

**А.П. Романова,
Белорусская медицинская академия
последипломного образования (Минск);
О.В. Агиевец,
Белорусская медицинская академия
последипломного образования (Минск)**

О МЕТОДОЛОГИИ «СЛЕДОВАНИЯ ЗА ДАННЫМИ» В ИЗУЧЕНИИ ПОКАЗАТЕЛЕЙ ОБЩЕСТВЕННОГО ЗДОРОВЬЯ

Приведены результаты исследования динамики смертности населения Республики Беларусь в долгосрочной ретроспективе с использованием методологии «следования за данными». Представлен порядок проведения исследования, описаны использованные методы. Научно обосновано выделение на протяжении 1959-2018 гг. шести периодов динамики смертности на основании статистически значимых различий темпов прироста показателей. Обозначена возможность соотнесения изменения темпов прироста показателей смертности с совокупностью социально-экономических, медицинских и экологических событий, в рамках выделенных по результатам моделирования, а не установленных авторами, периодов.

Ключевые слова: методология следования за данными, общественное здоровье, смертность, динамика смертности, моделирование динамики смертности.

**A.P. Romanova,
Belarusian Medical Academy of Postgraduate Education;
O.V. Aguietets,
Belarusian Medical Academy of Postgraduate Education**

ABOUT THE METHODOLOGY OF “FOLLOWING THE DATA” WHEN STUDYING THE HEALTH INDICATORS

The results of a study on the long-term retrospective mortality dynamics in the Republic of Belarus using the methodology of “following the data” are presented. The procedure of conducting the study is demonstrated, with the methods used being described. Six periods of mortality dynamics during 1959-2018 have been identified and scientifically grounded based on statistically significant differences in

the growth rates of indicators. The change in the growth rate of mortality rates is likely to correlate with a set of socio-economic, medical and ecological events within the periods identified by modeling and not established by the authors.

Keywords: methodology of “following the data”, public health, mortality, mortality dynamics, modeling the mortality dynamics.

Общественное здоровье формируется под влиянием социальных условий и факторов внешней среды, а в структуре факторов риска развития заболеваний, доля образа жизни составляет 50-55 %. [1]. Понятие образа жизни как исторически обусловленный тип деятельности человека в материальной и нематериальной (духовной) сферах жизни включает четыре категории: уровень жизни (экономическая категория), качество жизни (социологическая категория), стиль жизни (социально-психологическая категория) и уклад жизни (социально-экономическая категория) [2]. Информационное развитие общества существенно расширило горизонты и возможности исследований в различных сферах жизнедеятельности человека, что способствует разработке новых подходов к анализу уровня и качества жизни, равно как и к исследованию показателей здоровья населения. Показатели общественного здоровья – медико-демографические (смертность, рождаемость, естественный прирост), заболеваемость, инвалидность и физическое развитие – относят к индикаторам социально-экономического развития государств и эффективности систем здравоохранения. Применение новых подходов и использование возможностей математической обработки совокупности показателей здоровья населения, представляющих собой большие массивы данных, которые традиционно рассматриваются в разрезе пола, возраста и типа территории проживания, существенно расширяют возможности их исследования и позволяют по-новому взглянуть на развитие здоровья населения.

При проведении исследования динамики показателей смертности населения Республики Беларусь в долгосрочной ретроспективе за 1959-2018 гг. электронная «База данных, содержащая абсолютную численность умерших и среднегодовую численность населения БССР и Республики Беларусь по 5-летним возрастным группам с разделением по полу и типу территории проживания за 1959-2018 гг.», созданная в программе Excel, включила 23 760 значений в абсолютных значениях [3].

Исследование опиралось на методологию, основанную на данных (data-driven, «следование за данными»), которая является междисциплинарной областью и использует научные методы, процессы, алгоритмы и системы для извлече-

ния знаний из множества структурированных и неструктурированных данных [4; 5]. Для изучения динамики смертности в долгосрочной ретроспективе использовался метод моделирования [6; 7]. Временной ряд показателя общей смертности был исследован на предмет возможных статистически значимых временных тенденций/трендов на протяжении исследуемого периода, для чего был проведен регрессионный анализ. Он включал в себя выбор варианта регрессионной модели; оценку функции связи между зависимой и независимой переменными; проверку предположений, лежащих в основе модели; анализ производительности полученной модели на основе анализа остатков. Предположение о гомоскедастичности проверялось с помощью критерия Бройша-Пагана (англ. Breusch-Pagan test), также использовался критерий Дарбина-Уотсона (англ. Durbin-Watson test) на автокорреляцию первого порядка остатков модели [8]. Анализ остатков сопровождался расчётом абсолютной и относительной ошибок: максимальной, 90-го и 50-го квантилей. Для проведения анализа временных трендов была выбрана пуассоновская линейная модель с точками перелома (кусочно-линейная модель) [9; 10]. Выбор модели кусочно-линейной регрессии, наиболее точно описывающей динамику смертности населения Беларуси на протяжении 1959-2018 гг., проводился по критерию, предложенному Н. J. Kim et al. Оптимальное количество точек перелома определялось на основании перестановочного алгоритма [11]. Бейесовский информационный критерий (Bayesian information criterion – BIC) использовался для сравнения моделей между собой на основе баланса сложности и точности [12]. Валидация выбора модели проведена по алгоритму BIC [13] и BIC3 [14; 15]. На основании коэффициентов регрессии полученных моделей рассчитывался темп ежегодного прироста изучаемого показателя в процентах для каждого хронологического периода как $100(e^{\hat{\beta}} - 1)$, который интерпретировался как процентное возрастание средней величины соответствующего показателя за год (темп ежегодного прироста в процентах, ТЕП) [16; 17].

Расчеты по оценке показателей выполнялись в статистическом пакете R, Версия 3.6 [18], с подключением библиотеки epitools []; для моделирования кусочно-линейной регрессии и расчётов трендов использовалось специализированное программное обеспечение Joinpoint Regression Program [20], а также офисный пакет MS EXCEL 2010.

Созданная математико-статистическая динамическая модель продемонстрировала статистически значимое различие ($p < 0,05$) темпа ежегодного прироста показателя общей смертности (ОПС) в длительном временном интервале, что яви-

лось основанием для выделения на протяжении 1959-2018 гг. шести хронологически последовательных периодов динамики показателя общей смертности (Рис. 1)

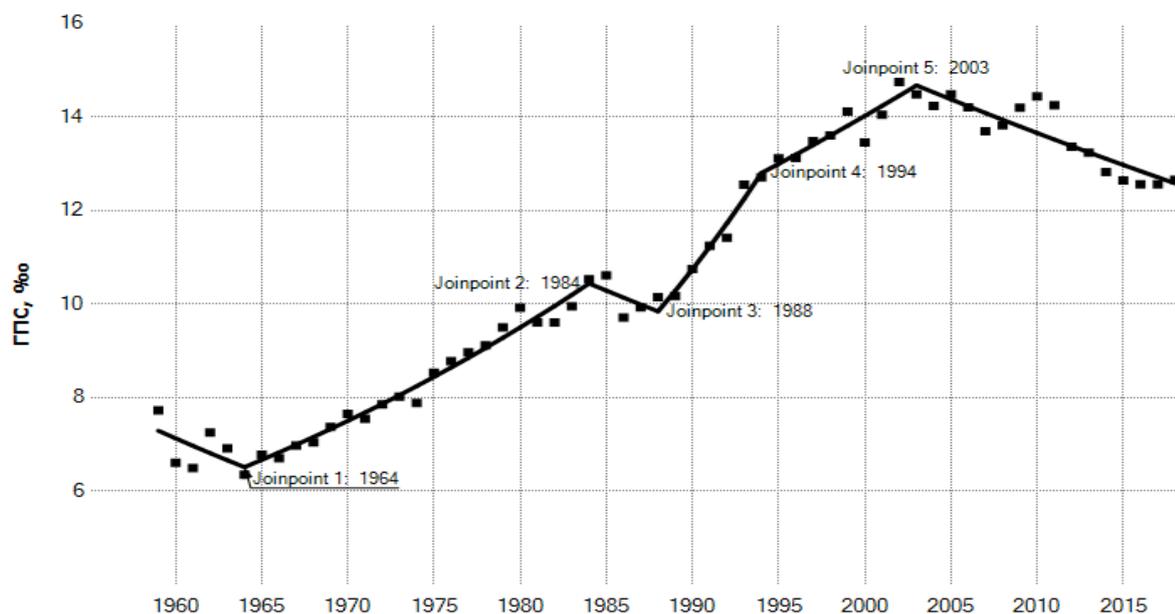


Рис. 1. Динамика общего показателя смертности в 1959-2018 гг.

Маркер – наблюдаемые ОПС, линии – модельные расчеты.

На протяжении 1959-1964 гг. ОПС населения не претерпевал статистически значимых изменений, $p = 0,071$, снижаясь в тенденции. В 1964-1984 гг. ОПС населения возрастал с ТЕР 2,4 (2,01; 2,7)%, $p < 0,001$. В течение последующих 4 лет (1984-1988 гг.) ОПС оставался статистически неизменным, $p = 0,484$. С 1988 г. до 1994 г. ОПС возрастал с ТЕР 4,5 (2,6; 6,4)%, $p < 0,001$, а затем продолжил рост в 1994-2003 гг. с ТЕР 1,5 (0,7; 2,3)%, $p < 0,001$. В 2003 году ОПС начал снижаться с ТЕР $-1,0$ ($-1,3$; $-0,7$)%, $p < 0,001$, и в 2018 году составил 12,66 (95%ДИ 12,59; 12,73)‰.

В отличие от подхода исследования динамики смертности, в хронологических периодах, выбираемых и устанавливаемых автором (например: между переписями населения, 5-летними или другой продолжительности периодами), границы установлены «от обратного», то есть границы периодов динамики смертности в долгосрочной ретроспективе формировали данные. Таким образом, использование методологии, основанной на данных (data-driven, «следование за данными»), и динамической информационной знаковой компьютерной модели позволило научно обосновать хронологические границы периодов динамики смертности населения БССР и Республики Беларусь в 1959-2018 гг. и соотнести изменения темпов прироста показателей смертности с совокупностью

социально-экономических, медицинских и экологических событий, происходивших в течение хронологических периодов.

Список литературы:

1. Лисицын Ю.П., Улумбекова Г.Э. Общественное здоровье и здравоохранение: учеб. Для студентов учреждений высш. проф. образования. 3-е изд., перераб. и доп. М.: ГЭОТАР-Медиа, 2011. 544 с.

2. Пилипцевича Н.Н., ред. Общественное здоровье и здравоохранение: учеб. пособие. Минск: Новое знание, 2015. 784 с.

3. Романова А.П. База данных содержащая абсолютную численность умерших и среднегодовую численность населения БССР и Республики Беларусь по 5-летним возрастным группам с разделением по полу и типу территории проживания за 1959-2018 гг. // База данных : зарег. в Гос. рег. информ. ресурсов, 22.06.2020, № 1052022830; Белорус. мед. акад. последиплом. образования. Электрон. дан. Минск, 2020. 1 электрон. опт. диск.

4. Dhar V. Data science and prediction // Communications of the ACM. 2013. Vol. 56, № 12. P. 64-73.

5. Leek J. The key word in «Data Science» is not data, it is science // <https://simplystatistics.org/posts/2013-12-12-the-key-word-in-data-science-is-not-data-it-is-science/>

6. Белоенко Е.Д., Беспальчук П.И., Мухля А.М. Приоритетная медицинская проблема – травматизм // Первый съезд врачей Республики Беларусь: тез. докл., Минск, 25-26 июня 1998 г. Минск, 1998. С. 86-87.

7. Кальченко Е.И. Менеджмент в здравоохранении: избранные понятия и термины. СПб.: [б.и.], 1995. 45 с.

8. Дрепер Н., Смит Г. Прикладной регрессионный анализ: в 2 кн. 2-е изд., перераб. и доп. М.: Финансы и статистика, 1986-1987. Кн. 1. 1986. 365 с.

9. Рокицкий П.Ф. Биологическая статистика. Изд. 2-е, испр. Минск: Высш. шк., 1967. 328 с.

10. Feder, P. I. On asymptotic distribution theory in segmented regression problems – identified case // The Annals of Statistics. 1975. Vol. 3, № 1. P. 49-83.

11. Kim H.J., Fay M.P., Feuer E.J., Midthune D.N. Permutation tests for joint point regression with applications to cancer rates // Statistics in medicine. 2000. Vol. 19, № 3. P. 335-351.

12. Schwarz G. Estimating the dimension of a model // The Annals of Statistics. 1978. Vol. 6, № 2. P. 461-464.

13. Kim H.J., Yu B., Feuer E.J. Selecting the number of change-points in segmented line regression // *Statistica Sinica*. 2009. Vol. 19, № 2. P. 597-609.
14. Kim J., Kim H.J. Consistent model selection in segmented line regression // *Journal of Statistical Planning and Inference*. 2016. Vol. 170. P. 106-116.
15. Zhang N.R., Siegmund D.O. A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data // *Biometrics*. 2007. Vol. 63, № 1. P. 22-32.
16. Вальчук Э.А., Гулицкая Н.И., Царук Ф.П. Основы организационно-методической службы и статистического анализа в здравоохранении. Минск: Харвест, 2007. 398 с.
17. Статистические методы анализа медико-демографических показателей // *Медицинская статистика и оргметодработа в учреждениях здравоохранения*. 2019. № 5. С. 6-26.
18. R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria // <https://www.R-project.org/>
19. Tomas J. Aragon (2017). epitools: Epidemiology Tools. R package version 0.5-10 // <https://CRAN.R-project.org/package=epitools>
20. Joinpoint regression program, version 4.8.0.1. April 2020 / Stat. research and applications branch, Nat. cancer institute.