



Научная статья

## ЗАВИСИМОСТЬ УРОВНЯ СМЕРТНОСТИ В РЕГИОНАХ ОТ РАСПРОСТРАНЕННОСТИ АКТИВНЫХ НОСИТЕЛЕЙ SARS-CoV-2 И РЕСУРСОВ ОРГАНИЗАЦИЙ ЗДРАВООХРАНЕНИЯ

В.С. Степанов

Центральный экономико-математический институт РАН, Россия, 117418, г. Москва, Нахимовский  
проспект, 47

Рассматривается ряд математических моделей развития эпидемий: логистического типа, SIR-модель и др. Сделан небольшой обзор статей с такими моделями динамики заболеваемости COVID-19. Эти модели нередко работают успешно на данных по странам, однако на региональном уровне возникают трудности из-за особенностей учета смертности в России. Здесь могут быть полезны регрессионные модели, имеющие преимущество на начальном этапе эпидемического процесса. Также в них учитываются экзогенные переменные, влияющие на смертность, например по обеспеченности больниц врачами, медсестрами, аппаратами ИВЛ, койко-местами. Осуществлено построение линейной модели регрессии, на основе которой можно оценивать уровень региональной смертности от COVID-19, а также распределять перечисленные выше ресурсы.

Модель строится по комплексу ресурсных показателей с включением информации по «активным случаям». Предварительно тройка переменных с ресурсами инфекционных отделений больниц сворачивалась линейным преобразованием в одну новую. Затем модель оценивалась по обучающей выборке, содержащей эндогенную переменную по смертности и четыре факторных, включая распространенность активных носителей вируса. При этом регионы включались в материал обучения с разными лагами, они входили в такие ежедневные отсчеты  $t$ , когда случаи смерти фиксировались редко. Затем оцененная модель применялась при других значениях  $t$ . Для регионов из материала обучения и еще некоторых она неплохо оценивает уровень COVID-смертности (на отдельных промежутках).

В результате была построена и оценена по точности регрессионная модель, которая связывает уровень смертности в регионе с распространенностью в нем активных носителей SARS-CoV-2 и показателями обеспеченности больниц ресурсами. Она может быть полезной при распределении этих ресурсов. Также ее можно использовать при построении моделей SIRD, SEIR, SEIRF на региональном уровне, при выборе в них параметров, связанных со смертностью. Дополнительный интерес может представлять и методология, которую похожим образом можно применять и для других эпидемических процессов.

**Ключевые слова:** модель регрессии, оценка смертности, COVID-19, коронавирусная инфекция, логистическое уравнение, SEIR, SIR, ИВЛ.

По вопросам пандемии, порожденной SARS-CoV-2, на ресурсе E-library и некоторых других сейчас имеется очень много работ. Например, в обзоре К.В. Жмеренецкого, Е.Н. Сазоновой рассмотрены эпидемиологические и прочие особенности COVID-19, в других исследованиях изучаются и обсуждаются проблемы сбора информации о заболеваемости и смертности, достоверности статистики [1–7]. В работе О.М. Драпкиной, И.В. Самородской был сделан призыв строить математические модели прогностического типа, чему, в частности, посвящена настоящая работа [3]. В статье С. Козловского кроме аспектов учета заболеваемости и смертности рассматривается и вопрос моделирования [5].

Далее мы сфокусируемся на проблематике математического моделирования эпидемиологиче-

ского процесса (ЭП), так как важность построения достаточно точной модели для COVID-19 признается многими [3, 5, 8]. На ее основе можно принимать более эффективные управленческие решения, что позволит снизить смертность, уменьшить негативные последствия эпидемии, сэкономить ресурсы. Различные подходы к построению моделей динамики ЭП описаны в обзоре М.А. Кондратьева, где рассматриваются нейронные сети, имитационное и статистическое моделирование и др. Если упростить его классификацию подходов, можно выделить два типа моделей – детерминированные и стохастические [9].

В первую очередь кратко опишем некоторые модели первого типа и их применение к COVID-19. При построении моделей, учитывающих особенности

развития ЭП, население территории (страны, ее региона) разбивается на несколько секторов: «восприимчивые к вирусу», «инфицированные» и др. Сначала предлагается блок-схема развития ЭП во времени, в которой имеются возможные переходы индивидов из одного сектора в другой с различными интенсивностями (параметрами модели). Последние оцениваются: некоторые – по литературным данным, другие – по статистике, собранной для данной территории. Переходы между секторами блок-схемы формализуются на основе системы дифференциальных уравнений. Так, в модели Барояна – Рвачёва, позднее реализованной в системе EpidMod, используется система нелинейных интегродифференциальных уравнений в частных производных с соответствующими граничными и начальными условиями, взятая по аналогии из гидродинамики. Модели, основанные на уравнениях в частных производных функции двух переменных (времени и возраста пациента), встречаются и сейчас [10]. Однако большинство моделей развития ЭП во времени  $t$  опирается на систему обыкновенных дифференциальных уравнений первого порядка. Весьма давно была создана модель SIR, ее аббревиатура получена от английского варианта слов «восприимчивый», «инфицированный», «выздоровевший» [5, 8, 9, 11]. Позднее были разработаны ее различные модификации в виде моделей SEIR, SEIRF, SIRD и др.; в двух последних явным образом учитывается снижение числа зараженных из-за смертности:  $D$  – из-за инфекции,  $F$  – от осложнений. В библиотеке E-library с конца 1990-х гг. имеется более сотни работ с такими моделями, в основном зарубежных. Модели SIR и SEIR сейчас также широко используются при анализе динамики процессов в социальных сетях, при изучении распространения в них конкретной информации [12].

Модель SIRD применяется в работе P.V. Khramov, A.A. Loginova к COVID-данным для РФ и ряда стран [11]. Аналогичная модель, правда, в ее конечно-разностной форме, была реализована в исследовании О.В. Друговой, Е.А. Павлова и др. для анализа динамики COVID-19 в Нижегородской области [13]. В качестве основных переменных здесь выбирались численности лиц, разбитых на четыре сектора: «риск быть зараженными» ( $S$ ), «инфицированные» ( $I$ ), «выздоровевшие» ( $R$ ), «умершие» ( $D$ ), при этом для секторов  $I$  и  $R$  предполагалось, что по 30 % лиц в них не могут быть выявлены. После подбора параметров модель подгонялась к реальным данным в начальной части роста COVID-19 в регионе, далее на ее основе выполнялся этап прогнозирования.

Модель Барояна – Рвачёва относится к виду SEIR, буква « $E$ » обозначает, что вводится еще один сектор «Инкубационный период» (от *exposed*). Он определяется временем от момента попадания инфекционной дозы вируса в организм человека до начала его заболевания COVID-19. В работе М.В. Тамма используется расширенный вариант модели SEIR, здесь также учитываются некоторые свойства терри-

тории (возрастная структура населения и др.) [14]. Затем, после подбора параметров, по модели анализируется ряд сценариев борьбы с новым ЭП в Москве. В исследовании А.В. Матвеева вариант SEIR-модели, модифицированной с учетом особенностей COVID-19, был реализован в среде AnyLogic [15]. В результате для двух столичных городов сделаны сценарные расчеты, даны рекомендации по оснащению больниц инфекционными койко-местами и аппаратами для искусственной вентиляции легких (ИВЛ), сделан прогноз численности умерших.

Другой вариант модели, уже со стохастическим подходом к оцениванию, применяется в работе A. Godio, F. Pace et al. [16]. При этом используется обобщение, сделанное для ЭП исследователями L. Peng, W. Yang et al. [17]. В итоге был создан вариант SEIR-модели, в котором используется ряд параметров. Они оцениваются по данным исследования A. Godio, F. Pace et al. [16], в частности, здесь есть коэффициент смертности  $k(t)$ , зависящий от отсчета  $t$  и эффективности лечения пациентов. Исследователями S. Pengpeng, C. Shengli et al. [18] предложен еще один вариант SEIR-модели с учетом пониженной заразности инфицированных лиц, которые не имеют симптомов COVID-19, и с варьированием продолжительности инкубационного периода. Неизвестные параметры в исследовании находятся подгонкой модели к данным по КНР за период с 23.01 до 10.02.2020. В итоге выполнен краткосрочный прогноз и делается оценка, когда в КНР ожидается пик заражений, также приводится код программы в MATLAB.

В работе A.V. Nikitina, I.A. Lyapunova et al. [19] рассмотрена SIR-модель и пара модификаций SEIR-модели. Предложены изменения для этих модификаций, учитывающие особенности приобретения иммунитета, а также запаздывание при выявлении зараженных лиц, исследована чувствительность. Эти результаты можно использовать при моделировании COVID-19. Также почти полвека назад в практику отечественных исследователей вошла SEIRF-модель, в Центре ИЭМ им. Н.Ф. Гамалеи к концу XX в. было создано программное обеспечение. В работах Z. Liu, P. Magal et al. и F. Ndaïrou, I. Area et al. [20, 21] эта модель применяется к данным по COVID-19, причем в последней работе она строится уже с 8-ю секторами; здесь имеется выражение для  $D(t)$ : численность лиц, умерших за сутки  $t$ , оно включает три параметра. Отметим, что модель (1), предложенная нами ниже, может быть полезна в процессе выбора значений параметра  $k(t)$  в работе A. Godio, F. Pace et al. или параметров в  $D(t)$  [16, 21].

Кроме подхода через построение и решение системы дифференциальных уравнений имеются и более простые способы моделирования. Они используются во многих работах, посвященных прогнозированию характеристик изучаемого ЭП на основе решения обыкновенного дифференциального уравнения П. Ферхюльста и его обобщений [22–25].

Этот математик очень давно применил его для анализа роста населения в динамике по  $t$ . Оно распространено в биофизике, сейчас его используют при моделировании роста числа  $N(t)$  лиц, инфицированных SARS-CoV-2, по данным КНР, Швеции [22]. В этой работе рассматриваются три модели, основанные на уравнении Ферхюльста, приводятся соответствующие решения, обсуждается вопрос прогнозирования. После дифференцирования величины  $N(t)$  можно оценить как время наступления пика ЭП, так и количество необходимых коек в больницах согласно исследованию А.А. Куркина, О.Е. Куркиной и др. [22]. Одна из рассмотренных обобщенных моделей приводит к дифференциально-разностному уравнению. В этой модели учитывается наличие инкубационного периода в развитии ЭП, а также то, что примерно 4 % лиц, заболевших COVID-19, может погибнуть.

Модели такого типа менее чувствительны к качеству опубликованной статистики, чем перечисленные выше модели секторного типа. Они строятся по ежесуточным отсчетам временного ряда, в частности для ряда по численности умерших от COVID-19 или заболевших, и включают ряд параметров. Например, в исследовании N. Fabiano, S.N. Radenović модель Ферхюльста подгонялась к данным по численности граждан Италии, заболевших COVID-19 [23]. В другой работе эта модель была модифицирована с учетом специфики данной инфекции, что позволило при введении ряда предположений получить точное аналитическое решение [24]. Параметры этого решения были оценены на китайских данных, после чего получилось хорошее согласие модели с динамикой ЭП как для заболевших, так и для общей численности смертей в КНР [24].

Модель Ферхюльста используется также и в дискретном варианте, в котором имеется два параметра. Такой вид логистического уравнения после работ американского ученого М. Фейгенбаума стал широко использоваться при изучении различных процессов. Применительно к численности инфицированных в десятке стран и двух мегаполисах мира дискретная модель Фейгенбаума использовалась в работе Э.М. Кольцовой, Е.С. Куркиной и др., при этом ее параметры были оценены по общему числу заболевших в КНР [25]. Для Москвы здесь было рассмотрено четыре варианта роста общего числа зараженных, самый легкий из них был назван «израильским». По этому варианту к концу ЭП ожидается 300 тысяч москвичей, инфицированных SARS-CoV-2, что вполне правдоподобно согласуется с 248 тысячами на 10 августа с приростом по 700 случаев за сутки.

Наконец, в исследовании В.В. Бояринцева, Р.С. Пальмина и др. [8] выявляются закономерности развития ЭП на основе анализа общего числа  $I(t)$  инфицированных лиц, проживающих на конкретной территории. В результате при комплексе гипотез-допущений предложена рекуррентная формула расчета  $I(t)$  на  $t$ -й день ЭП. На ее основе получена зави-

симость по  $t$  для функции распределения доли инфицированных лиц, затем сделаны прогнозы по странам и для Москвы.

Однако многие исследователи ЭП на основе SIR и ее модификаций, отмечают неустойчивость получаемых решений, что препятствует получению хороших прогнозов ЭП на период свыше недели [5, 8]. Также о приближенных результатах прогнозирования сообщается в исследовании А.В. Матвеева [15]. В статье С. Козловского, О. Болдырева указывается, что многие параметры для моделей ЭП, если их применять для регионов, необходимо оценивать по региональной статистике [5]. А качество ее публикаций специалисты по данным считают низким [4, 5, 7]. Экспертно ими оценивается, что уровень ее искажений растет с конца апреля 2020 г. М.В. Тамм также не доверяет статистике по смертности [5, 14]. Эталонным регионом по открытости и достоверности данных по COVID-смертности всеми признается г. Москва. По регионам ситуацию со смертностью лучше освещают Росстат и региональные органы ЗАГС. Последние на своих сайтах ежемесячно сообщают об общих случаях смерти за месяц без указания ее причин, правда, такие публикации выходят в свет с существенными задержками [4, 6, 7]. Кроме того, некоторые параметры на начальном этапе развития ЭП неизвестны; именно поэтому на этой стадии необходимо применение максимально простых моделей, содержащих минимальное количество параметров [22]. К ним можно отнести модели регрессионного анализа и временных рядов.

Если рассмотреть недетерминированные модели ЭП, в которых существующая зашумленность данных учитывается естественным образом, то в исследовании Д.В. Мелика-Гусейнова, Н.Н. Карякина и др. [26] были построены и оценены несколько регрессионных моделей для анализа COVID-смертности на уровне всей РФ. Так, модель 1 в варианте 2 имеет вид  $\ln(Y) = \ln(X) \cdot \beta_1 + \beta_0$ , где  $Y$  – количество новых летальных случаев за прошедшие сутки, а  $X$  – количество новых заболеваний за эти сутки;  $\beta_0, \beta_1$  – параметры этой модели, подлежащие оцениванию. После подстановки оценок  $\beta_0, \beta_1$  из исследования Д.В. Мелика-Гусейнова, Н.Н. Карякина и др. по методу наименьших квадратов (МНК) получается линейная модель регрессии с коэффициентом детерминации  $R^2 = 70\%$  [26]. Кроме того, предложена модель 2 со значением  $R^2 = 80\%$ , она строится, исходя из модели 1, после введения еще одной переменной  $X_2$ : «количество тяжелых случаев протекания COVID-19 в  $j$ -м регионе».

В работе И.А. Лакмана, А.А. Агапитова и др. предлагается достаточно сложно создаваемая модель ARIMA, а также адаптивная модель Хольта – Уинтерса – ARIMA или АРПСС – вид «авторегрессии – скользящего среднего» Бокса-Дженкинса [9, 27]. С их помощью вместе с использованием модели SIRD с высокой точностью можно сделать прогноз краткосрочной заболеваемости, смертности и выжи-

ваемости COVID-пациентов на конкретной территории. Однако для SIRD-модели предполагается, что используемые коэффициенты, отвечающие за вероятности инфицирования, излечения и смерти, постоянны во времени  $t$ , что не согласуется с российской практикой. Это приводит к тому, что для получения хороших прогнозов модель приходится довольно часто перенастраивать [27].

Из-за вышеперечисленных трудностей на уровне российских регионов моделирование применяется более редко, чем на уровне стран, РФ и Москвы. Кроме того, некоторые статистические модели имеют преимущества по сравнению с секционными или логистическими моделями. Поэтому иногда сочетают оба подхода, как это было сделано в вышеуказанных работах, где вместе с SEIR-моделью применяется аналог распределения Эрланга [9, 27–29]. Среди статистических моделей весьма привлекательны модели линейной множественной регрессии [9]. Их положительными свойствами является хорошо разработанная теория, например, можно построить доверительные интервалы, вычислить коэффициенты эластичности и др. Такую модель легко реализовать при соответствующей квалификации пользователя в комплексах программ: IBM SPSS, Statistica, EViews, Deductor. Также она есть в Excel, в системах компьютерной математики MATLAB, Wolfram Mathematica и др., среде Python или R (RStudio). Наконец, здесь не сложно учесть дополнительные факторы, которые могут влиять на уровень COVID-смертности. Например, добавить переменную, описывающую тяжесть протекания болезни у пациента; показатель доступности для него медицинской помощи; переменную по потреблению им табака и электронных сигарет, или другие показатели.

Необходимо также отметить, что моделированием ЭП, в частности COVID-19, сейчас занимаются известные научные центры и университеты. Например такие, как Центр моделирования инфекционных заболеваний (CMMID) из Школы гигиены и тропической медицины, Центр эпидемиологии Имперского колледжа (Лондон), университеты Женевы и Базеля, многие университеты из США и других крупных стран [5, 14]. Так, американские Центры по контролю и профилактике заболеваемости еженедельно публикуют оценки развития ЭП, полученные 24–26 группами ученых, в основном из США, для общего числа умерших от COVID-19 [30]. Эти оценки приводятся на дату  $t$  в виде точечных и интервальных прогнозов на четыре недели. Используется широкий набор моделей; так, 11 групп исследователей применяют различные виды модели SEIR (и ее стохастические варианты SEIS, SEIRD, SEIRX); пять групп – SIR-модель и ее модификацию SLIR.

Реже применяются агентные модели, нейронные сети, модели теории временных рядов и регрессионного анализа, как нелинейного, так и линейного, с применением «гребневой» оценки матрицы ковариаций (модель UM) [9, 30]. Ряд групп исследует ситуацию и на уровне штатов страны. Также некоторые иностранные эпидемиологи моделируют динамику COVID-19 в России [5, 31].

**Цель исследования** – построение линейной модели множественной регрессии, на основе которой можно оценивать региональный уровень смертности от COVID-19, выполнять ее точечный прогноз на ранней стадии ЭП. Также с ее помощью можно более обоснованно распоряжаться некоторыми ресурсами организаций здравоохранения в регионах России.

**Материалы и методы.** Модель строилась по данным о распространенности активных носителей SARS-CoV-2 в начальном периоде развития ЭП в регионе, а также с учетом показателей обеспеченности его больниц ресурсами или мощностями. Сначала была сформирована таблица «объект – признак», объектами были многомерные наблюдения (или  $p$ -мерные векторы) над  $n = 40$  регионами РФ, исходными признаками были  $p = 7$  показателей. Из них два изменяются ежесуточно: уровень COVID-смертности населения и соотношение числа активных носителей вируса со среднегодовой численностью жителей за 2019 г., иначе говоря, их распространенность («фактор», или «экзогенная переменная»  $X_1$ )<sup>1</sup> [32]. Остальные пять брались за 2018, 2019 гг. из данных Росстата<sup>2</sup>, исследования Ю. Апухтиной, С. Зобовой и данных статьи А.А. Соколова [33, 34]. Они описывали обеспечение региональных больниц:  $X_2$  – аппаратами для ИВЛ в отделениях реанимации (2019) в расчете на  $10^5$  жителей;  $X_3$  – врачами из санитарно-противоэпидемической группы (2018) на  $10^5$  жителей; а также врачами-инфекционистами (2019), средним медперсоналом (2018), исключая сестер-акушеров; койко-местами «инфекционного типа» (2019).

Для обоснования выбора этих переменных отметим, что в некоторых исследованиях были перечислены факторы, которые влияют на летальность в ЭП [3]. Среди них оказались состояние системы здравоохранения в период, предшествующий ЭП, а также доступность и качество медицинской помощи. Переменные  $X_2, X_3, \dots$  отчасти отражают это. Многие из них учитываются в процессе принятия решений за рубежом [13]. Так, в одном из университетов Германии была разработана система «Cogo-naVIS» для мониторинга ресурсов госпиталей страны – их коечного фонда и аппаратов кислородной терапии [35]. После онлайн-запроса на фоне карты местности здесь сразу показывается загрузка коек и аппаратов в конкретном госпитале с его координатами.

<sup>1</sup> Российский статистический ежегодник. 2019 // Росстат. – М., 2019. – 708 с.

<sup>2</sup> Здравоохранение в РФ. Приложение к сборнику (информация в разрезе субъектов Российской Федерации) [Электронный ресурс] // Росстат. – 2019. – URL: <https://rosstat.gov.ru/folder/210/document/13218> (дата обращения: 06.08.2020).

тами, приводятся диаграммы загрузки в течение времени суток для близко лежащих к нему госпиталей. Это позволяет заранее избежать появления «узких мест» в процессе доставки пациентов в госпиталь, а также оптимизировать их распределение по госпиталям. В институте ИМБЕ из Вашингтонского университета разработаны модели для прогнозирования частоты случаев смерти пациентов, а также уровня занятости коек и аппаратов ИВЛ [30, 36]. В колледже из Нью-Йорка совместно с учеными из трех университетов в среде Excel создан калькулятор для оценки нагрузки, порожденной COVID-пациентами, в электронной таблице учитывается мощность больниц и загрузка аппаратов ИВЛ. В итоге здесь моделируется нагрузка на инфекционные и реанимационные отделения больницы конкретного региона, причем с учетом загрузки коек, как обычных, так и для интенсивной терапии [37].

Предварительно перечисленные выше переменные, описывающие обеспеченность инфекционных отделений больниц ресурсами (врачи-инфекционисты, медсестры, инфекционные койки), сворачивались по методу главных компонент в новую переменную. При этом использовалась небольшая модификация, предложенная ранее в работе S.A. Aivazian [38, 39]. Алгоритм расчета этой переменной, обозначенной ниже  $(X_4)^*$ , включал серию шагов. Сначала каждая из переменных (показателей) отображалась на  $[0, 10]$ , где 10 соответствует ее максимальному значению в таблице обучающих данных, а 0 – минимальному. Затем вычисляется оценка  $S$  ковариационной матрицы трех переменных. Далее для матрицы  $S$  находятся все ее собственные числа и векторы, один из последних задает искомые веса для расчета величины  $(X_4)^*$ .

Выборка объема  $n = 40$  регионов получилась после исключения трех федеральных городов, Московской области и трех граничащих областей, а также Астраханской, Брянской, Волгоградской, Воронежской, Калининградской, Кемеровской, Мурманской, Орловской, Пензенской, Псковской, Ростовской областей; всех АО; республик Калмыкия, Коми, Марий Эл, Чувашская, многих из Северного Кавказа и некоторых других; Алтайского, Краснодарского, Ставропольского краев; еще регионов, где весной не публиковалась смертность. Статистика для многих из них критикуется в инфографике по коронавирусу в России [7].

В итоге строилась неадаптивная линейная регрессионная модель

$$Y = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \beta_3 \cdot X_3 + \beta_4 \cdot (X_4)^* + \varepsilon, \quad (1)$$

где  $Y$  – уровень смертности от COVID-19;

$X_1, \dots, (X_4)^*$  – четверка вышеописанных «факторов»;

$\varepsilon$  отражает влияние на  $Y$  неучтенных переменных;

$\beta_0, \dots, \beta_4$  – неизвестные параметры, оцениваемые по выборке [9].

Величина  $\varepsilon$  в формуле (1) определяется как случайная; по определению ее математическое ожидание  $E(\varepsilon) = 0$  и дисперсия  $\text{var}(\varepsilon) = \sigma^2$ , причем она

неизвестна. Оценки для  $\beta_0, \dots, \beta_4$  и  $\sigma^2$  находятся по МНК в среде Excel по таблице с обучающими данными. Обозначим оценки  $\beta$ -коэффициентов через  $b_0, \dots, b_4$ . Эта таблица включала наборы значений  $\{Y, X_1, X_2, X_3, (X_4)^*\}$  для всех  $n$  регионов, причем эти наборы входили в таблицу с разными отсчетами по  $t$ . Предполагалось, что новое число смертей  $M$ , опубликованных за сутки  $t$  для конкретного региона, является случайной величиной, приближенно имеющей биномиальное распределение, а  $m$  выявленных за сутки  $t$  активных случаев COVID-19 среди  $N$  его жителей интерпретировались как  $m$  независимых испытаний Бернулли над совокупностью из  $N$  объектов, из которых  $M$  обладают важным для нас свойством «был летальный случай» [32]. С ростом  $m$  распределение числа  $M$  приближается к пуассоновскому закону, однако вероятность смерти  $P_t$  из-за COVID-19 в регионе за сутки  $t$  точно неизвестна. Для ее оценки рассматривалась последовательность за 7–10 дней с приростами числа смертей за сутки  $t$ . Затем на основе биномиального закона строился 95%-ный доверительный интервал, и его середина бралась за оценку  $P_t$ . Далее бралось среднее значение  $P^*$  этих оценок по указанной последовательности дней. Как известно, смесь пуассоновских распределений также имеет распределение Пуассона, но с другим параметром интенсивности. В итоге при формировании обучающей выборки лаг во временном ряде выбирался так, чтобы значение оценки  $P^*$  вероятности COVID-смерти лежало в  $(3 \cdot 10^{-5}, 9 \cdot 10^{-5})$ , т.е. было менее одного летального случая на 1 млн человек. Например, Ульяновская область входит в таблицу с данными за дату 11.05.2020, а ее значение  $P^*$ , рассчитанное за период 06.05–17.05.2020, равно 0,000049. Сделаем следующее:

*Замечание 1.* Так как три факторные переменные  $X_2, X_3$  и  $(X_4)^*$ , описывающие обеспеченность больниц ресурсами, не изменяются (точнее, в нашем распоряжении нет данных об изменении), то выражение (1) можно записать для каждого  $j$ -го региона в виде суммы двух слагаемых. Первое – константа  $C(j)$ , по определению равная  $b_0 + b_2 \cdot X_2 + b_3 \cdot X_3 + b_4 \cdot (X_4)^*$ , где  $X_2, X_3, (X_4)^*$  – переменные  $j$ -го региона,  $j = 1, \dots, n$ . Величина  $C(j)$  показывает, каким ожидается уровень смертности при совсем малой распространенности активных случаев в  $j$ -м регионе, когда  $X_1 \approx 0$ . Очевидно, что чем лучше развиты ресурсы больниц в регионе, тем меньше будет значение  $C(j)$ . Величина  $C(j)$  не зависит от  $t$ , так как параметры в итоге оцениваются для (1) один раз. Второе слагаемое в (1) содержит фактор  $X_1(j)$ , который уже зависит от  $t$ . В результате модель (1) после оценки неизвестных  $\beta_0, \dots, \beta_4$  имеет вид

$$Y(j) = C(j) + b_1 \cdot X_1(j), \quad (2)$$

где  $Y(j)$  – уровень COVID-смертности на 1 млн человек;

$X_1 = X_1(j)$  – распространенность активных случаев с SARS-CoV-2 в расчете на  $10^4$  жителей из  $j$ -го региона.

Индекс  $t$  здесь опущен для упрощения записи. Формулу (2) очень легко использовать для ежедневной оценки  $Y$ . Для этого надо взять в сводной таблице официального интернет-ресурса по вопросам коронавируса за дату  $t$  общие численности: инфицированных (И), умерших (У), выздоровевших (В), найти значение  $X_1 = 10 \cdot [И - (В + У)] / N$ , где  $N$  – численность населения из табл. 2, и затем подставить  $X_1$  в (2) с использованием МНК-оценки  $b_1$  для  $\beta_1$  из (1) [32].

*Замечание 2.* Однако, если в  $i$ -м регионе имело место изменение одной из переменных  $X_2$ ,  $X_3$  и  $(X_4)^*$ , и у пользователя имеется такая информация, то вместо (2) следует использовать формулу (1) с оценками  $b_0, \dots, b_4$  вместо  $\beta_0, \dots, \beta_4$  (табл. 1).

В итоге в начале июня по сформированной выше выборке были найдены МНК-оценки  $b_0, \dots, b_4$  параметров (1), равные 4,6721; 0,4494; –0,04774; –0,05668; –0,3308; оценка  $\sigma$  равна 0,03474. Достигнутые уровни значимости для каждого фактора и свободного члена были ниже 0,001, при расчете их через «Пакет анализа» в Excel. Оценка коэффициента детерминации  $R^2$ , поправленная для отсутствия смещения, получилась выше 95 %, что, отчасти, связано с особенностью формирования обучающей выборки. Остатки для формулы (2) распределены приближенно нормально, их нормальность проверялась через  $d$ -статистику критерия, описанного Л.Н. Большевым, Н.В. Смирновым [40]. Она рассчитывается как среднее значение абсолютного отклонения, найденного по  $n$  наблюдениям для остатков в формуле (2), которое берется в соотношении с корнем из выборочной дисперсии. Эта статистика распределена примерно нормально с параметрами  $E(d)$ ,  $var(d)$  согласно формулам, приведенным в [40]. Численные значения для статистики критерия  $d$  и ее параметров при расчете их по остаткам (2) и выборке объема  $n = 40$  оказались:  $d = 0,7954$ ,  $E(d) = 0,8029$ ,  $(var(d))^{0.5} = 0,0324$ , поэтому гипотеза о нормальности не отвергается, так как  $d \approx E(d)$ , причем отклонение от  $E(d)$  меньше 1/4 стандарта этой статистики.

**Результаты и их обсуждение.** Сначала приведем табл. 1 к замечанию 2, сделанному выше. Для трех регионов здесь показан спад величины  $Y$  при росте  $X_2$ . Отметим, что недорогой аппарат ИВЛ «Авента-М» стоит 1,8 млн руб.

Здесь вместо фактических значений по числу аппаратов ИВЛ в регионах на дату  $t$  (табл. 1) мы взяли аналогичное их количество, требуемое по нормативам, и рассчитали значение показателя  $X_2$

для обоих случаев с учетом данных Росстата [33, 34]. Далее, подставив значение  $X_2$  в формулу (1) и заменив предварительно здесь  $\beta_0, \dots, \beta_4$  на  $b_0, \dots, b_4$ , оценили, как с ростом значений  $X_2$  снизился бы уровень смертности  $Y$  на момент  $t$ . Из столбца 5 и расчета по модели следует, что ожидается весьма заметное снижение  $Y$ : от 22 до 75 %.

Модель формулы (1) с оценками  $b_0, \dots, b_4$  или формула (2) может с удовлетворительной точностью ежедневно оценивать уровень смертности, порожденный COVID-19 в ряде регионов. Ниже показаны данные для 12 регионов и достигнутая для них точность по последовательности суток, взятых из промежутка по  $t$  из табл. 2. При этом ряд регионов не входил в материал обучения: Республика Карелия, Тамбовская и Курганская области. Расчеты в табл. 2 сделаны по формуле (2) с фокусом на период со второй половины мая до июля. В столбце 5 табл. 2 приводится средняя ошибка аппроксимации (MRPE):

$$\delta = \frac{100}{K} \cdot \sum_{t=1}^K \left| \frac{Y_t - Y_t^*}{Y_t^*} \right|, \quad (3)$$

рассчитанная по модулям относительных ошибок в ходе сравнения оценки  $Y$ , полученной по формуле (2), с ее сглаженным истинным значением  $Y_t^*$  (%). Расчет был сделан по  $K$  суткам, взятым из промежутка, который показан в столбце 4 табл. 2.

В столбце 5 табл. 2 в скобках стоит выборочное стандартное отклонение, найденное по указанной последовательности из  $K$  суток. На  $t$ -м шаге такого расчета оценка  $Y_t$  по формуле (2) сравнивалась со сглаженным значением  $Y_t^*$  (с возможной его предварительной коррекцией согласно столбцу 6 табл. 2; но если там «0», то данные [32] не изменялись). Коррекция «вверх» показателя COVID-смертности в некоторых регионах вполне оправдана, поэтому если в столбце «Примечание» указан не ноль, то это число летальных случаев, добавленных к данным [4–7, 32, 41]. В серии графиков приводится динамика этой смертности в регионах от информационного центра Минздрава в сравнении с графиками из официального интернет-ресурса по вопросам коронавируса [32, 41]. Для 31 российского региона они идентичны, шесть – публикуют информацию с задержкой, 19 – иногда занижали значения относительно расследования «Медиазоны», а 29 – приводят в данные, сильно отличающиеся по многим промежуткам. Аналогичные табл. 2 значения  $\delta$  можно получить и для прочих

Таблица 1

Изменение  $Y$  при улучшении обеспечения регионов аппаратами ИВЛ

Область из РФ	Наличие ИВЛ, шт.	Норматив Минздрава	Улучшение фактора $X_2$	Уменьшение значений $Y$	$t$
1	2	3	4	5	6
Воронежская	348	860	15 → 37	2.43 → 1.38	24.04
Тверская	118	527	9.24 → 41.3	4.7 → 3.84	26.04
Ростовская	643	1521	15.3 → 35	4.5 → 3.57	05.05

Таблица 2

Результаты расчета константы  $C(j)$  и средней ошибки аппроксимации  $\delta$ 

Регион	$C(j)$ из (2)	$N(j)$ , тыс. чел. <sup>3</sup>	Промежуток по $t$	Ошибка $\delta$ (%) из (3)	Примечание
Республика Карелия	1.060	620.3	06.05–05.07	5.0 (2.9)	0... +1
Республика Крым	1.163	1912.8	27.05–02.06	6.2 (2.2)	0
Республика Татарстан	2.164	3896.4	30.05–08.07	4.5 (3.9)	0
Архангельская обл.	0.120	1149.6	02.06–09.06	3.3 (2.3)	0
Вологодская обл.	2.074	1172.2	15.05–19.05	5.2 (3.2)	+1
Курганская обл.	3.191	840.1	15.05–02.07	5.3 (3.0)	+1... +3
Липецкая обл.	1.333	1147.1	03.06–14.06	2.3 (1.6)	+1
Оренбургская обл.	0.901	1970.4	04.05–25.05	5.6 (2.7)	0
Рязанская обл.	1.541	1117.8	04.06–16.06	3.5 (2.0)	+4
Тамбовская обл.	1.970	1024.7	27.05–14.06	3.0 (2.3)	+2
Ульяновская обл.	1.529	1242.5	07.05–01.06	3.5 (2.2)	+1
Челябинская обл.	2.10	3484.4	18.05–27.05	1.0 (0.6)	+7...0 [7]

почти 30 регионов для промежутка «апрель – первая половина мая» (они входили в материал обучения с более ранними отсчетами по  $t$ ).

Эта модель обеспечивает точность из столбца 5 табл. 2, пока оценка вероятности смерти  $P^*$  имеет примерно те же значения, что и в материале обучения. Например, для Ульяновской области в промежутке 06.05–17.05.2020 оценка  $P^*$  была 0,00049 (с учетом данных столбца 6 табл. 2), а на отрезке 18.05–02.06.2020, для дат из которого выполнялся прогноз по формуле (2), она стала около 0.0004.

Опишем чуть подробнее результаты для ряда регионов, полученные по формуле (2) на различных промежутках по  $t$ , и приведем графики на рис. 1.

По оси ординат везде – уровень смертности в  $j$ -м регионе  $Y$ . Он находится как общее число летальных случаев в расчете на 1 млн человек. Сначала на рис. 1, а, приводятся графики для Республики Татарстан. Пунктиром показан график, построенный по формуле (2) для суток  $t$  из (27.05.2020, 10.07.2020); непрерывной линией – уровень смертности<sup>4</sup> [32].

Далее на рис. 1, б, идут графики для Архангельской области<sup>5</sup>. Непрерывная линия показывает сглаженные данные по факту; мелкие точки лежат на экспоненте, подогнанной к  $Y$ , пунктир – согласно формуле (2) [32].

Затем представлены графики для Липецкой области (рис. 1, в). График, показанный длинным пунктиром и точками, рассчитан по формуле (2). Значения на графике ступенчатого вида считались так: бралось число летальных случаев, далее с 17.04.2020 к ним добавлялся еще один, вероятно, еще неучтенный случай смерти, затем сумма умножалась на  $10^3$ , делилась на  $N(j) = 1147,1$  (это обозначается в столбце 6 табл. 2 кодом «+1») [32]. Гладкая кривая на рис. 1, в – МА (moving average – «скользящее среднее») – рассчитана по точкам ступенчатого графика через подвижное центрированное 7-дневное окно.

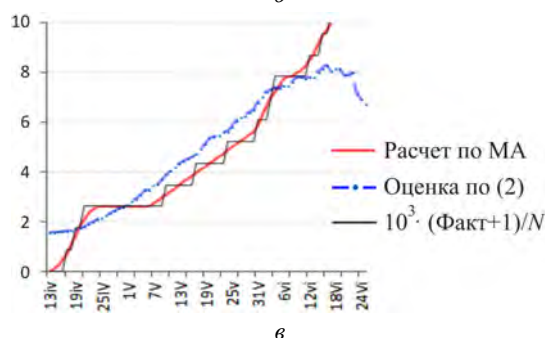
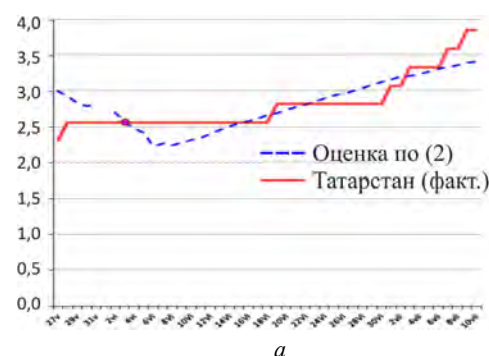


Рис. 1. Динамика: а – для  $Y$ : Республика Татарстан; б – для  $Y$ : Архангельская обл.; в – для  $Y$ : Липецкая обл.

<sup>3</sup> Российский статистический ежегодник. 2019 // Росстат. – М., 2019.

<sup>4</sup> Там же.

<sup>5</sup> Там же.

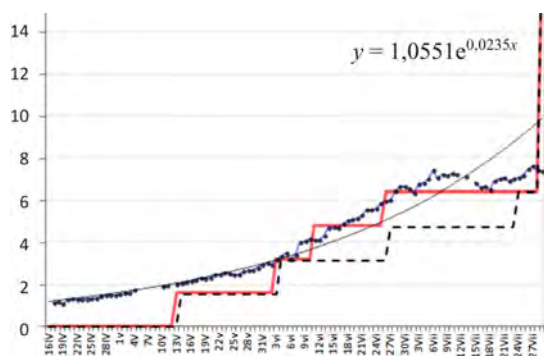


Рис. 2. Динамика показателя смертности  $Y$  в Республике Карелия и ее оценка по формуле (2)

Далее на рис. 2 приводятся графики для Республики Карелия с 16.04 по 29.07.2020. Точки темного цвета получены после подстановки значений  $X_k$  в формулу (2). К этим точкам подгонялся экспоненциальный тренд, показанный тонкой линией.

Константа в его формуле есть 1,055, что соответствует почти нулевой распространенности активных случаев COVID-19, когда  $C(j) = 1,06$  из табл. 2. Ступенчатая линия (пунктир) показывает уровень смертности, аналогичная непрерывная линия показывает  $Y$ , что была скорректирована согласно столбцу 6 табл. 2; они чуть сдвинуты, чтобы не было слияний. Начиная с 11.06.2020 при расчете скорректированной линии нами добавлялся еще один летальный случай, который, вероятно, был учтен с запозданием. Об этом дополнительном случае сообщают официальные региональные власти, это можно увидеть на репозитории «Медиазоны» [7]. О сдвигах по  $t$  при публикации случаев смерти сообщается и в других источниках [6, 7, 26, 41]. Это предположение можно попытаться

обосновать логически, анализируя уровень смертности в приграничных областях: Ленинградской, Вологодской и Архангельской. Причем его надо рассмотреть при близких значениях  $X_1$ , что были в республике во вторую декаду июня (между 6,8 и 9). Из графиков на рис. 2 следует, что модель, описанная формулами (1), (2) достаточно хорошо оценивает чуть подправленный «вверх» уровень смертности на отрезке 06.06, ..., 05.07.2020 (если вместо ступенчатой линии провести линию, сглаженную по скользящему среднему). Аналогичная работа была выполнена для Курганской области. Примерно похожие действия при корректировке данных о смертности проводят аналитики ряда организаций [5, 7], например в компании Data Insight, где в отделе исследований строится индекс недооценки смертности [5].

**Выводы.** Таким образом, построена и оценена по точности регрессионная модель, которая связывает уровень смертности в регионах с распространенностью активных случаев COVID-19, а также с обеспеченностью больниц ресурсами. Она может быть полезной в начальном периоде распространения ЭП, когда в регионе изредка наблюдаются совсем малые ежедневные приросты смертности в течение 7–10 дней, и они подчиняются распределению Пуассона. Кроме того, модель может быть полезна при построении моделей SEIR, SIRD и др. на этапе подбора параметров для них, связанных со смертностью. Также ее можно применять при распределении ресурсов организаций здравоохранения.

**Финансирование.** Исследование не имело спонсорской поддержки.

**Конфликт интересов.** Автор заявляет об отсутствии конфликта интересов.

### Список литературы

1. COVID-19: Только научные факты / К.В. Жмеренецкий, Е.Н. Сазонова, Н.В. Воронина, Г.С. Томилка, О.А. Сенькевич, В.С. Гороховский, С.В. Дьяченко, И.П. Кольцов, М.Б. Куцый // Дальневосточный медицинский журнал. – 2020. – № 1. – С. 5–22.
2. Данилова И.А. Заболеваемость и смертность от COVID-19. Проблема сопоставимости данных // Демографическое обозрение. – 2020. – Т. 7, № 1. – С. 6–26.
3. Методические аспекты оценки заболеваемости, распространенности, летальности и смертности при COVID-19 / О.М. Драпкина, И.В. Самородская, М.Г. Сивцева, Е.П. Какорина, Н.И. Брико, С.Н. Черкасов, В.А. Цинзерлинг, П.Г. Мальков // Кардиоваскулярная терапия и профилактика. – 2020. – Т. 19, № 3. – С. 302–309.
4. Иванов С. Смертность от COVID-19 на фоне других всплесков смертности XX века // Демографическое обозрение. – 2020. – Т. 7, № 2. – С. 143–151.
5. Козловский С., Болдырев О. Можно ли предсказать развитие пандемии коронавируса? Объясняем на примере России [Электронный ресурс] // BBC Russia. – URL: <https://www.bbc.com/russian/features-52762747> (дата обращения: 30.07.2020).
6. Глава Росстата рассказал, как считают жертв COVID-19 [Электронный ресурс] // BBC Russia. – URL: <https://www.bbc.com/russian/features-53156041> (дата обращения: 02.08.2020).
7. Коронавирус в России: Инфографика [Электронный ресурс] // Медиазона. – 2020. – URL: <https://zona.media/coronagraph> (дата обращения: 30.07.2020).
8. Метод прогнозирования параметров эпидемического процесса, вызванного COVID-19 / В.В. Бояринцев, Р.С. Пальмин, С.А. Пальмин, С.Ф. Перцев // Кремлевская медицина. Клинический вестник. – 2020. – № 2. – С. 14–21.
9. Кондратьев М.А. Методы прогнозирования и модели распространения заболеваний // Компьютерные исследования и моделирование. – 2013. – Т. 5, № 5. – С. 863–882.
10. Magpantay F.M.G., Kosovalić N., Wu J. An age-structured population model with state-dependent delay: derivation and numerical integration // SIAM Journal on Numerical Analysis. – 2014. – Vol. 52, № 2. – P. 735–756. DOI: 10.1137/120903622
11. Khrapov P.V., Loginova A.A. Comparative analysis of the mathematical models of the dynamics of the coronavirus COVID-19 epidemic development in the different countries // International Journal of Open Information Technologies. – 2020. – Vol. 8, № 5. – P. 17–22.

12. Системно-динамическое моделирование сетевых информационных операций / В.А. Минаев, М.П. Сычев, Е.В. Вайц, К.М. Бондарь // Инженерные технологии и системы. – 2019. – Т. 29, № 1. – С. 20–39.
13. Статистические и динамические аспекты прогнозирования распространения COVID-19 в Нижегородской области / О.В. Другова, Е.А. Павлов, А.П. Баврина, А.С. Благоданова, Н.В. Саперкин, О.В. Ковалишена // Медицинский альманах. – 2020. – Т. 63, № 2. – С. 27–36.
14. Тамм М.В. Коронавирусная инфекция в Москве: прогнозы и сценарии // Фармакоэкономика. Современная фармакоэкономика и фармакоэпидемиология. – 2020. – Т. 13, № 1. – С. 43–51. DOI: 10.17749/2070-4909.2020.13.1.43-51
15. Матвеев А.В. Математическое моделирование оценки эффективности мер против распространения эпидемии COVID-19 // Национальная безопасность и стратегическое планирование. – 2020. – Т. 1, № 29. – С. 23–39.
16. Godio A., Pace F., Vergnano A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence // Int. J. Environ. Res. & Public Health. – 2020. – Vol. 17, № 10. – P. 3535. DOI: 10.3390/ijerph17103535
17. Epidemic analysis of COVID-19 in China by dynamical modeling / L. Peng, W. Yang, D. Zhang, C. Zhuge, L. Hong // Med. Rxiv. Epidemiol. – 2020. – Vol. 18. – 11 p.
18. Pengpeng S., Shengli C., Peihua F. SEIR Transmission dynamics model of 2019 nCoV coronavirus with considering the weak infectious ability and changes in latency duration // Med. Rxiv. – 2020. – № 20. – 5 p. DOI: 10.1101/2020.02.16.20023655
19. Nikitina A.V., Lyapunova I.A., Dudnikov E.A. Study of the spread of viral diseases based on modifications of the SIR model // Computational mathematics and information technologies. – 2020. – Vol. 1, № 1. – P. 19–30. DOI: 10.23947/2587-8999-2020-1-1-19-30
20. Predicting the Cumulative Number of Cases for the COVID-19 Epidemic in China From Early Data / Z. Liu, P. Magal, O. Seydi, G.F. Webb // Populations and Evolution. – 2020. – Vol. 1, № 10. – 10 p. DOI: 10.20944/preprints202002.0365.v1
21. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan / F. Ndairou, I. Area, J.J. Nieto, D.F.M. Torres // Chaos, Solitons & Fractals. – 2020. – Vol. 135, № 6. – P. 109846. DOI: 10.1016/j.chaos.2020.109846
22. Куркин А.А., Куркина О.Е., Пелиновский Е.Н. Логистические модели распространения эпидемий // Труды НГТУ им. П.Е. Алексеева. – 2020. – Т. 129, № 2. – С. 9–18.
23. Fabiano N., Radenović S.N. On COVID-19 diffusion in Italy: data analysis and possible outcome // Vojnotehnički glasnik. – 2020. – Vol. 68, № 2. – P. 216–224. DOI: 10.5937/vojtehg68-25948
24. Cherniha R., Davydovych V. A mathematical model for the coronavirus COVID-19 outbreak [Электронный ресурс] // ArXiv. – 2020. – URL: <https://arxiv.org/abs/2004.01487v2> (дата обращения: 25.07.2020).
25. Кольцова Э.М., Куркина Е.С., Васецкий А.М. Математическое моделирование распространения эпидемии коронавируса COVID-19 в Москве // Computational nanotechnology. – 2020. – Т. 7, № 1. – С. 99–105.
26. Регрессионные модели прогнозирования количества летальных исходов при новой коронавирусной инфекции / Д.В. Мелик-Гусейнов, Н.Н. Карякин, А.С. Благоданов, В.И. Климко, А.П. Баврина, О.В. Другова, Н.В. Саперкин, О.В. Ковалишена // Современные технологии в медицине. – 2020. – Т. 12, № 2. – С. 6–13.
27. Возможности математического прогнозирования коронавирусной инфекции в Российской Федерации / И.А. Лакман, А.А. Агапитов, Л.Ф. Садикова, О.В. Черненко, С.В. Новиков, Д.В. Попов, В.Н. Павлов, Д.Ф. Гареева [и др.] // Артериальная гипертензия. – 2020. – Т. 26, № 3. – С. 288–294.
28. Fontes E. Моделирование в COMSOL Multiphysics распространения вируса COVID-19 [Электронный ресурс] // COMSOL. – 2020. – URL: <https://www.comsol.ru/blogs/modeling-the-spread-of-covid-19-with-comsol-multiphysics/> (дата обращения: 30.07.2020).
29. Getz W.M., Dougherty E.R. Discrete stochastic analog of Erlang epidemic models // J. of Biological Dynamics. – 2018. – Vol. 12, № 1. – P. 16–38. DOI: 10.1080/17513758.2017.1401677
30. The CDC portal: Forecasts of total deaths at the USA [Электронный ресурс] // Centers for Disease Control and Prevention. – 2020. – URL: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html> (дата обращения: 15.07.2020).
31. Мы изучили 858 ваших вопросов про коронавирус [Электронный ресурс] // Meduza. – 2020. – URL: <https://www.meduza.io/feature/2020/05/13> (дата обращения: 15.07.2020).
32. Официальный интернет-ресурс по вопросам коронавируса [Электронный ресурс] // Стопкоронавирус.рф. – 2020. – URL: <https://стопкоронавирус.рф> (дата обращения: 17.07.2020).
33. Алухтина Ю., Зобова С. Исследование о том, сколько больничных коек может спасти российскую медицину [Электронный ресурс] // Проект. – 2020. – URL: <https://www.proekt.media/research/koronavirus-regiony> (дата обращения: 01.07.2020).
34. Соколов А.А. Готово ли российское здравоохранение к борьбе с коронавирусом [Электронный ресурс] // Ведомости. – 2020. – URL: <https://www.vedomosti.ru/society/articles/2020/04/09/827471-gotovo-rossiiskoe> (дата обращения: 02.12.2020).
35. Free and occupied beds for Covid-19 patients in Germany [Электронный ресурс] // Coronavis. – 2020. – URL: <https://coronavis.dbvis.de/en/> (дата обращения: 01.07.2020).
36. Murray C.J.L. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months // Med. Rxiv. – 2020. – № 30. – 26 p. DOI: 10.1101/2020.03.27.20043752
37. COVID caseload calculator C5V [Электронный ресурс] // Weill Cornell Medicine. – 2020. – URL: <https://phs.weill.cornell.edu/cornell-covid-caseload-calculator-c5v> (дата обращения: 01.08.2020).
38. Aivazian S.A. Quality of life and living standards analysis: an econometric approach. – Berlin – Boston: De Gruyter, 2016. – 399 p. DOI: 10.1515/9783110316254
39. Степанов В.С. Индикатор уровня благосостояния населения Крыма и близких ему регионов: связь с факторными переменными // Вестник ЦЭМИ. – 2019. – № 2. – С. 8. DOI: 10.33276/S265838870004976-6
40. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – 3-е изд. – М.: Наука, 1983. – 416 с.
41. С апреля в коронавирусных больницах умерло не меньше 74,9 тысяч человек. Как регионы скрывают эти данные [Электронный ресурс] // Медиазона. – 2020. – URL: <https://zona.media/news/2020/11/30/75k> (дата обращения: 02.12.2020).

*Степанов В.С. Зависимость уровня смертности в регионах от распространенности активных носителей SARS-CoV-2 и ресурсов организаций здравоохранения // Анализ риска здоровью. – 2020. – № 4. – С. 12–22. DOI: 10.21668/health.risk.2020.4.02*

UDC 616.9–036.8; 51-76; 519.237  
DOI: 10.21668/health.risk/2020.4.02.eng

Read  
online



Research article

## DEPENDENCE BETWEEN MORTALITY IN REGIONS AND PREVALENCE OF ACTIVE SARS-CoV2 CARRIERS AND RESOURCES AVAILABLE TO PUBLIC HEALTHCARE ORGANIZATIONS

V.S. Stepanov

The Central Economics and Mathematics Institute of the Russian Academy of Sciences, 47 Nakhimovskii Ave., Moscow, 117418, Russian Federation

*The paper dwells on certain mathematical models showing how epidemics develop, namely, logistic ones, SIR-model, and some others. There is also a review of articles that focus on such models showing dynamics of incidence with COVID-19 infection. These models are often successfully applied for data collected in a whole country but on a regional level there are difficulties due to peculiarities of calculating mortality figures in Russia. In this case regression models can be useful with their obvious advantage at the initial stage in an epidemic process. They also include exogenous variables that influence mortality, for example, a number of doctors and nurses per a hospital, how well hospitals are equipped with ALV devices, and a number of available beds in them.*

*Our research goal was to build up a linear regression model that could be used as a basis for estimating regional mortality caused by COVID-19 as well as for more efficient distribution of all the resources mentioned above.*

*The model is built as per a set of resource parameters including data on «active cases». Preliminary three variables that showed data on resources available to communicable diseases departments in hospitals were transformed into a new single one via linear transformation. Then the model was tested on a training sample containing an endogenous variable on mortality and four factor ones including prevalence of active virus carriers. Regions were included into training data with different lags; they were included into such daily samples when death cases were registered rarely. Then the estimated model was applied with other values. It turned out to be quite efficient in estimating COVID-induced mortality for regions from trainings samples as well as for several others (for certain intervals).*

*As a result, we built a regression model and estimated its precision; the model showed a relation between mortality in a region and prevalence of active SARS-CoV2 carriers and availability of resources to hospitals in it. It can be useful when these resources are distributed. It can also be used to build SIRD, SEIR, and SEIRF models at a regional level when choosing parameters in them related to mortality. A methodology itself that can be similarly applied for other epidemic processes also deserves certain attention.*

**Key words:** regression model, mortality estimation, COVID-19, coronavirus infection, logistic equation, SEIR, SIR, ALV.

### References

1. Zhmerenetskii K.V., Sazonova E.N., Voronina N.V., Tomilka G.S., Sen'kevich O.A., Gorokhovskii V.S., D'yachenko S.V., Kol'tsov I.P., Kutsyi M.B. COVID-19: scientific facts. *Dal'nevostochnyi meditsinskii zhurnal*, 2020, no. 1, pp. 5–22 (in Russian).
2. Danilova I.A. Morbidity and mortality from COVID-19. The problem of data comparability. *Demograficheskoe obozrenie*, 2020, vol. 7, no. 1, pp. 6–26 (in Russian).
3. Drapkina O.M., Samorodskaya I.V., Sivtseva M.G., Kakorina E.P., Briko N.I., Cherkasov S.N., Tsinkerling V.A., Mal'kov P.G. COVID-19: urgent questions for estimating morbidity, prevalence, case fatality rate and mortality rate. *Kardiovaskulyarnaya terapiya i profilaktika*, 2020, vol. 19, no. 3, pp. 302–309 (in Russian).
4. Ivanov S. Mortality from COVID-19 against the backdrop of other twentieth century mortality bursts. *Demograficheskoe obozrenie*, 2020, vol. 7, no. 2, pp. 143–151 (in Russian).
5. Kozlovskii S., Boldyrev O. *Mozhno li predskazat' razvitiye pandemii koronavirusa? Ob"yasnyayem na primere Rossii* [Can we predict how coronavirus pandemic will develop? Let us explain using Russia as an example]. *BBC Russia*. Available at: <https://www.bbc.com/russian/features-52762747> (30.07.2020).
6. Glava Rosstat rasskazal, kak schitayut zhertv COVID-19 [Head of Rosstat has told how victims of COVID-19 are calculated in the country]. *BBC Russia*. Available at: <https://www.bbc.com/russian/features-53156041> (02.08.2020).
7. Koronavirus v Rossii: Infografika [Coronavirus in Russia: graphic data]. *Mediazona*, 2020. Available at: <https://zona.media/coronagraph> (30.07.2020).
8. Boyarintsev V.V., Pal'min R.S., Pal'min S.A., Pertsev S.F. Nique for predicting parameters of the epidemic process caused by COVID-19. *Kremlevskaya meditsina. Klinicheskii vestnik*, 2020, no. 2, pp. 14–21 (in Russian).
9. Kondrat'ev M.A. Forecasting methods and models of disease spread. *Komp'yuternye issledovaniya i modelirovaniye*, 2013, vol. 5, no. 5, pp. 863–882 (in Russian).
10. Magpantay F.M.G., Kosovalic N., Wu J. An age-structured population model with state-dependent delay: derivation and numerical integration. *SIAM Journal on Numerical Analysis*, 2014, vol. 52, no. 2, pp. 735–756. DOI: 10.1137/120903622
11. Khrapov P.V., Loginova A.A. Comparative analysis of the mathematical models of the dynamics of the coronavirus COVID-19 epidemic development in the different countries. *International Journal of Open Information Technologies*, 2020, vol. 8, no. 5, pp. 17–22.

© Stepanov V.S., 2020

**Vladimir S. Stepanov** – Candidate of Physical and Mathematical Sciences, Senior researcher (e-mail: [\\_stepanov@cemi.rssi.ru](mailto:_stepanov@cemi.rssi.ru), [vladstep0355@gmail.com](mailto:vladstep0355@gmail.com); tel.: +7 (916) 872-52-07; ORCID: <https://orcid.org/0000-0002-4478-376X>).

12. Minaev V.A., Sychev M.P., Vaitis E.V., Bondar' K.M. System-dynamic modeling of network information operations. *Inzhenernyye tekhnologii i sistemy*, 2019, vol. 29, no. 1, pp. 20–39 (in Russian).
13. Drugova O.V., Pavlov E.A., Bavrina A.P., Blagonravova A.S., Saperkin N.V., Kovalishena O.V. Statistic and dynamic aspects of the prediction of the COVID-19 spread in Nizhny Novgorod region. *Meditinskii al'manakh*, 2020, no. 2 (63), pp. 27–36 (in Russian).
14. Tamm M.V. COVID-19 in Moscow: prognoses and scenarios. *Farmakoekonomika. Sovremennaya farmakoekonomika i farmakoepidemiologiya*, 2020, vol. 13, no. 1, pp. 43–51. DOI: 10.17749/2070-4909.2020.13.1.43-51
15. Matveev A.V. The mathematical modeling of the effective measures against the COVID-19 spread. *Natsional'naya bezopasnost' i strategicheskoe planirovanie*, 2020, vol. 1, no. 29, pp. 23–39 (in Russian).
16. Godio A., Pace F., Vergnano A. SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *Int. J. Environ. Res. & Public Health*, 2020, vol. 17, no. 10, pp. 3535. DOI: 10.3390/ijerph17103535
17. Peng L., Yang W., Zhang D., Zhuge C., Hong L. Epidemic analysis of COVID-19 in China by dynamical modeling. *MedRxiv Epidemiol*, 2020, 11 p. (in Russian).
18. Pengpeng S., Shengli C., Peihua F. SEIR Transmission dynamics model of 2019 nCoV coronavirus with considering the weak infectious ability and changes in latency duration. *MedRxiv*, 2020, no. 20, 5 p. DOI: 10.1101/2020.02.16.20023655
19. Nikitina A.V., Lyapunova I.A., Dudnikov E.A. Study of the spread of viral diseases based on modifications of the SIR model. *Computational mathematics and information technologies*, 2020, vol. 1, no. 1, pp. 19–30. DOI: 10.23947/2587-8999-2020-1-1-19-30
20. Liu Z., Magal P., Seydi O., G.F. Webb Predicting the Cumulative Number of Cases for the COVID-19 Epidemic in China From Early Data. *Populations and Evolution*, 2020, vol. 1, no. 10, 10 p. DOI: 10.20944/preprints202002.0365.v1
21. Ndaïrou F., Area I., Nieto J.J., Torres D.F.M. Mathematical modeling of COVID-19 transmission dynamics with a case study of Wuhan. *Chaos, Solitons & Fractals*, 2020, vol. 135, no. 6, pp. 109846. DOI: 10.1016/j.chaos.2020.109846
22. Kurkin A.A., Kurkina O.E., Pelinovskii E.N. Logistic models of epidemic growth. *Trudy NGTU im. R.E. Alekseeva*, 2020, no. 2 (129), pp. 9–18 (in Russian).
23. Fabiano N., Radenović S.N. On COVID-19 diffusion in Italy: data analysis and possible outcome. *Vojnotehnički glasnik*, 2020, vol. 68, no. 2, pp. 216–224. DOI: 10.5937/vojtehg68-25948
24. Cherniha R., Davydovych V. A mathematical model for the coronavirus COVID-19 outbreak. *ArXiv*, 2020. Available at: <https://arxiv.org/abs/2004.01487v2> (25.07.2020).
25. Kol'tsova E.M., Kurkina E.S., Vasetskii A.M. Mathematical modeling of the spread of COVID-19 in Moscow. *Computational nanotechnology*, 2020, vol. 7, no. 1, pp. 99–105 (in Russian).
26. Melik-Guseinov D.V., Karyakin N.N., Blagonravov A.S., Klimko V.I., Bavrina A.P., Drugova O.V., Saperkin N.V., Kovalishena O.V. Regression models predicting the number of deaths from the new coronavirus infection. *Sovremennye tekhnologii v meditsine*, 2020, vol. 12, no. 2, pp. 6–13 (in Russian).
27. Lakman I.A., Agapitov A.A., Sadikova L.F., Chernenko O.V., Novikov S.V., Popov D.V., Pavlov V.N., Gareeva D.F. [et al.]. COVID-19 mathematical forecasting in the Russian Federation. *Arterial'naya gipertenziya*, 2020, vol. 26, no. 3, pp. 288–294 (in Russian).
28. Fontes E. Modelirovanie v COMSOL Multiphysics rasprostraneniya virusa COVID-19 [Modeling the spread of the COVID-19 virus in COMSOL Multiphysics]. *COMSOL*, 2020. Available at: <https://www.comsol.ru/blogs/modeling-the-spread-of-covid-19-with-comsol-multiphysics/> (30.07.2020) (in Russian).
29. Getz W.M., Dougherty E.R. Discrete stochastic analog of Erlang epidemic models. *J. of Biological Dynamics*, 2018, vol. 12, no. 1, pp. 16–38. DOI: 10.1080/17513758.2017.1401677
30. The CDC portal: Forecasts of total deaths at the USA. *Centers for Disease Control and Prevention*, 2020. Available at: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/forecasting-us.html> (15.07.2020).
31. My izuchili 858 vashikh voprosov pro koronavirus [We have examined 858 questions you have about coronavirus]. *Meduza*, 2020. Available at: <https://www.meduza.io/feature/2020/05/13> (15.07.2020) (in Russian).
32. Ofitsial'nyi internet-resurs po voprosam koronavirusa [The official web-site on issues related to coronavirus]. *Stopkoronavirus.rf*, 2020. Available at: <https://stopkoronavirus.rf> (17.07.2020) (in Russian).
33. Apukhtina Yu., Zobova S. Issledovanie o tom, skol'ko bol'nichnykh koek mozhet spasti rossiiskuyu meditsinu [Research on a number of available beds in hospitals that can save public healthcare in Russia]. *Proekt*, 2020. Available at: <https://www.proekt.media/research/koronavirus-regiony> (01.07.2020) (in Russian).
34. Sokolov A.A. Gotovo li rossiiskoe zdavookhranenie k bor'be s koronavirusom [Is public healthcare in Russia ready to fight coronavirus?]. *Vedomosti*, 2020. Available at: <https://www.vedomosti.ru/society/articles/2020/04/09/827471-gotovo-rossiiskoe> (02.12.2020) (in Russian).
35. Free and occupied beds for COVID-19 patients in Germany. *Coronavis*, 2020. Available at: <https://coronavis.dbvis.de/en/> (01.07.2020).
36. Murray C.J.L. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator-days and deaths by US state in the next 4 months. *MedRxiv*, 2020, no. 30, 26 p. DOI: 10.1101/2020.03.27.20043752
37. COVID caseload calculator C5V. *Weill Cornell Medicine*, 2020. Available at: <https://phs.weill.cornell.edu/cornell-covid-caseload-calculator-c5v> (01.08.2020).
38. Aivazian S.A. Quality of life and living standards analysis: an econometric approach. Berlin/Boston, De Gruyter Publ., 2016, 399 p. DOI: 10.1515/9783110316254
39. Stepanov V.S. integral indicator of the living conditions at the Crimea Republic and some other territories: dependence from factorial variables. *Vestnik TsEMI*, 2019, no. 2, pp. 8. DOI: 10.33276/S265838870004976-6
40. Bol'shev L.N., Smirnov N.V. Tablitsy matematicheskoi statistiki [Mathematical statistics tables]. 3-th edition. Moscow, Nauka Publ., 1983, 416 p. (in Russian).
41. S aprelya v koronavirusnykh bol'nitsakh umerlo ne men'she 74,9 tysyach chelovek. Kak region skryvayut eti dannye [Not less than 74.9 thousand people have died in coronavirus hospitals since April. How do regions manage to hide these data?]. *Mediazona*, 2020. Available at: <https://zona.media/news/2020/11/30/75k> (02.12.2020).

Stepanov V.S. Dependence between mortality in regions and prevalence of active SARS-COV2 carriers and resources available to public healthcare organizations. *Health Risk Analysis*, 2020, no. 4, pp. 12–22. DOI: 10.21668/health.risk.2020.4.02.eng

Получена: 18.08.2020

Принята: 23.11.2020

Опубликована: 30.12.2020