

КОНСТРУИРОВАНИЕ ЭМПИРИЧЕСКИХ ПОДХОДОВ ДЛЯ ОПРЕДЕЛЕНИЯ АНОМАЛИЙ В МЕДИЦИНСКИХ СТАТИСТИЧЕСКИХ ПРЕДОСТАВЛЕННЫХ

Аннотация. Данное исследование охватывает анализ аномалий в показателях смертности распространяясь на широкий спектр заболеваний согласно Международной классификации болезней. С помощью муниципальных данных и микроданных, исследование применяет эконометрическое моделирование для выявления и объяснения причин высоких показателей смертности. В исследовании выделены аномалии различного класса, а также эмпирические способы их поиска на основе реальных данных. На основе полученных результатов создан системы поддержки принятия управленческих решений в сфере здравоохранения на основе данных для профильных органов государственной власти.

Ключевые слова: аномалии в медицинских данных, смертность от заболеваний, эконометрическое моделирование, муниципальные данные, микроданные, управленческие решения, система здравоохранения.

Subbotovsky Dmitry

DESIGN AND EMPIRICAL APPROACHES TO IDENTIFY ANOMALIES IN MEDICAL STATISTICAL DATA

Annotation. This study covers the analysis of anomalies in mortality rates extending to a wide range of diseases according to the International Classification of Diseases. Using municipal data and microdata, the study applies econometric modelling to identify and explain the causes of high mortality rates. The study highlights anomalies of different class and empirical ways to find them based on real data. On the basis of the results obtained, a data-driven management decision support system in the field of health care is created for the relevant public authorities.

Keywords: Anomalies in medical data, disease mortality, econometric modelling, municipal data, microdata, management decisions, health care system.

В 2020 год смертность от сердечно сосудистых заболеваний в Дальнереченском муниципальном районе Приморского края (ПК) составила 238.8 на 10000 человек, тогда как в среднем в муниципалитетах региона величина данного показателя была ниже 81 со стандартным отклонением 35.4. При этом сразу в трех районах уровень смертности был ниже 50, например, 47.7 в Спасском муниципальном районе.

Исследовательская проблема заключается в необходимости объяснения того, почему на той или иной территории наблюдается соответствующий уровень смертности от конкретных направлений заболеваний (Международная классификация болезней (МКБ-10)), т.е. соответствующей вариации. Так, с одной стороны, высокий уровень смертности может быть обусловлен влиянием соответствующих внутренних факторов, характеризующих территорию, например, таких как социально-экономические факторы (уровень доходов, образование), демографические факторы (половозрастной состав), экологические характеристики территории или внешних факторов, например, влияние соседних территорий. С другой стороны, вариация может объясняться эффективностью функционирования сложившейся на территории системы здравоохранения, как в части действующей инфраструктуры для оказания медицинского обслуживания населения, так и результативностью реализации различных программ, направленных на ее улучшение, например, призванных обеспечить сокращение показателей смертности населения от конкретных заболеваний. Но какие именно факторы, как и в какой степени повлияли на наблюдаемые показатели на конкретной территории? Речь идет о подходах к обнаружению и объяснению аномалий в медицинских данных. Под аномалией понимается необъясненное отклонение наблюдаемых показателей от ожидаемых, другими словами, речь идет о противоречии между теоретическим предсказанием и эмпирическими данными.

Выявление и объяснение аномалий в медицинских данных позволит профильным органам региональной власти:

1. Определять настоящие причины того почему на той или иной территории наблюдается соответствующий показатель заболеваемости или смертности. Территория, на которой высокие показатели смертности из-за социально-экономических факторов может ну-

ждаться в мероприятиях в рамках соответствующих государственных программ по борьбе с заболеваниями, совершенно отличных от мероприятия, необходимых на территории, где высокие показатели смертности вызваны половозрастными показателями.

2. Объяснять возможные причины неэффективности существующих мер (например, ввиду того, что они воздействуют не на те факторы, которые вносят основной вклад в наблюдаемый уровень соответствующих показателей) и формировать правки.

3. Оценивать эффективность функционирования системы здравоохранения на территории при прочих равных условиях (социально-экономические, демографические, экологические и иные условия).

4. Адресно определять участки (вплоть до улиц), требующие адресного выяснения причин аномалий (например, посредством выездных проверок) и реализацию мер управленческого воздействия (повышение квалификации управленцев, врачей и другие кадровые решения, закупка и наладка работы оборудования, проведение профилактических мероприятий).

Цель, задачи и гипотезы исследования

Цели: моделирование показателей смертности в муниципалитетах ПК для выявления и объяснения аномалий, необходимых в разработке системы поддержки принятия управленческих решений в интересах Министерства Здравоохранения ПК и ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава РФ.

Задачи:

1. Сбор, систематизация и подготовка к моделированию данных
2. Конструирование теоретических и эмпирических подходов для определения аномалий различных классов и их последующего объяснения (в том числе в формате рекомендационной системы).

Гипотезы исследования:

1. Рейтинг территорий характеризующий эффективность функционирования системы здравоохранения на отдельных территориях региона, в том числе определяющий приоритетность внимания профильных органов власти, сформированный на основе эконометрического моделирования показателей заболеваемости и смертности с учетом оценки аномалий и их объяснения на основе влияния внешних и внутренних факторов будет отличаться от общепринятых рейтингов, основанных на сравнении относительного уровня соответствующих показателей.

2. Эконометрическое моделирование показателей заболеваемости и смертности с учетом оценки аномалий и их объяснение на

основе влияния внешних и внутренних факторов и возможности органов власти влиять на них продемонстрирует необходимость направления расходования средств в контексте территорий и соответствующих мер отличные от существующих во многом стандартных, не учитывающих особенности местности с учетом современного исследовательского инструментария мероприятий в существующих государственных программах в сфере здравоохранения.

Обзор литературы

На сегодняшний день в литературе представлен широкий круг исследований, описывающих влияние отдельных социально-экономических, демографических, экологических и иных факторов на показатели заболеваемости и смертности от заболеваний согласно Международной классификации болезней. Особенно можно выделить влияние таких факторов, как пол [1,2,3,4,5] возраст [6,7,8,9] урбанизация [10,11], экология [12], доходы [13,14] образование [15] обеспеченность профильными врачами [16]. Так в вышеприведенных исследованиях, авторы часто демонстрируют как применение тех или иных квазиэкспериментальных методов позволяет выяснить истинное причинно-следственное влияние конкретного фактора на моделируемый показатель, например, смертности от онкологических заболеваний [17]. Проведём сопоставительный обзор литературы по 6 направлениям заболеваний (МКБ-10), на примере которых будет происходить выявление аномалий в нашей работе.

В таблице 1 приведено описание влияния характеристик территорий на показатели смертности от различных направлений заболеваний. Знак «-» означает отрицательное ожидаемое влияние фактора на соответствующее заболевание, «+» положительное ожидаемое влияние фактора, «+/-» противоречивые результаты, пустые ячейки отсутствие информации о влиянии данного фактора либо его незначимость в большинстве исследований с учетом изученной литературы.

Таблица 1. Описание влияния характеристик территорий на показатели смертности от различных направлений заболеваний

Фактор	Инфекционные и паразитарные заболевания	Онкологические заболевания	Заболевания системы кровообращения	Заболевания органов дыхания	Заболевания органов пищеварения	Травмы, отравления
Доход	-	-	-	-	+/-	-

Пол Ж>М	-	-	-	-	-	-
Возраст	+	+	+	-	+	+
Курение	-	+	+	+	+	
Алкоголь	+	+	+			+
Здоровый Образ жизни	-	-	-	-	-	
Мед. Персонал	-	-	-	-	-	-
Соц. Факторы (брач- ность)	+/-	-	-	+/-	-	+

Источник: составлено авторами

Пол является одним из самых важных детерминантов риска и смертности от заболеваний, так как он обуславливает многие факторы, влияющие на здоровье. Например, рак имеет более высокую заболеваемость и смертность среди мужчин, чем среди женщин, что может быть связано с половыми гормонами, пищевыми привычками и практикой скрининга. Как отмечают Барайбар et al [1], половые и гендерные различия раке многофакторны и включают биологические, экологические и поведенческие факторы, которые влияют на развитие и прогрессирование заболевания. Аналогично, респираторные и инфекционные [18,19] заболевания имеют более высокую нагрузку среди мужчин, чем среди женщин, особенно в странах с низким и средним уровнем дохода, где мужчины более подвержены инфекции, менее склонны обращаться за диагностикой и лечением, и более вероятно имеют сопутствующие заболевания, такие как ВИЧ и курение. Согласно Хортон et al [5], нагрузка туберкулеза выше среди мужчин, чем среди женщин, с оценочными 60% случаев заболевания и 64% смертей, происходящих среди мужчин в 2015 году. С другой стороны, рак молочной железы распространен и смертелен среди женщин, а курение может повысить риск рака молочной железы у женщин, особенно до менопаузы. Бьеркаас et al [17], сообщают, что женщины, которые курили не менее 2 сигарет в день, имели на 20% повышенный риск рака молочной железы по сравнению с никогда не курившими.

Возраст - еще один ключевой фактор, влияющий на возникновение и исход определенных заболеваний, так как он отражает снижение иммунной системы, функции органов и механизмов клеточного ремонта. Например, время до диагностики туберкулеза больше у пожилых пациентов, что может привести к задержке лечения и худшим результатам. Аббара et al [6] обнаружили, что медианное время до диагностики составляло 28 дней у пациентов в возрасте ≥ 65 лет по сравнению с 14 днями у тех, кто был моложе 65 лет. Точно так же заболеваемость и смертность от инфаркта миокарда (сердечного приступа) увеличиваются с возрастом, особенно у пожилых женщин. Кристенсен et al [7] отмечают, что заболеваемость болезнями крови и сердца увеличивалась с возрастом у обоих полов, но более резко у женщин, чем у мужчин. Уровень смертности был выше у женщин, чем у мужчин во всех возрастных группах. Кроме того, травмы и инфекционные заболевания также более распространены и смертельны среди пожилых людей, иммунная система которых ослабевает при увеличении возраста.

Социально-экономический статус - еще один важный фактор, влияющий на состояние здоровья и доступность медицинской помощи у индивидов и населения. Например, пациенты с туберкулезом в Республике Македония сталкиваются с неравенством в доступе к медицинской помощи из-за бедности, безработицы, низкого уровня образования и отсутствия медицинского страхования. Никовска и Тозиц [2] сообщают, что пациенты с заболеваниями органов дыхания были более склонны к бедности, безработице, низкому уровню образования и отсутствию страховки по сравнению с контрольной группой. Клегг et al [20] отмечают, что анализ выявил значительную корреляцию между уровнем смертности от онкологии и уровнем дохода, образования, безработицы и степенью урбанизации населения.

Подводя итог обзору литературы стоит отметить, что для дальнейшего моделирования смертности по выбранным ранее шести классам заболеваний будут использованы некоторые общие для всех моделей факторы, такие как доход, доля сельского населения по полу и возрасту, возраст, половой состав, брачность и разводимость, как детерминанты социального благополучия, обеспеченность региона врачами, медсестрами, мощность имеющихся госпиталей, вместимость существующих мед учреждений, соответствующие коэффициенты смертности и рождаемости.

Данные и модели

Данные

В исследовании рассматриваются ежегодные данные для 31 муниципалитета Приморского края, для которых есть соответствующие ежегодные наблюдения в период с 2011 по 2020 гг. Описание переменных приведено в таблице 2.

Таблица 2. Описание переменных

№	Обозначение	Фактор	Способ расчета
1	dheart_p	Смертность от ССЗ	Количество смертей на 1000 чел. населения
2	dneoplasms_p	Смертность от новообразований	Количество смертей на 1000 чел. населения
3	dbreath_p	Смертность от заболеваний легких	Количество смертей на 1000 чел. населения
4	ddigestion_p	Смертность от заболеваний ЖКТ	Количество смертей на 1000 чел. населения
5	dinfections_p	Смертность от инфекций	Количество смертей на 1000 чел. населения
6	dexternalres_p	Смертность от внешних причин	Количество смертей на 1000 чел. населения
7	doctors	Врачи	Количество на 1 чел.
8	hospbed	Койки	Число больничных коек (на 10000 чел. насел.)
9	city	Урбанизация	Доля городского населения
10	incom	Доходы	Среднемесячная номинальная начисленная заработная плата по муниципальным образованиям (без субъектов малого предпринимательства)
11	sex	Пол	Количество женщин на 1 мужчину
12	old	Доля пожилых	Доля лиц старше трудоспособного возраста
13	marriage_p	Брачность	Количество на 1 чел.
14	divorce_p	Разводимость	Количество на 1 чел.
15	ill	Заболеваемость	Количество заболевших на 1000 чел.

Уравнения регрессии:

Для проверки сформированных ранее гипотез будут использованы модели с пространственно-авторегрессионными эффектами (spatial autoregression) на основе фиксированных эффектов (m1).

$$\text{dheart_p}_{it} = \beta_0 + \alpha_1 \rho * W * \text{dheart_p}_{it} + \beta_1 * \text{doctors}_{it} + \beta_2 * \text{hospbed}_{it} + \beta_3 * \text{city}_{it} + \beta_4 * \text{income}_{it} + \beta_5 * \text{sex}_{it} + \beta_6 * \text{old}_{it} + \beta_7 * \text{marriage_p}_{it} + \beta_8 * \text{ill}_{it} + \epsilon_{it} \quad (m1)$$

где W – матрица, характеризующая пространственную компоненту в модели, ρ – коэффициент, отражающий наличие пространственных эффектов.

В данном случае можно воспользоваться возможностью оценки соответствующих фиксированных эффектов территорий. Именно данный показатель позволит привести характеристики территорий «к общему знаменателю» и даст возможность сравнить сложившиеся там показатели без учета их влияния (Смертность от ССЗ, Смертность от новообразований, Смертность от заболеваний легких, Смертность от заболеваний ЖКТ, Смертность от инфекций, Смертность от внешних причин).

Результаты

Полученные результаты позволяют подтвердить авторские гипотезы. Сформированные на их основе цифровые сервисы, которые уже доступны (<https://datalens.yandex/b4ej7g6pbtvuz?tab=0A>) для использования заинтересованными пользователями, позволяют увидеть, что рейтинг регионов на основе оценок на основе фиксированных эффектов (т.е. уровня смертности от соответствующего направления заболевания при прочих равных условиях) существенно отличается от рейтинга регионов, сформированных на основе исходных показателей смертности. При этом результаты факторного анализа могут объяснить данную вариацию в части интерпретации вклада различных факторов в формирование итоговых показателей и их отличие для муниципалитета от среднерегионального уровня.

Выводы и рекомендации

В исследовании осуществлено моделирование показателей смертности в муниципалитетах ПК для выявления и объяснения аномалий, необходимых в разработке системы поддержки принятия управленческих решений в интересах Министерства Здравоохранения ПК и ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава РФ.

Получили предварительное подтверждение исследовательские гипотезы. В частности рейтинг территорий характеризующий эффективность функционирования системы здравоохранения на отдельных территориях региона, в том числе определяющий приоритетность внимания профильных органов власти, сформированный на основе эконометрического моделирования показателей заболеваемости и смертности с учетом оценки аномалий и их объяснения на основе влияния внешних и внутренних факторов будет отличаться от общепринятых рейтингов, основанных на сравнении относительного уровня соответствующих показателей (таблица 3 и 4). Несмотря на то, что интерпретировать значения коэффициентов нельзя в логике причинно-следственных связей, все же видна вариация их вклада в соответствующие показатели, что и объясняет необходимость их учета при планировании тех или иных программ развития (подтверждение гипотезы 2).

На основе полученных результатов сформирован прототип аналитического дашборда: <https://datalens.yandex/b4ej7g6pbvtvuz> (будет дополняться)

Основной научный результат заключается в разработке алгоритмов поиска и объяснения аномалий в медицинских данных, отличающихся от уже существующих использованием и сопоставлением нескольких семейств подходов (сравнение со средними, эконометрическое моделирование показателей и их сопоставление с реальными наблюдениями, работа с паттернами и фиксированными эффектами) и применения микроданных, предоставленных МИАЦ и сотовыми операторами, позволяющих выявлять аномалии на уровне произвольно сформированных территорий, для которых отсутствуют агрегированные статистические данные (поселки, села, поселения, жилищные комплексы, улицы и др.).

Со временем будет накапливаться база данных, в которой будут храниться:

- Алгоритмы выявления и объяснения аномалий в данных МИАЦ (симптомы управленческой болезни)
- Объяснения или подходы к объяснению аномалий, в том числе тех, которые были выявлены на основе данных, но объяснены, например, посредством выездных проверок (диагноз управленческой болезни)
- Управленческие решения, которые были приняты на конкретной территории (лечение управленческой болезни)
- Результаты реализации конкретных управленческих решений

Библиографический список

1. Baraibar I., Ros J., Saoudi N. и др. Sex and gender perspectives in colorectal cancer // *ESMO Open*. 2023. Т. 8, № 2. С. 101204.
2. Nikovska G., Tozija F. Social determinants of equity in access to healthcare for tuberculosis patients in Republic of Macedonia – results from a case-control study // *International Journal of Health Policy and Management*. 2014. Т. 3, № 4. С. 199–205.
3. Amiri L., Torabi M., Deardon R., Pickles M. Spatial modeling of individual-level infectious disease transmission: tuberculosis data in Manitoba, Canada // *Stat Med*. 2021. Т. 40, № 7. С. 1678–1704.
4. Chung-Delgado Kocfa, Revilla-Montag Alejandro, Guille'n-Bravo Sonia, Bernabe-Ortiz Antonio Weight variation over time and its relevance among multidrug-resistant tuberculosis patients // *International Journal of Infectious Diseases*. 2014. Т. 23.
5. Horton Katherine C., Sumner Tom, Houben Rein M. G. J. A Bayesian Approach to Understanding Sex Differences in Tuberculosis Disease Burden // *Practice of Epidemiology*. 2018. Т. 187, № 11.
6. Abbara A., Simon M. Collin, Onn M. Kon Time to diagnosis of tuberculosis is greater in older patients: a retrospective cohort review // *ERJ Open Res*. 2019. Т. 5, № 4.
7. Christensen D. M., Strange J. E. и др. Age- and sex-specific trends in the incidence of myocardial infarction in Denmark, 2005 to 2021 // *Atherosclerosis*. 2022. Т. 346.
8. Fox Gregory J., Barry Simone E., Britton Warwick J., Marks Guy B. Contact investigation for tuberculosis: a systematic review and meta-analysis // *European Respiratory Journal*. 2013. Т. 41, № 1. С. 140–156.
9. Cluver L., Orkin M., Moshabela M., Kuo C., Boyes M. The hidden harm of Home-Based Care: Pulmonary tuberculosis symptoms among children providing home medical care to HIV/ AIDS-affected adults in South Africa // *NIH Public Access*. 2013. Т. 25, № 6. С. 748–755.
10. Afshar N., English D.R., Milne R.L. Milne Rural–urban residence and cancer survival in high-income countries: A systematic review // *Cancer*. 2019. Т. 125, № 13. С. 2172–2184.
11. Бойцов С. А., Болотова Е. В., Самородская И. В., Иноземцев Е. С. Взаимосвязь региональных уровней смертности от ишемической болезни сердца с социально-экономическими факторами в Российской Федерации // *Проблемы социальной гигиены, здравоохранения и истории медицины*. 2016. Т. 24, № 2. С. 68–73.
12. Шахгельдян К. И., Гельцер Б. И., Гмарь Д. В. и др. Проблемы анализа данных медицинской статистики // *Проблемы социальной*

гигиены, здравоохранения и истории медицины. 2018. Т. 26, № 3. С. 132–136.

13. Hagen T. P., Häkkinen U., Iversen T. и др. Socio-economic inequality in the use of procedures and mortality among ami patients: quantifying the effects along different paths // Health Econ. 2015. Т. 2. С. 102–115.

14. Arnold M., Rutherford M.J., Bardot A. и др. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study // The Lancet Oncology. 2019. Т. 20, № 11. С. 1493-1505.

15. Vaccarella S., Georges D., Bray F. и др. Socioeconomic inequalities in cancer mortality between and within countries in Europe: a population-based study // The Lancet Regional Health – Europe. 2023. Т. 25. С. 100551.

16. Nagapetyan A., Drozd A., Subbotovsky D. How to determine the optimal number of cardiologists in the region? // Mathematics. 2023. Т. 11, № 21. С. 4422.

17. Bjerkaas E., Parajuli R., Engeland A. и др. Social inequalities and smoking-associated breast cancer – Results from a prospective cohort study // Preventive Medicine. 2015. Т. 73. С. 125-129.

18. World Health Organization. Chronic respiratory diseases // World Health Organization. 2020. (<https://www.who.int/health-topics/chronic-respiratory-diseases>)

19. Alcaide J., Altet N., Plans P. et al. Cigarette smoking as a risk factor for tuberculosis in young adults: a case-control study // Tubercle and Lung Disease. 1996. Vol. 77, No. 2. P. 112-116. ([https://doi.org/10.1016/S0962-8479\(96\)90019-8](https://doi.org/10.1016/S0962-8479(96)90019-8))

20. Clegg, Limin X et al. Impact of socioeconomic status on cancer incidence and stage at diagnosis: selected findings from the surveillance, epidemiology, and end results: National Longitudinal Mortality Study. Cancer causes & control : CCC vol. 20,4 (2009): 417-35. doi:10.1007/s10552-008-9256-0

Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации, проект № FZNS-2023-0016 «Устойчивое развитие региона: эффективные экономические механизмы организации рынков и предпринимательские компетенции населения в условиях неопределенности (баланс безопасности и риска)».