

**Д.Б. ЕГОРОВ,**

старший преподаватель кафедры медицинской информатики и физики, ФГБОУ ВО «Тюменский государственный медицинский университет» Минздрава России, г. Тюмень, Россия, e-mail: dmdel@mail.ru

С.Д. ЗАХАРОВ,

к.ф.-м.н., доцент кафедры медицинской информатики и физики, ФГБОУ ВО «Тюменский государственный медицинский университет» Минздрава России, г. Тюмень, Россия, e-mail: s_zah@rambler.ru

А.О. ЕГОРОВА,

к.фарм.н., доцент кафедры фармацевтических дисциплин, ФГБОУ ВО «Тюменский государственный медицинский университет» Минздрава России, г. Тюмень, Россия, e-mail: alexa_kizim@mail.ru

СОВРЕМЕННЫЕ МЕТОДЫ АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ВРЕМЕННЫХ РЯДОВ И ИХ ПРИМЕНЕНИЕ В МЕДИЦИНЕ

УДК 614.2:51-76:004.9

DOI: 10.37690/1811-0193-2020-1-21-26

Егоров Д.Б., Захаров С.Д., Егорова А.О. Современные методы анализа и прогнозирования временных рядов и их применение в медицине (ФГБОУ ВО «Тюменский государственный медицинский университет» Минздрава России, г. Тюмень, Россия)

Аннотация. Статья представляет собой обзор российских и зарубежных научных публикаций, связанных с применением в медицине методов анализа и прогнозирования временных рядов. Рассмотрены 112 основных публикаций за последние 5 лет, расположенные на Интернет-ресурсах e-library и PubMed. Приведены примеры применения таких методов как экспоненциальное сглаживание, регрессионный анализ, метод ARIMA и их варианты для анализа временных рядов. Представлены различные подходы к математическому моделированию временного ряда. Результаты статьи могут применяться для выбора метода анализа и прогнозирования временных рядов в зависимости от поставленных задач.

Ключевые слова: временной ряд, анализ и прогнозирование временного ряда, экспоненциальное сглаживание, регрессионный анализ, ARIMA.

UDC 614.2:51-76:004.9

Egorov D.B., Zakharov S.D., Egorova A.O. Modern methods of analysis and forecasting of time series and use in medicine (Tyumen State Medical University)

Abstract. The article is a review of Russian and foreign scientific publications related to the use of methods of analysis and forecasting of time series in medicine. 112 major publications over the past 5 years, located on the Internet resources e-library and PubMed, are considered. Examples of the application of such methods as exponential smoothing, regression analysis, the ARIMA method and their variants for time series analysis are shown. Various approaches to mathematical modeling of the time series are presented. The results of the article can be used to select a method of analysis and forecasting time series depending on the tasks.

Keywords: time series, analysis and forecasting of the time series, exponential smoothing, regression analysis, ARIMA.

ВВЕДЕНИЕ

На сегодняшний день анализ и прогнозирование временных рядов (ВР) используется во многих научных исследованиях. Основной целью анализа ВР является получение исчерпывающей информации, выявление всех возможных зависимостей из представленных данных: общий тренд развития, циклические и сезонные процессы, а также стохастическую (случайную) составляющую.

Основное развитие анализа и прогнозирования ВР получило в эконометрике – науке, изучающей конкретные количественные и качественные взаимосвязи экономических объектов и процессов при помощи математических методов и моделей. Наиболее интенсивное развитие данная

тема получила в 70-е годы 20 века после выхода в свет книги «Time Series Analysis: Forecasting and Control», авторов Box G., Jenkins G. В своей работе американские ученые подробно описывают применение метода ARIMA (autoregressive integrated moving average), который и по сей день является актуальным. В дальнейшем исследования других авторов, основанные на применении ARIMA моделей, получили высокую оценку (Нобелевская премия по экономике Р. Энгл и К. Грэнджер, 2003 год).

Однако актуальность применения методов анализа ВР выходит далеко за рамки данной науки и успешно применяется в других научных сферах, в том числе в медицине. Если ограничиться при анализе ВР только описательной статистикой, то итоги будут слишком усредненные, а применение результатов такого анализа приведет к грубому искажению данных и тем более прогноза. Опираясь на такие выводы недопустимо. Результаты медицинских научных исследований обязательно должны анализироваться адекватными статистическими методами. От достоверности выводов таких научных работ зависит корректность решений принимаемых в области здравоохранения, что влияет на здоровье населения страны.

Говоря о временных рядах, чаще всего подразумеваются дискретные наблюдения с равными промежутками времени. Однако существуют статьи, где рассматривается преобразование неэквидистантного ряда в последовательность с равными промежутками времени посредством интерполяции последовательности кубическими β-сплайнами [15].

Далеко не всегда в исследованиях анализов ВР используют сложные статистические методы. В некоторых исследованиях достаточно расчета динамики в абсолютных и относительных цифрах по сравнению с предыдущими значениями – оценки темпов роста и прироста. Такие исследования характерны для статей отчетно-оценочного уровня [10, 20, 27]. В любом случае актуальность анализа и прогнозирования временных рядов в медицине не вызывает сомнений.

Целью данного исследования является литературный обзор научных источников на предмет применения различных методов и способов анализа временных рядов в медицинских исследованиях.

МЕТОДЫ И СРЕДСТВА

Используя контент-анализ, было рассмотрено 112 литературных источников в научных журналах

России и зарубежья, посвященных применению методов анализа и прогнозирования временных рядов в медицине за последние 5 лет.

Для поиска статей применялись Интернет-ресурсы: научная электронная библиотека eLibrary.ru и национальная медицинская библиотека США PubMed.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Традиционно под временным рядом понимают упорядоченную во времени последовательность величин

$$Y(t) = Y_1, Y_2 \dots, Y_t,$$

где t – момент времени [22].

Для анализа и прогнозирования временных рядов использовались несколько математических научно-практических подходов. На сегодняшний день используется широкий спектр возможностей при анализе ВР, исходя из целей исследования, от простейшего описания динамики [10] до прогнозирования методами искусственных нейронных сетей [25].

Одним из первоочередных методов анализа ВР является графический анализ. Невооруженным глазом на основе своего субъективного мнения исследователь может предположить тенденции развития ВР и выдвинуть гипотезы о прогнозе.

Анализ временного ряда может использоваться как промежуточный этап исследования с последующим корреляционным [1, 9, 16, 33], кросс-корреляционным [11, 12, 19, 23, 24], факторным, кластерным [26] и другими видами анализов.

Одним из самых простых способов анализа ВР является сглаживание ВР методом скользящей средней. Смысл данной методики состоит в том, что высчитывается среднее значение не по всем данным сразу, а взвешенно для каждой точки \tilde{Y}_t с учетом k соседних точек:

$$\tilde{Y}_t = \frac{Y_{t-k} + Y_{t-k+1} + \dots + Y_t + \dots + Y_{t+k-1} + Y_{t+k}}{2k + 1}$$

Метод скользящей средней в основном применяется как один из этапов подготовки представления данных для дальнейшего анализа. Данная методика нечасто применяется в научных медицинских исследованиях.

Часто авторами используется метод экспоненциального сглаживания временных рядов в различных вариациях, как метод аппроксимации данных.



$$\tilde{Y}_1 = Y_1,$$

$$\tilde{Y}_t = \tilde{Y}_{t-1} + \alpha \cdot (Y_t - \tilde{Y}_{t-1}), \text{ при } t > 1,$$

где α – коэффициент сглаживания, $0 < \alpha < 1$.

Например, используя данный метод, был проведен анализ заболеваемости населения алкогольными психозами и алкоголизмом на уровне РФ, Центрального Федерального округа и Воронежской области, выявлены районы с неблагоприятной обстановкой [6].

Примером частного варианта экспоненциального сглаживания является модель Брауна, которая применяется в работах В.В. Котина (2014 г.) [7] при прогнозировании заболеваемости скарлатиной. Модель Брауна может отображать развитие не только в виде линейной тенденции, но также в виде случайного процесса, не имеющего тенденции, а также в виде изменяющейся параболической тенденции. В этой же работе применяется еще один метод экспоненциального сглаживания – метод Хольта-Уинстера. Его суть заключается в получении прогностической модели, учитывающий экспоненциальный тренд и аддитивную сезонность.

Однако в некоторых работах авторы ограничиваются только методом Хольта [21], что диктуется особенностями предмета исследования. Метод Хольта характеризуется как еще один вариант экспоненциального сглаживания, используя для оценки влияния линейного тренда на исходный ВР.

Наиболее часто применяемым методом анализа ВР и его прогнозирования является регрессионный анализ. Говоря о регрессионных методах анализа ВР, в основном исследуют линейную зависимость, которая дает понятие о направлении тренда и количественную оценку характера развития процесса. Для некоторых работ этого достаточно [5, 17].

Однако в некоторых работах говорится о недостаточности данного метода, так как ВР может иметь нелинейную зависимость, например, при анализе временных рядов при прогнозировании природно-очаговых инфекций [4]. Среди нелинейных зависимостей часто применяются другие варианты описания тренда ВР: параболическая, логарифмическая, гиперболическая, экспоненциальная, степенная и другие модели [26].

Многие авторы, используя современные представления об анализе ВР, представляют математическую модель в виде аддитивных (1) [4] или мультипликативных (2) [3, 8] компонентов, где основными составляющими являются сезонная $S(t)$, циклическая $C(t)$, трендовая $G(t)$ и стохастическая $E(t)$.

$$Y(t) = C(t) + S(t) + G(t) + E(t); \quad (1)$$

$$Y(t) = C(t) * S(t) * G(t) * E(t). \quad (2)$$

Одним из распространенных вариантов анализа ВР является последовательная элиминация всех возможных зависимостей до получения, так называемого, «белого шума». Данное понятие подразумевает временной ряд со средним значением равным нулю, постоянной дисперсией и нулевой корреляцией последовательных наблюдений, то есть нулевой автокорреляцией [18].

Для определения модели временного ряда последовательно идентифицируют сезонные и циклические компоненты, далее детерминированную составляющую, и, в конечном итоге, обрабатывают остаточный ряд с помощью метода авторегрессионного проинтегрированного скользящего среднего (ARIMA).

Сезонные и циклические составляющие временного ряда могут быть найдены с помощью коррелограмм, которые показывают численно и графически автокорреляционную функцию. Другой полезный метод исследования периодичности состоит в исследовании частной автокорреляционной функции, в которой устраняется зависимость между промежуточными наблюдениями. Частная автокорреляция дает более чистую картину периодических зависимостей.

Для глубокого анализа ВР и получения адекватного прогноза рекомендуется применять метод ARIMA, как к стохастической составляющей аддитивной или мультипликативной модели, так и для анализа исходного ВР только данным методом.

Идея данного анализа заключается в нахождении зависимости текущего значения от показателей предыдущих временных отрезков. При зависимости только от предыдущего значения, говорят об авторегрессии первого уровня AR ($p=1$) от двух предыдущих – AR ($p=2$) и т.д. Помимо самих значений ВР, зависимость может быть найдена и в ошибках значений каждой временной точки. В случае, если текущая ошибка зависит только от предыдущей, то это первый порядок скользящего среднего MA ($q=1$), если ошибка зависит от двух предыдущих ошибок – MA ($q=2$) и т.д. Таким образом, частным случаем модели ARIMA является модель ARMA с параметрами (p, q). Данной моделью можно пользоваться только если ВР является стационарным [28, 30].

Понятие «Стационарность» временного ряда подразумевает наличие постоянного среднего значения, неизменяемой дисперсии, а ковариация зависит



только от временного интервала между отдельными наблюдениями [18, 31]. Для достижения стационарности необходимо брать разности ряда до тех пор, пока он не станет стационарным (часто также применяют логарифмическое преобразование для стабилизации дисперсии). Число разностей, которые были взяты, чтобы достичь стационарности, определяются параметром d . В итоге получаем единую модель ARIMA (p, d, q), которую принято считать процессом авторегрессии порядка p и d раз проинтегрированного скользящего среднего порядка q [29].

Достаточно много работ посвящено данному методу при анализе ВР, так как он является наиболее точным и дает наиболее правдоподобный прогноз. ARIMA-метод является достаточно гибким. С помощью данного метода можно проанализировать и сезонную, и циклические составляющие. Данная методика является расширением данного метода и называется SARIMA с параметрами $(p, d, q)(P, D, Q)m$, где P, D, Q являются аналогичными параметрами для сезонной составляющей, а m – период сезонной зависимости. Применение SARIMA в области медицины описано в статьях Бухарбаевой Л.Я., Егоровой Ю.В. (2018) [2] для прогнозирования региональной заболеваемости гриппом и ОРВИ; Разводовского Ю.Е. с соавт. (2015) для прогнозирования уровня суицидов [13]; Yan CQ (2019) для прогнозирования заболеваемости туберкулезом в Китае [35], Xinya Yu (2019) о заболеваемости шистосомозом в Китае [34] и др.

Существуют и принципиально другие подходы к анализу ВР, такие как вейвлет-преобразование, как обобщение спектрального анализа, например, в работе Рагозина Э.Р. и Рагозина О.В. (2016) [14] для выявления циркасаптанной ритмической активности в спектре ритмов вызовов скорой помощи. В некоторых работах прогноз строится на основе нейросетевых моделей ВР [25, 32].

ВЫВОДЫ

Таким образом, изучив научные публикации в области медицины за последние 5 лет, определили, что математическая статистика позволяет широко применять методы анализа временных рядов. Их можно встретить во многих областях медицины. Однако, так или иначе, все исследования касаются вопросов организации здравоохранения: оценка уровня смертности больных злокачественными новообразованиями полости рта или потери здоровья населения от инфаркта миокарда или другие. Выбор методов зависит от конечных целей самого исследования. Для некоторых работ достаточно описания динамики в относительных и абсолютных показателях по отношению к предыдущему временному отрезку, некоторые же используют сложный математический аппарат с помощью общеизвестных программ статистической обработки.

Если необходимо дать характеристику только тренду развития процесса во времени, то достаточно использовать методы экспоненциального сглаживания в различных вариантах или регрессионного анализа линейного или нелинейного. В случае наличия сезонных или циклических составляющих применяют методы автокорреляционной или частной автокорреляционной функции. Чаще применяются аддитивные и мультипликативные модели временного ряда.

Наиболее полно временной ряд можно проанализировать с помощью методов ARIMA. Данные способ в сочетании с вариантом SARIMA помогает привести временной ряд к стационарному виду и выявить все зависимости временного ряда.

Имея представление о современных способах анализа и прогнозирования временных рядов, авторы исследования могут выбрать наиболее оптимальный метод математического моделирования процесса в зависимости от поставленных задач.

ЛИТЕРАТУРА



1. *Беляева В.А.* Анализ временных рядов заболеваемости инфарктом миокарда с учетом гендерной специфики и солнечной активности // Современные тенденции развития науки и технологий. – 2015. – № 7–1. – С. 63–67.
2. *Бухарбаева Л.Я., Егорова Ю.В.* Прогнозирование региональной заболеваемости гриппом и ОРВИ на основе авторегрессионных моделей // Управление экономикой: методы, модели, технологии. Материалы XVIII Международной научной конференции. 2018. – Уфа: Уфимский государственный авиационный технический университет, 2018. – С. 301–305.
3. *Воробей О.В.* Исследование моделей анализа и прогнозирования временного ряда на примере смертности населения по городу // Современные проблемы гуманитарных и естественных наук.



- Материалы XXXIV международной научно-практической конференции. 2017. – М.: Научно-информационный издательский центр «Институт стратегических исследований», 2017. – С. 61–68.
4. *Евстегнеева В.А., Честнова Т.В., Смолянинова О.Л.* Анализ временных рядов в прогнозировании природно-очаговых инфекций // Вестник новых медицинских технологий. Электронное издание. – 2015. – № 1. – С. 9 DOI: 10.12737/7241.
 5. *Истомина А.А., Бадеников В.Я., Истомин А.Л.* Анализ и прогнозирование продаж лекарственных средств в аптечных организациях // Математические методы в технике и технологиях – ММТТ. – 2016. – № 3. – С. 136–138.
 6. *Косолапов В.П., Манакин И.И.* Анализ заболеваемости населения алкогольными психозами и алкоголизмом на основе результатов прогностического и классификационного моделирования // Моделирование, оптимизация и информационные технологии. – 2016. – № 1. – С. 1–13.
 7. *Котин В.В.* Прогнозирование заболеваемости: динамические модели и временные ряды // Биотехносфера. – 2014. – № 3. – С. 45–47.
 8. *Маххамова З.Р., Овсянникова Н.М., Голубова Т.Н., Ткаченко И.Ю.* Прогноз первичной заболеваемости туберкулезом легких (числа случаев) в республике Крым // Научные ведомости белгородского государственного университета. Серия: Медицина. Фармация. – 2017. – № 19. – С. 127–138.
 9. *Моисеева О.В., Сохлакова А.В.* Математические технологии в прогнозировании эпидемиологических показателей по туберкулезу у детей // Фтизиатрия и пульмонология. – 2015. – № 2. – С. 225–233.
 10. *Морев М.В., Короленко А.В.* Оценка демографических и социально-экономических потерь вследствие преждевременной смертности населения России и Вологодской области // ПРОБЛЕМЫ ПРОГНОЗИРОВАНИЯ. – 2018. – № 2. – С. 110–123.
 11. *Осипова Е.В., Аверина А.С., Суховская В.В., Протопопова Н.В., Баженов А.А., Дружинина Е.Б., Лабьгина А.В., Даренская М.А.* Влияние факторов солнечной и геомагнитной активности на развитие спонтанной родовой деятельности // ACTA BIOMEDICA SCIENTIFICA. – 2018. – № 3. – С. 82–87.
 12. *Разводовский Ю.Е., Зотов П.Б., Кондричин С.В.* Суициды и фатальный дорожно-транспортный травматизм в России: сравнительный анализ трендов // Суицидология. – 2016. – № 4. – С. 3–10.
 13. *Разводовский Ю.Е., Смирнов В.Ю., Зотов П.Б.* Прогнозирование уровня суицидов с помощью анализа временных серий // Суицидология. – 2015. – № 3. – С. 41–49.
 14. *Сказкина В.В., Киселев А.Р., Боровкова Е.И., Пономаренко В.И., Прохоров М.Д., Караваев А.С.* Оценка синхронизованности контуров вегетативной регуляции кровообращения по длительным временным рядам // Нелинейная динамика. – 2018. – № 1. – С. 3–12.
 15. *Рагозина Э.Р. и Рагозина О.В.* Циркасаптаный компонент временных вариаций вызовов скорой помощи в г. Ханты-Мансийске // Север России: стратегии и перспективы развития материалы II Всероссийской научно-практической конференции. – Сургут: Сургутский государственный университет, 2016. – С. 144–149.
 16. *Соколов С.В., Масалева О.В.* Оценка влияния патогенности погоды суток на частоту обращаемости больных бронхиальной астмой по городу Сургуту // Вопросы науки и образования. – 2018. – № 29. – С. 94–98.
 17. *Табакеев М.В., Шаповалова Э.Б., Максимов С.А., Артамонова Г.В.* Потери здоровья населения города Кемерово от инфаркта миокарда за 2006–2012 гг. по индексу *Daly* // Комплексные проблемы сердечно-сосудистых заболеваний. – 2014. – № 1. – С. 24–25.
 18. *Уотшем Т.Дж., Паррамоу К.* Количественные методы в финансах/ пер. с англ. под ред. М.Р. Ефимовой. – М.: Финансы, ЮНИТИ, 1999. – 527 с.
 19. *Федорова О.И.* Влияние гелиогеофизических циклов на частоту возникновения инсультов в Барнауле // Известия алтайского государственного университета. – 2014. – № 3–2. – С. 80–84.
 20. *Харькова Т.Л., Кваша Е.А., Ревич Б.А.* Сравнительная оценка смертности населения в российских и зарубежных мегаполисах // Проблемы прогнозирования. – 2018. – № 6. – С. 150–159.
 21. *Чопоров О.Н., Золотухин О.В., Болгов С.В.* Алгоритмизация интеллектуального анализа данных о распространенности заболеваний на региональном и муниципальном уровнях // Моделирование, оптимизация и информационные технологии. – 2015. – № 2. – С. 1–9.
 22. *Чураков Е.П.* Прогнозирование эконометрических временных рядов. – М.: Финансы и статистика, 2008. – 208 с.
 23. *Шельгин К.В.* Заболеваемость алкогольными психозами, смертность от отравлений алкоголем и патология печени в России в зависимости от финансовой доступности алкоголя // Журнал неврологии и психиатрии им. С.С. Корсакова. – 2015. – № 8. – С. 76–78.



24. Шельгин К.В., Сумароков Ю.А., Малявская С.И. Смертность от самоубийств в арктической зоне Российской Федерации // Социальные аспекты здоровья населения. – 2018. – № 1. – С. 4. DOI: 10.21045/2071-5021-2018-59-1-4.
25. Широкоступ С.В., Лукьяненко Н.В., Баландович Б.А., Сафьянова Т.В. Оценка потенциальной эпидемиологической опасности заболеваемости клещевыми инфекциями в Республике Алтай // Тихоокеанский медицинский журнал. – 2019. – № 3. – С. 33–36.
26. Шуметов В.Г. Методические аспекты типологизации регионов по показателям демографической нагрузки (на примере центрального федерального округа) // Human progress. – 2016. – № 1. – С. 1–25.
27. Юмагузин В.В., Винник М.В. Оценка уровня смертности от внешних причин (на примере Республики Башкортостан) // Проблемы прогнозирования. – 2017. – № 1. – С. 125–138.
28. Chatfield C. Time-series forecasting / Chris Chatfield. Chapman & Hall/CRC – 2000. – London. – p. 265.
29. Franses P.H. A Concise Introduction to Econometrics. – Cambridge: Cambridge university press, 2004. – 117 p.
30. Kirchgässner G., Wolters J. Introduction to Modern Time Series Analysis. – Berlin: Springer-Verlag, 2007. – 274 p.
31. Mariano R.S., Yiu-Kuen T. Econometric forecasting and High-frequency data analysis. – Singapore: World Scientific Publishing, 2008. – 189 p.
32. Soliman M., Lyubchich V., Gel Y.R. Complementing the power of deep learning with statistical model fusion: Probabilistic forecasting of influenza in Dallas County, Texas, USA // Epidemics. – 2019. – № 28. – P. 1–9.
33. Wang L., Liang C., Wu W., Wu S., Yang J., Lu X., Cai Y., Jin C. Epidemic Situation of Brucellosis in Jinzhou City of China and Prediction Using the ARIMA Model // Canadian Journal of Infectious Diseases and Medical Microbiology. – 2019. – № Article ID1429462. – 9 pages, <https://doi.org/10.1155/2019/1429462>.
34. Xinya Yu, Zhuang Chen, Longxing Qi. Comparative study of SARIMA and NARX models in predicting the incidence of schistosomiasis in China. Mathematical Biosciences and Engineering, 2019, 16(4): 2266–2276. doi: 10.3934/mbe.2019112.
35. Yan C.Q., Wang R.B., Liu H.C., Jiang Y., Li M.C., Yin S.P., Xiao T.Y., Wan K.L., Rang W.Q. Application of ARIMA model in predicting the incidence of tuberculosis in China from 2018 to 2019 // Zhonghua Liu Xing Bing Xue Za Zhi. – 2019. – № 40. – P. 633–637 DOI: 10.3760 / cma.j.issn.0254–6450.2019.06.006.

Новости отрасли

КОНФЕРЕНЦИЯ «ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В МЕДИЦИНЕ»

Правительство Ямало-Ненецкого автономного округа (ЯНАО) совместно с ассоциацией разработчиков и пользователей искусственного интеллекта в медицине «Национальная база медицинских знаний» (НБМЗ) и Министерством здравоохранения Российской Федерации проведут 27 марта 2020 г. 2-ю региональную научно-практическую конференцию «Искусственный интеллект в медицине».

На конференции будут представлены результаты различных Российских проектов внедрения систем поддержки принятия врачебных решений (СППВР), использующих методы искусственного интеллекта, опыт их разработки и внедрения. Также планируется несколько экспертных докладов, в том числе от Министерства здравоохранения, Росздравнадзора и «Национальной базы медицинских знаний». Конференция включена в план научно-практических мероприятий Минздрава России.

Более подробная информация: <http://conf2020.nbmz.ru/>