery Risk Development in Young Adults Study. *Clin Infect Dis*. 2011; 53( 11): 291-294.
8. Paulino J, Martins A, Machado M, Gomes M, Gaio AR, Duarte R. Tuberculosis in native- and foreign-born populations in Portugal. *Int J Tuberc Lung Dis*. 2016; 20( 3 ): 357-362.
9. Васильева И.А., Белиловский Е.М., Борисов С.Е., Стерликов С.А. Глобальные отчеты всемирной организации здравоохранения по туберкулезу: формирование и интерпретация. *Туберкулез и болезни легких*. 2017; 5: 7‒ 16.
10. Baltagi B.H. *Econometric Analysis of Panel Data*. 5th ed.: John Wiley and Sons, Chichester; 2013.
11. Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. *Methods Ecology and Evolution*. 2010; 1( 1): 3–14.
12. Flach P. *Machine Learning. The Art and Science of Algorithms that make Sense of Data*. Cambridge University Press, Cambridge; 2012.
13. Davies D.L., Bouldin D.W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979; 1( 2): 224-227.
14. Groemping U. Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software*. 2007; 17( 1).
15. Краснов В.А., Ревякина О.В., Филимонов П.Н., Степанов Д.В., Мальцев А.В. Влияние отдельных демографических и географических факторов на оказание противотуберкулезной помощи в регионах Сибири и Дальнего Востока. *Туберкулез и болезни легких*. 2012; 8: 010-015.
16. Oren E., Winston C., Pratt R., Robinson V., Narita M. Epidemiology of Urban Tuberculosis in the United States, 2000-2007. *Am J Public Health*. 2011; 101( 7 ): 1256-1263.
17. Краснов В.А., Ревякина О.В., Филимонов П.Н., Степанов Д.В. Туберкулез: общие закономерности эпидемического процесса в России и за Уралом. *Туберкулез и болезни легких*. 2016; 10: 5-11.

**REFERENCES**

1. Ladefoged K., Rendal T., Skifte T., Andersson M., Soborg B., Koch A. Risk factors for tuberculosis in Greenland: case-control study*. Int J Tuberc Lung Dis*. 2011; 15( 1 ): 44-53.
2. Lönnroth K., [Jaramillo E.](https://elibrary.ru/author_items.asp?refid=377436532&fam=Jaramillo&init=E), [Williams B. G.](https://elibrary.ru/author_items.asp?refid=377436532&fam=Williams&init=B+G), et al. Drivers of tuberculosis epidemics: The role of risk factors and social determinants. *Soc Sci Med*. 2009; 68: 2240-2246.
3. Gel'tser B.I., Shakhgel'dyan K.I., Gmar D.V., Arshinov A.Yu., Medvedev V.I., Trankovskaya L.V. The accessibility of phthisiologic in the Primorskii kraii*. Problemy social'noj gigieny, zdravoohranenija i istorii mediciny*. 2017; 4: 229-233. (in Russian)
4. [Lönnroth K.](https://elibrary.ru/author_items.asp?refid=377436538&fam=L%C3%B6nnroth&init=K), [Williams B. G.](https://elibrary.ru/author_items.asp?refid=377436538&fam=Williams&init=B+G), [Jaramillo E.](https://elibrary.ru/author_items.asp?refid=377436538&fam=Jaramillo&init=E) et al. Alcohol use as a risk factor for tuberculosis -a systematic review*.* [*BMC Public Health*](https://elibrary.ru/contents.asp?titleid=13747). 2008; 8: 289.
5. de Castro BD, Costa Pinto R, Cláudio de Albuquerque B, Sadahiro M, Ueleres Braga J. The Socioeconomic factors and the indigenous component of tuberculosis in Amazonas. *PLOS ONE*. 2016; 11( 6).
6. Przybylski G., Dabrowska A., Pilaczynska-Cemel M., Krawiecka D. Unemployment in TB patients - ten-year observation at Regional Center of Pulmonology in Bydgoszcz, Poland. *Med Sci Mon*. 2014; 20: 2125-2131.
7. Nahid P, Horne DJ, Jarlsberg LG, Reiner AP, et al. Racial Differences in Tuberculosis Infection in United States Communities: The Coronary Artery Risk Development in Young Adults Study. *Clin Infect Dis*. 2011; 53( 11): 291-294.
8. Paulino J, Martins A, Machado M, Gomes M, Gaio AR, Duarte R. Tuberculosis in native- and foreign-born populations in Portugal. *Int J Tuberc Lung Dis*. 2016; 20( 3 ): 357-362.
9. Vasil'eva I.A., Belilovskiy E.M., Borisov S.E., Sterlikov S.A. WHO Global Tuberculosis Reports: Compilation and Interpretation. *Tuberkulez i bolezni legkikh.* 2017; 5: 7‒ 16. (in Russian)
10. Baltagi B.H. *Econometric Analysis of Panel Data*. 5th ed.: John Wiley and Sons, Chichester; 2013.
11. Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. *Methods Ecology and Evolution*. 2010; 1( 1): 3–14.
12. Flach P. *Machine Learning. The Art and Science of Algorithms that make Sense of Data*. Cambridge University Press, Cambridge; 2012.
13. Davies D.L., Bouldin D.W. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1979; 1( 2): 224-227.
14. Groemping U. Relative Importance for Linear Regression in R: The Package relaimpo. *Journal of Statistical Software*. 2007; 17( 1).
15. Krasnov V.A., Revyakina O.V., Filimonov P.N., Stepanov D.V., Mal'tsev A.V. Influence of Some Demographic and Geographic Factors on Antituberculosis Care in the Regions of Siberia and The Far East*. Tuberkulez i bolezni legkikh.* 2012; 8: 010-015. (in Russian)
16. Oren E., Winston C., Pratt R., Robinson V., Narita M. Epidemiology of Urban Tuberculosis in the United States, 2000-2007. *Am J Public Health.* 2011; 101( 7 ): 1256-1263.
17. Krasnov V.A., Revyakina O.V., Filimonov P.N., Stepanov D.V. Tuberculosis: Common Consistent Epidemic Patterns in Russia and Behind The Urals. *Tuberkulez i bolezni legkikh.* 2016; 10: 5-11. (in Russian)

### **Список авторов**

1. Шахгельдян Карина Иосифовна, д.т.н., директор Института информационных технологий Владивостокского государственного университета экономики и сервиса, заведующий лабораторией “Анализ больших данных в медицине и здравоохранении” Школы биомедицины Дальневосточного федерального университета. Почтовый адрес: Россия, 690014, Владивосток, Гоголя 41, ВГУЭС, carinash@vvsu.ru, +7924-231-44-91
2. Гельцер Борис Израйльевич, д.м.н., профессор, член-корр. РАН, директор департамента клинической медицины Школы биомедицины Дальневосточного федерального университета. Почтовый адрес: Россия, 690091, г. Владивосток, ул. Суханова, 8., ДВФУ, Школа биомедицины, e-mail: boris.geltser@vvsu.ru, +7914-703-38-04
3. Емцева Елена Дмитриевна, к.ф.-м.н., доцент кафедры Математики и моделирования Владивостокского государственного университета экономики и сервиса. Почтовый адрес: Россия, 690014, Владивосток, Гоголя 41, ВГУЭС, emtseva@mail.ru, +7902-482-28-55.
4. Середа Виктор Григорьевич, к.м.н., доцент департамента фундаментальной и клинической медицины школы Биомедицины Дальневосточного федерального университета. Почтовый адрес: Россия, 690091, г. Владивосток, ул. Суханова, 8., ДВФУ, Школа биомедицины, e-mail: drsereda@mail.ru, +7914-792-21-12.
5. Мазелис Лев Соломонович, д.э.н., заведующий кафедрой Математики и моделирования Владивостокского государственного университета экономики и сервиса. Почтовый адрес: Россия, 690014, Владивосток, Гоголя 41, ВГУЭС, lev.mazelis@vvsu..ru, +7902-524-62-54.
6. Сахарова Ольга Борисовна, к.м.н., доцент департамента клинической медицины Школы биомедицины Дальневосточного федерального университета. Почтовый адрес: Россия, 690091, г. Владивосток, ул. Суханова, 8., ДВФУ, Школа биомедицины, e-mail: sakharova.ob@dvfu.ru, +7924-320-54-14

Для переписки: Шахгельдян Карина Иосифовна, Почтовый адрес: Россия, 690014, Владивосток, Гоголя 41, ВГУЭС, carinash@vvsu.ru, +7924-231-44-91