

1. НАУЧНЫЕ СТАТЬИ

1.1. ОБ АНАЛИЗЕ, МОДЕЛИРОВАНИИ И ПРОГНОЗЕ СТАТИСТИКИ ЭПИДЕМИИ КОРОНАВИРУСА В РОССИИ

Абрамов С. М., Травин С.О.

В марте 2020 года стало очевидно, что эпидемия коронавируса в России будет иметь достаточно значительные масштабы и продолжительность. Для составления личных планов или планов небольших учреждений надо было как-то отвечать в каждый момент времени на простые вопросы: что нас ждет в ближайшее время? На какой фазе эпидемии мы находимся? Когда все это закончится? По сути, стало необходимым строить суждения о динамике эпидемии, рассматривая поступающую официальную статистику об эпидемии, моделируя и прогнозируя поведение этой статистики. Данная работа посвящена методам и результатам такого анализа, моделирования и прогноза.

Введение

В марте 2020 года стало очевидно, что эпидемия коронавируса в России будет иметь достаточно значительные масштабы и по количеству заболевших, и по продолжительности. Для составления личных планов и планов учреждений надо было как-то отвечать в каждый момент времени на простые вопросы о динамике развития эпидемии:

Что нас ждет в ближайшее время?

На какой фазе эпидемии мы находимся?

Когда все это закончится?

Для выработки ответов на эти вопросы была доступна только официальная ежедневная статистика: количество выявленных заболеваний, выздоровлений и смертей. По сути, встал вопрос: как можно анализировать и прогнозировать данную статистику. В силу объективных¹ и субъективных² факторов, доступная статистика далека от неизвестных истинных данных про эпидемию. Поэтому, строго говоря, *предстояло моделировать и прогнозировать не динамику эпидемии, а динамику официальной статистики об эпидемии*. Именно так и написано в заголовке данной статьи. По этой же причине мы не пытались привязать численные параметры применяемых моделей к объясняемым величинам, а всюду, где только возможно (а возможным оказалось везде), мы ушли от численного интегрирования дифференциальных уравнений, обычно лежащих в основе известных моделей 1–2, и сразу воспользовались аналитическими формулами из них вытекающими. Далее в подразделах введения мы дадим общее описание используемых нами подходов и методов решения проблемы.

Excel — инструмент анализа, моделирования и прогноза

Работа велась в инициативном порядке. В качестве инструмента для реализации анализа, моделирования и прогноза использовался Microsoft Excel 3–6, и в дальнейшем такой выбор не разочаровал.

*Описываемые в данной работе исходные данные, методы и результаты доступны как Excel-файл (с макро, *.xlsm) по ссылкам 7–8. Все исходные данные берутся для России, и результаты их обработки справедливы для России. Пока длится эпидемия коронавируса в России, данный файл будет обновляться ежедневно, обычно около полудня. То есть, по указанной ссылке файл каждый день будет новый — обновленный — его можно просмотреть в веб-браузере либо скачать для просмотра в Microsoft Excel.*

Исходные данные для анализа, моделирования и прогноза

В качестве входных данных берется официальная статистика. В последнее время эта информация публикуется на сайте «стопкоронавирус.рф»⁹. Входные данные с этого сайта берутся один раз в день. При этом (Рисунок 1, Таблица 1) с сайта копируется ровно четыре значения T_i , V_i , H_i , и D_i . (см. рис.1). Таким образом, исходная официальная статистика является временным рядом (time-value): $\{... (T_i, V_i, H_i, D_i) ... \}$.

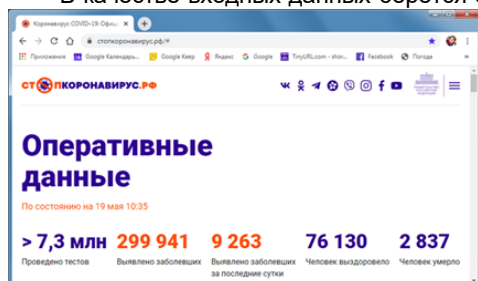


Рисунок 1. Входные данные с официального сайта «стопкоронавирус.рф»

¹ — Несовершенство тестов и методик, наличие невыявляемых (в том числе, бессимптомных) больных и т.п.

² — Нежелание показывать «плохую» статистику, возможные манипуляции с задержкой передачи данных.

Таблица 1. Входные данные с официального сайта «стопкоронавирус.рф»

Обозначение для формул — пояснение (название параметра в Excel-файле)	Пример данных с сайта
T_i — Дата и время объявления данных (дата)	«По состоянию на 19 мая 10:35»
V_i — Нарастающий итог, число зарегистрированных заболеваний (заболеваний)	«299 941 Выявлено заболевших»
H_i — Нарастающий итог, число выздоровевших (выздоровело)	«76 130 Человек выздоровело»
D_i — Нарастающий итог, число умерших (умерло)	«2 837 Человек умерло»

Прямое вычисление дополнительных величин

По входным данным T_i , V_i , H_i , и D_i дополнительно рассчитываются следующие величины:

- $E_i = H_i + D_i$ — «выздоровело + умерло», сколько человек закончило болеть. В конце эпидемии можно ожидать, что все выявленные заболевшие либо умрут, либо выздоровеют³: $V_i = E_i = H_i + D_i$, если i соответствует окончанию эпидемии. Кроме того, каждое выявленное заболевание (каждая единица в V_i), судя по всему, в среднем за 13–18–23 дня завершается либо как единица к «умерло» D_i , либо как единица к «выздоровело» H_i . То есть, $E_i = V_{i-shift}$, где $shift$ — число (количество дней) в диапазоне 13–15–23. Эта величина $shift$ также вычисляется в Excel-файле (вкладка «Входные данные», колонка «Дней лечения») и видно, что по мере развития эпидемии это число растет: от 14–13 дней – в начале апреля до 23–24 дней – в середине мая. Указанный факт иллюстрирует рост нагрузки на медицинскую систему: скорее всего, не само лечение затягивается, а срок установки и регистрации факта выздоровления.
- $S_i = V_i - E_i$ — «пока больны», сколько человек пока еще не закончило болеть.
- $\dot{V}_i = \frac{V_i - V_{i-1}}{T_i - T_{i-1}}$ — «заболеваний в день», — средняя скорость изменения V_i за сутки. Здесь надо учесть, что дата-время в Excel означает вещественное число, равное количеству суток, прошедших с начала 1900 года. А разница двух дат-времен $T_i - T_{i-1}$ дает нам вещественное число, означающее количество (необязательно целое) суток между ними. Например:
 - $T_{12} = 27.03.2020\ 11:00:00, V_{12} = 1\ 036$
 - $T_{13} = 28.03.2020\ 21:00:00, V_{13} = 1\ 264$
 - $T_{13} - T_{12} = 1.417, V_{13} - V_{12} = 228$
 - $\dot{V}_{13} = \frac{228}{1.417} = 160.94$ — заболеваний в день, средняя скорость изменения V_i за 13-е сутки
- $\dot{H}_i, \dot{D}_i, \dot{E}_i, \dot{S}_i$ — «выздоровело в день», «умерло в день», «(выздоровело + умерло) в день», «пока больны в день» и другие скорости (изменения величин за сутки) — определяются и вычисляются аналогично \dot{V}_i .

Общее описание динамики эпидемии, модель и прогноз

Эпидемия коронавируса в России является завозной. То есть, весь процесс можно описать следующим образом. Среди всего населения появляется несколько больных людей, которые, контактируя со здоровыми людьми, заражают их. В эпидемиологии используют параметр R_0 — среднее количество здоровых людей, которых заражает один больной. Считается, что для коронавируса $R_0 = 2 \dots 3$.

Итак, изначальные больные заражают (в среднем) каждый по R_0 здоровых. Эти люди заболевают и заражают (в среднем) каждый по R_0 других здоровых и так далее... Описание соответствует тому, что наблюдается в любых иных ветвящихся процессах (цепных реакциях). А это значит, что процесс будет развиваться в геометрической прогрессии или экспоненциально во времени: V_i и \dot{V}_i будут расти по экспоненте, а показатель экспоненты будет определяться реальным значением R_0 в текущих обстоятельствах⁴.

Экспонента является быстрорастущей и неограниченной функцией. Значит, V_i не может неограниченно долго развиваться по экспоненте — иначе бы V_i превысила общую численность населения. Таким образом, скорость изменения V_i — величина \dot{V}_i «заболеваний в день», — которая в начале эпидемии

³ — Из-за национальных особенностей медицинской статистики это не совсем так. Некоторые выявленные заболевания по разным причинам могут не попасть в дальнейшем ни в категорию «выздоровело», ни в категорию «умерло». Но мы пока этим пренебрегаем, рассчитывая, что таких исключений будет немного. Подробнее данные нюансы рассматриваются ниже, в разделе «Моделирование и прогнозирование количества смертей в эпидемии коронавируса в России», стр. 18.

⁴ — На значение R_0 могут влиять самые различные факторы: погодные условия, плотность и «коммуникабельность» населения, действия властей — блокады мест с заражениями, иные изоляционные и карантинные меры. В конечном итоге на скорость распространения эпидемии влияет не сам итоговый параметр числа зараженных, а темп заражения в единицу времени (например, за сутки). Очевидно R_0 есть не что иное, как произведение темпа передачи инфекции на среднюю продолжительность болезни, от момента подхвата инфекции до полного выздоровления или смерти больного.

тоже растет по экспоненте, постепенно должна замедлять свой рост⁵, пройти свой максимум и постепенно уменьшаться до нуля. Характер поведения \dot{V}_i («шляпа») и V_i (S-образный профиль) очевиден из этих, самых простых соображений, и он должен быть примерно таким, как показано ниже — см. *Рисунок 2*. Наконец, можно ожидать, что графики величин E_i «выздоровело + умерло» и \dot{E}_i «(выздоровело + умерло в день)» будут совпадать с графиками величин V_i и \dot{V}_i , сдвинутыми вправо на величину $shift$ средней продолжительности болезни. Имея величины V_i , \dot{V}_i , E_i и \dot{E}_i , легко определить и $S_i = V_i - E_i$ «пока больны» и $\dot{S}_i = \dot{V}_i - \dot{E}_i$ «пока больны в день».

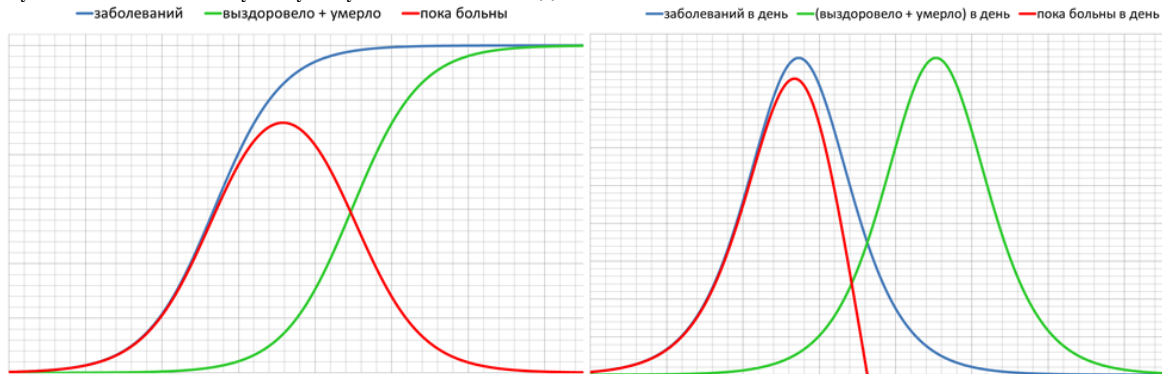


Рисунок 2. Примерные ожидаемые (из самых общих соображений) формы графиков величин V_i «заболеваний», E_i «выздоровело + умерло», S_i «пока больны» (слева), \dot{V}_i «заболеваний в день» и \dot{E}_i «(выздоровело + умерло в день)», \dot{S}_i «пока больны в день» (справа), ось абсцисс — время, ось ординат — значения величин

Ключевые события и фазы эпидемии. И из этих же, самых общих соображений, мы можем выделить ключевые события, которые легко распознать на графиках и которые разбивают все течение эпидемии на различные этапы. Вот эти события:

- максимум \dot{S}_i — пиковое превышение госпитализации над выпиской;
- максимум \dot{V}_i — пиковая нагрузка на процесс госпитализации;
- максимум S_i — пиковое количество больных;
- $\dot{V}_i = \dot{E}_i$ — перекрестье, после него в день прекращают заболеть больше, чем выявляют новых больных; перелом в эпидемии, нагрузка на медицинскую систему затем начинает падать;
- максимум \dot{E}_i — пиковая нагрузка на процесс выписки;
- максимум V_i — заболели все, кому было суждено;
- максимум E_i — все закончили болеть, конец эпидемии.

В начале эпидемии мы имеем фактическую статистику, которая соответствует начальному отрезку течения эпидемии. И мы можем нанести на координатную плоскость только самые начальные точки графиков. Дальнейший анализ основан на моделировании и прогнозе.

Основная идея моделирования. Будем использовать некоторую математическую функцию $f(t)$ от времени t , природа которой бы соответствовала динамике эпидемии (это обсуждается ниже) и $f(T_i)$ приближало бы величины V_i . Как правило, такие функции, кроме аргумента (t), содержат некоторые параметры, значения которых подбирают так, чтобы на накопленной статистике $\{(T_i, V_i) \mid i = [1 \dots n]\}$ функция легла бы максимально тесно к имеющимся данным: $f(T_i) \approx V_i$, для $i = [1 \dots n]$. Понятие «легла бы максимально тесно» можно формализовать по-разному. Мы будем использовать метод наименьших квадратов (МНК): значения параметров функции $f(t)$ подбирают так, чтобы минимизировать величину: $\sum_{i=1}^n (f(T_i) - V_i)^2$.

Переход от модели к прогнозу. После подбора модели — подбора функции $f(t)$, приближающей максимально тесно имеющиеся данные, — мы предполагаем, что и для последующих моментов наблюдения — за рамками статистики, уже в будущем, $i > n$, — будет тоже выполняться $f(T_i) \approx V_i$.

Иногда мы можем ожидать выполнения условия $f(T_i) \approx V_i$ только для ближайших моментов времени, например, на предстоящую неделю. Тогда наша модель позволяет построить только краткосрочный прогноз.

Если же мы можем ожидать, что $f(T_i) \approx V_i$ для большого отрезка времени, то мы можем продолжить процесс моделирования до нахождения в модельном времени (в будущем) всех ключевых точек и фаз эпидемии. И тем самым определить даты наступления ключевых точек и ожидаемые значения величин V_i , E_i , S_i , \dot{V}_i , \dot{E}_i , \dot{S}_i в эти даты. Тем самым мы получим прогноз временных и объемных масштабов эпидемии.

⁵ — Можно считать, что у нас соответственно изменяются коэффициенты в экспоненте. Или уменьшается R_0 .

На этом мы закончим введение с описанием общих подходов и методов. Далее, в разделах основной части статьи, будут рассмотрены использованные нами модели и результаты прогнозирования на их основе.

Простейшая модель — экспонента

В начале развития эпидемии с высокой точностью с течением времени V_i (как и \dot{V}_i) растет по экспоненте. И хорошей моделью является функция $f(t) = A \cdot \exp(C \cdot t)$. Остается подобрать константы A и C из соображения «функция легла бы тесно с имеющимися данными: $f(T_i) \approx V_i$ ».

На таком подходе был реализован краткосрочный прогноз. Расчетная схема даже была упрощена: взяв логарифм от левой и правой части условия $f(T_i) \approx V_i$, мы получим: $C \cdot T_i + \ln A \approx \ln V_i$. Построим линейную регрессию $C \cdot t + \ln A$ для набора точек на двумерной плоскости $\{(T_i, \ln V_i) \mid i = 1..n\}$ и определим, таким образом, коэффициенты C и A в нашей модели.

При реализации такого подхода надо учитывать: он применим только на начальном этапе эпидемии и на коротком отрезке времени — только при этих условиях с высокой точностью с течением времени V_i (как и \dot{V}_i) растет по экспоненте. Поэтому, для построения линейной регрессии бралась не вся накопленная статистика, а только за семь последних дней $\{(T_i, \ln V_i) \mid i = (n - 6)..n\}$, и построенную модель можно рекомендовать использовать для прогноза только на следующую неделю.

Однако даже такая простая модель на начальном этапе развития эпидемии давала достаточно точный краткосрочный прогноз: применение ее к фактическим данным $\{\dots (T_i, V_i) \dots\}$ за семь дней 1–7.04.2020 позволило предсказать значения V_i на следующие 7 дней с относительной погрешностью $[-5.6\% \dots - 0.3\%]$.

Вторая модель — линейное уменьшение коэффициента экспоненты

Как было отмечено во введении, по мере развития эпидемии скорость роста V_i должна уменьшаться. Поэтому, если мы моделируем V_i экспонентой — функцией $f(t) = A \cdot \exp(C \cdot t)$, — то коэффициент C не должен быть постоянным, он должен уменьшаться. В соответствии с этим, второй соавтор предложил модель, которая использовалась нами до 26.04.2020. В этой модели значения V_i приближались экспонентой с линейно уменьшающимся коэффициентом $C(t) = C_0 - C_1 \cdot t$.

То есть, $f(t) = A \cdot \exp((C_0 - C_1 \cdot t) \cdot t)$, параметры A , C_0 , C_1 подбирались методом наименьших квадратов при помощи встроенного солвера Excel⁶, минимизировалась величина $\sum_{i=1}^n (f(T_i) - V_i)^2$, рассчитываемая по всей накопленной статистике $\{(T_i, V_i) \mid i = 1 \dots n\}$. Модель позволяла не только строить более точные краткосрочные прогнозы, но и моделировать весь процесс до момента завершения эпидемии, получая оценки сроков и значений величин для всех ключевых точек в эпидемии. Однако мы не будем подробно на этом останавливаться, поскольку эта модель была заменена более совершенной, описанной в следующем разделе.

Третья модель — гиперболический тангенс

В дальнейшем вторым соавтором была предложена и обоснована новая модель. Дело в том, что все (почти без исключения) саморазгоняющиеся процессы на поле с ограниченными ресурсами, от автокатализа до размножения кроликов, подчиняются одному и тому же закону: они описываются одной и той же функцией $f(t)$. Эпидемия в замкнутом обществе людей — не исключение. В работах по экологии и смежным областям эту функцию называют сигмоидой, логистической функцией, функцией Ферми–Дирака. Сигмоида $f(t)$ — гладкая монотонная возрастающая нелинейная функция, график которой (см. Рисунок 3, синюю или зеленую линию) напоминает форму буквы «S», и которая растет от 0 до некоторой величины $2N_0$. Сигмоида через гиперболический тангенс (th) выражается следующим образом:

$$f(t) = N_0 \cdot \left(1 + \operatorname{th}(k \cdot (t - \tau))\right) \quad (1)$$

Суть третьей модели состоит в том, что мы будем приближать V_i именно этой функцией, подбирая МНК оптимальные значения параметров N_0 , k , τ .

Физический смысл параметров N_0 , k , τ прозрачен:

- Параметр N_0 — это половина общего числа всех зараженных во время эпидемии.
- Параметр k — отвечает за темпы протекания эпидемии — это коэффициент сжатия или растяжения времени t . Карантинные меры могут влиять на k .
- Параметр τ — сдвиг по времени. В разных странах коронавирус завезли в разное время.

Предложенная модель тесно связана с природой эпидемии. А именно: неограниченная обычная экспонента $f(t) = A \cdot \exp(C \cdot t)$ является решением дифференциального уравнения: $\dot{f} = C \cdot f$ — скорость (роста) f прямо пропорциональна текущей величине f . Переформулируем для эпидемии: скорость появления новых заболеваний тем выше, чем больше у нас уже заболевших. Это верно на начальных этапах, но неверно, когда мы приближаемся к насыщению — к предельному числу заболеваний в данной эпидемии. Легко проверить прямой подстановкой (выкладки опускаем), что логистическая функция (1) является решением дифференциального уравнения:

⁶ — Меню/ Данные/ Поиск решения.

$$\dot{f} = \frac{k}{N_0} \cdot (2N_0 - f) \cdot f \quad (2)$$

Из (2) видно, что если f много меньше $2N_0$, то величину $C = \frac{k}{N_0} \cdot (2N_0 - f)$ можно считать постоянной и уравнение (2), по сути, совпадает с дифференциальным уравнением экспоненты $\dot{f} = C \cdot f$ с коэффициентом $C = 2k$ — скорость (роста) f прямо пропорциональна текущей величине f . Но по мере приближения f к насыщению — к $2N_0$, — скорость f будет уменьшаться, стремясь к нулю.

Как указал Аркадий Климов в частной переписке, уравнение (2) является уравнением развития эпидемии в самой простой модели эпидемии «SI» — когда рассматриваются только категории S (susceptible — восприимчивый) и I (infected — зараженный), без рассмотрения категорий R (recovered — выздоровел) и E (exposed — заражающий) ². Таким образом, использование функции (1) для приближения величин V_i — это переход к классической модели SI или SIR развития эпидемии. По нашему мнению, для России рассматривать более сложные модели, например, SEIR, не имеет смысла, потому что:

- входные данные для моделирования недостаточно полны и достоверны — имеются проблемы с диагностикой и регистрацией (как заболевания, так и выздоровления или причин смерти), и присутствует искажение статистики по внутривидовым мотивам;
- по сути, приходится моделировать не саму эпидемию (для нее математическую модель построить можно), а ту информацию об эпидемии, которую службы страны смогут и захотят сделать публичной.

Прогнозирование количества выявленных заболевших

После 26.04.2020 прогноз количества выявленных заболевших в нашем Excel-файле был переведен на использование функции $f(t)$, задаваемой уравнением (1). При этом, параметры N_0 , k и τ находятся методом наименьших квадратов (МНК) при помощи встроенного солвера Excel. То есть, параметры N_0 , k и τ находятся из условия минимизации величины $\sum_{i=1}^n (f(T_i) - V_i)^2$, рассчитываемой по всей накопленной статистике $\{(T_i, V_i) \mid i = 1 \dots n\}$.

После подбора параметров функции $f(t)$, приближающей максимально тесно имеющиеся данные, — мы предполагаем, что для последующих⁷ моментов наблюдения — за рамками статистики, уже в будущем, $i > n$, — будет тоже выполняться прогноз $V_i = f(T_i)$.

Прогнозирование остальных значений

Мы уже отмечали, что график величины E_i «выздоровело + умерло» скорее всего будет совпадать с графиком величины V_i , сдвинутым вправо на некоторую величину *shift* — среднюю продолжительность болезни. Практика показала (обсуждено выше), что средняя продолжительность болезни не является постоянной: по мере развития эпидемии растет нагрузка на систему здравоохранения и это, по всей видимости, влечет увеличение продолжительности болезни. Поэтому, разумнее приближать величины E_i своей логистической функцией:

$$g(t) = N_0 \cdot \left(1 + \text{th}(k_1 \cdot (t - \tau_1))\right), \quad (3)$$

которая имеет тот же самый параметр N_0 , что и $f(t)$ в формуле (1), и свои собственные параметры k_1 и τ_1 — коэффициент растяжения/сжатия времени и сдвиг по времени. Параметры k_1 и τ_1 находятся МНК из условия минимизации величины $\sum_{i=1}^n (g(T_i) - E_i)^2$, рассчитываемой по всей накопленной статистике $\{(T_i, E_i) \mid i = 1 \dots n\}$.

После подбора параметров функции $g(t)$, приближающей максимально тесно имеющиеся данные, — мы предполагаем, что для последующих моментов наблюдения — за рамками статистики, уже в будущем, $i > n$, — будет тоже выполняться прогноз $E_i = g(T_i)$.

Рассчитав прогнозные значения $\{... (T_i, V_i, E_i) \dots\}$ на любой момент T_i будущего, мы вычисляем прогнозы величины $S_i = V_i - E_i$ — «пока больны». Ниже (Рисунок 3) показаны графики V_i , E_i , S_i . Круглые точки на графике — значения из реальной статистики, непрерывные линии — аппроксимирующие функции и прогноз на будущее.

Посмотрите на Рисунок 3. Реальная статистика объемных величин V_i , E_i , S_i имеет разумное поведение, и поэтому при помощи МНК удастся тесно прижать аппроксимирующие линии к реальной статистике. И это позволяет относиться с большим доверием к последующим прогнозам.

⁷ — Если T_n время последних доступных статистических данных — «за сегодня», — то удобно последующие моменты времени в будущем определить с шагом «одни сутки»: $T_{n+1} = T_n + 1$, $T_{n+2} = T_n + 2$ и т.д.

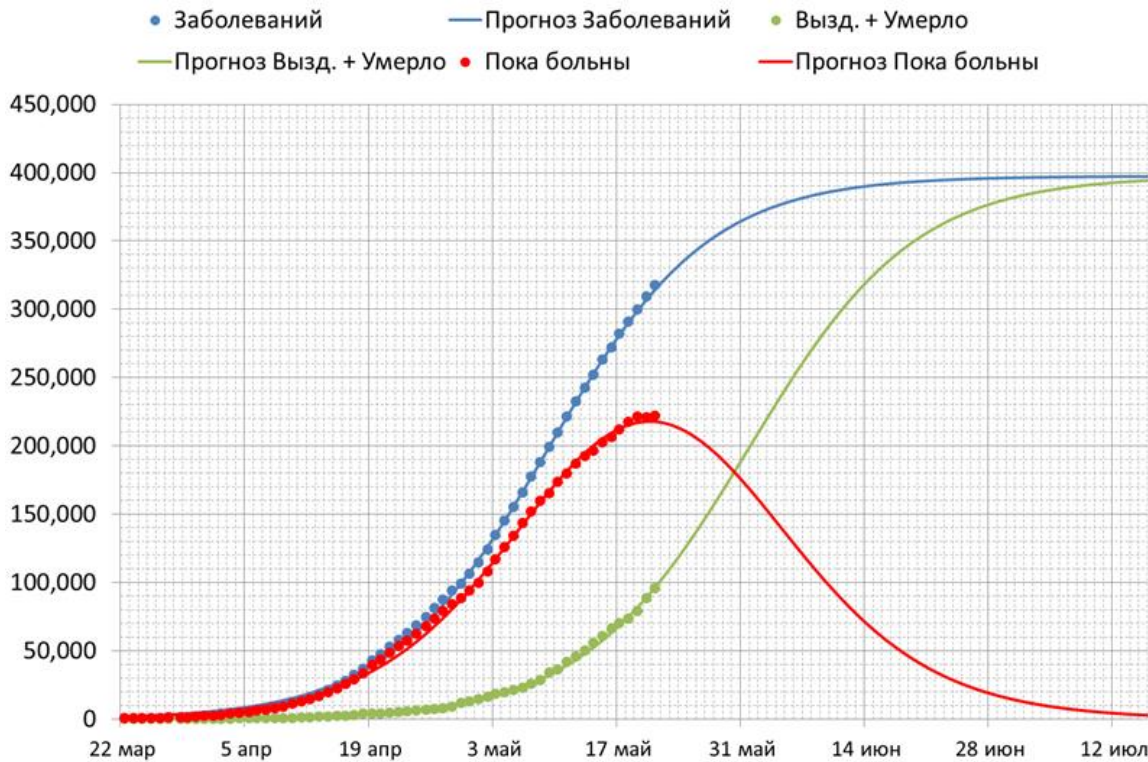


Рисунок 3. Прогноз объемных величин V_i , E_i , S_i , выполненный на основе статистики 16.03.2020–21.05.2020. Круглые точки — значения из реальной статистики, непрерывные линии — аппроксимирующие функции и прогноз на будущее

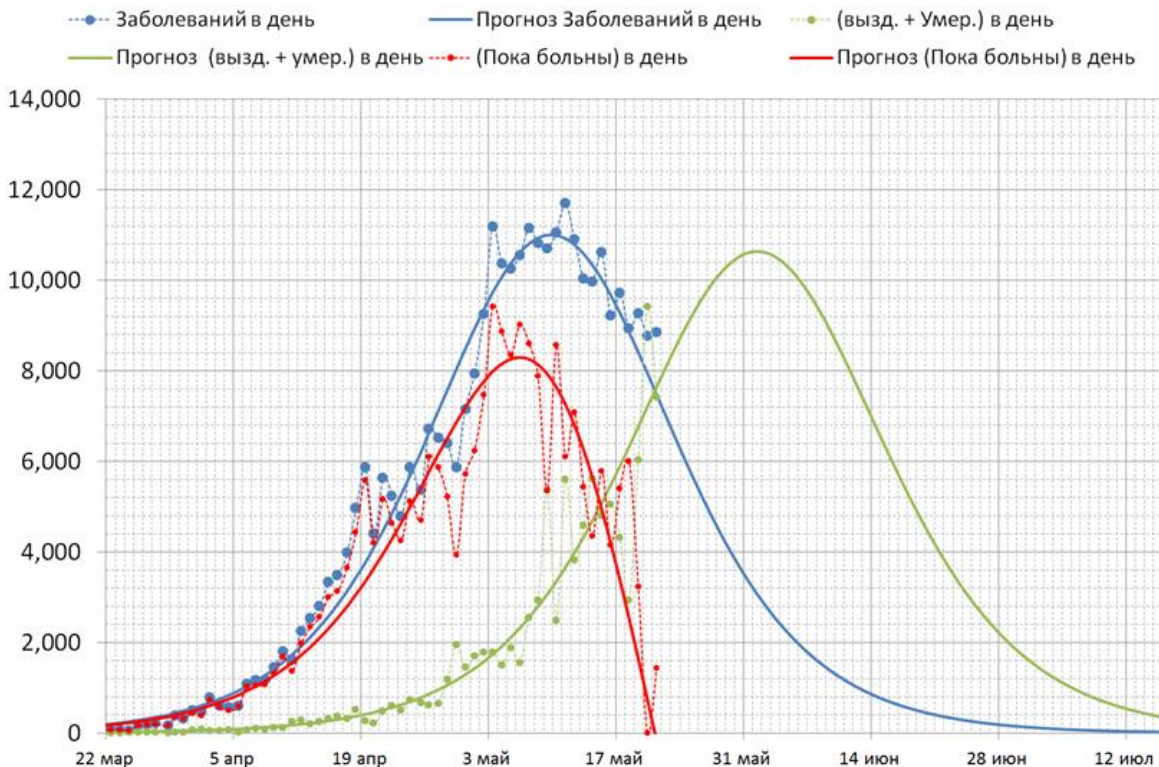


Рисунок 4. Прогноз скоростей \dot{V}_i , \dot{E}_i , \dot{S}_i , выполненный на основе статистики 16.03.2020–21.05.2020. Круглые точки — значения из реальной статистики, непрерывные линии — аппроксимирующие функции и прогноз на будущее

Высокую точность краткосрочного прогноза, построенного на базе третьей модели — на базе логистических функций, — подтвердила и практика. Так, аппроксимирующая функция $f(t)$, построенная по

формуле (1) с МНК для набора статистики за период 16.03.2020–14.05.2020, в последующие 7 дней (целая неделя) предсказывала фактические значения V_i «заболеваний» с точностью от +1.2% до +3.7%. Зная $\{... (T_i, V_i, E_i, S_i) ... \}$ на любой момент T_i будущего, мы вычисляем прогнозы скоростей изменения этих величин, доля изменения величин за отрезок времени $[T_i, T_{i+1}]$ на длину этого отрезка.

Ниже (Рисунок 4) показаны графики $\dot{V}_i, \dot{E}_i, \dot{S}_i$. Круглые точки на графике — значения из реальной статистики. Непрерывные линии — аппроксимирующие функции и прогноз на будущее. Видно, что реальная статистика скоростей $\dot{V}_i, \dot{E}_i, \dot{S}_i$ имеет большой разброс в разные стороны от аппроксимирующих линий. И такое хаотичное поведение трудно чем-либо объяснить, кроме проблем с входными данными: вероятные трудности с диагностикой (как заболевания, так и выздоровления или причин смерти) и/или искажение статистики по внутриполитическим мотивам.

Поэтому, приемлемый по точности краткосрочный прогноз следует ожидать только на небольшой период и только для объемных величин V_i, E_i, S_i , как это было проиллюстрировано выше, но не для среднесуточных скоростей $\dot{V}_i, \dot{E}_i, \dot{S}_i$.

Дальние прогнозы, их точность и смысл

Формально мы можем продлить аппроксимирующие функции третьей модели (Рисунок 3 и Рисунок 4) сколь угодно далеко в будущее — до завершения эпидемии. Но как доверять таким прогнозам? На что можно достоверно рассчитывать и использовать? Ясно, что действовать надо осторожно, понимая, что на дальнюю перспективу мы можем получить только *общее представление* об объемах и сроках (датах).

Мы используем такой подход: смотрим на построенные модели и прогнозы и выписываем *ключевые события* из модели эпидемии — их моменты времени (даты) и достигнутые при этом величины (значения). Выше (Рисунок 3 и Рисунок 4) эти события отмечены красными значками с номером, а пояснения, значения и даты ключевых событий приведены ниже (Таблица 2).

Таблица 2. Основные результаты дальнего прогноза, выполненного 21.05.2020 на массиве данных за отрезок времени 16.03.2020–21.05.2020: значения величин и даты наступления шести ключевых событий в прогнозе протекания эпидемии

№	Краткое пояснение	обозначение	ключевого события	и Значение	Дата
1	«Макс Заболело» — Достигнуто максимальное значение числа заболевших. Поставлен последний диагноз в эпидемии. Больше заболевших не будет появляться (Рисунок 3, метка 1)			397,443	11 сентября
2	«Все выздоровеют» — Немного неточно сказано: все, кто заболел — все закончат болеть, либо выздоровеют, либо умрут. Больных больше не останется (Рисунок 3, метка 2)			397,443	8 октября
3	«Макс Пока больны» — Максимальное значение тех, кто заболел, но еще не закончил болеть — пока не выздоровел и не умер (Рисунок 3, метка 3)			217,934	20 мая
4	«Макс Пока больны в день» — Максимальное значение скорости изменения (в день) величины «Пока больны» (Рисунок 4, метка 4)			8,295 в день	6 мая
5	«Макс Заболеваний в день» — Максимальное значение скорости изменения (в день) величины «Заболеваний» (Рисунок 4, метка 5)			10,996	9 мая
6	«Перекрестье» — Скорость изменения (в день) числа заболеваний сравнялась со скоростью изменения (в день) числа тех, кто прекратил болеть (выздоровел + умер). Перелом в эпидемии. Далее число тех, кто еще болеет, будет всегда уменьшаться (Рисунок 4, метка 6)			—	21 мая

Таким образом, главные результаты дальнего прогноза, выполненного 21.05.2020, это:

- Прогнозы четырех чисел «Макс Заболело», «Макс Пока больны», «Макс Пока больны в день», «Макс Заболеваний в день» — дают общую оценку (по порядку величины) значений. Не следует здесь ожидать высокой точности.
- Прогнозы для шести дат (добавлена буква D от английского слова Day — День) ключевых событий в течение эпидемии: «D Макс Пока больны в день», «D Макс Заболеваний в день», «D Макс Пока больны», «D перекрестья», «D Макс Заболело», «D Все выздоровеют».

Использование фазового пространства (f, \dot{f}) для мониторинга и дальнего прогноза

Пусть $f(t)$ — функция, описывающая количество выявленных заболеваний в момент времени t , $\dot{f}(t)$ — первая производная по времени, скорость изменения в момент времени t функции $f(t)$. Будем на числовой плоскости xOy для разных моментов времени t наносить точки (x, y) , где $x = f(t)$, $y = \dot{f}(t)$ — получим параметрическую кривую, определяемую двумя этими уравнениями. Наша числовая плоскость xOy будет фазовым пространством (f, \dot{f}) , а параметрическая кривая — траекторией нашей функции, описывающей количество выявленных заболеваний.

Предположим, что эпидемия в точности описывается логистической функцией (1), которая удовлетворяет дифференциальному уравнению (2):

$$\dot{f} = \frac{k}{N_0} \cdot (2N_0 - f) \cdot f.$$

Траектория на фазовой плоскости отвечает соотношениям $x = f(t)$, $y = \dot{f}(t)$. Получаем следующее описание траектории:

$$y = \frac{k}{N_0} \cdot (2N_0 - x) \cdot x. \quad (4)$$

Это есть парабола! То есть, в предположении, что эпидемия в точности описывается логистической функцией (1), траектория эпидемии на фазовой плоскости (f, \dot{f}) , является параболой (4) ветвями вниз, с корнями $x = 0$ и $x = 2N_0$. Наконец, заметим, что $N = 2N_0$ — это максимальное значение логистической функции, то есть, максимальное (финальное) количество выявленных заболеваний.

Данные рассуждения дают нам новый подход к дальнейшему прогнозу максимального (финального) количества выявленных заболеваний.

- Нанесем на числовую плоскость xOy точки (V_i, \dot{V}_i) .
- Рассмотрим квадратный трехчлен $h(x) = D \cdot (N - x) \cdot x$, с неизвестными параметрами D и N , описывающий параболу с ветвями вниз и корнями 0 и N .
- Подберем при помощи МНК параметры D и N так, чтобы парабола тесно легла бы к нашим точкам (V_i, \dot{V}_i) . То есть, параметры D и N подбираются из условия минимизации величины $\sum_{i=1}^n (h(V_i) - \dot{V}_i)^2$, рассчитываемой по всей накопленной статистике $\{(V_i, \dot{V}_i) \mid i = 1 \dots n\}$.
- Найденное значение параметра N является еще одним дальним прогнозом максимального (финального) количества выявленных заболеваний. Будем его обозначать «**Макс Заболело** (f, f')».

В целом так все и реализовано в Excel-файле (Рисунок 5). Но, с учетом высоких погрешностей в статистике в средних скоростях за день \dot{V}_i , вместо них используются более сглаженные данные — средние скорости за неделю: $\hat{V}_i = \frac{V_i - V_{i-6}}{T_i - T_{i-6}}$.

Практика показала, что прогноз «**Макс Заболело** (f, f')» является, скорее всего, оценкой сверху истинной величины максимального (финального) количества выявленных заболеваний, в то время как прогноз «**Макс Заболело**» — оценкой снизу. Это удобно, иметь для анализа и верхнюю, и нижнюю оценку — Рисунок 7, стр. 14.

Анализ траекторий эпидемий на фазовой плоскости (f, \dot{f}) — популярный метод мониторинга процесса протекания эпидемий. С этой точки зрения отметим электронный ресурс 3, на котором представлены анимированные траектории прохождения эпидемий в разных странах на фазовой плоскости «число выявленных заболевших — скорость (число выявленных заболевших за последнюю неделю)». Если на этом ресурсе выбрать линейную шкалу (а не логарифмическую) и ту или иную страну, то можно увидеть траекторию прохождения эпидемии в этой стране.

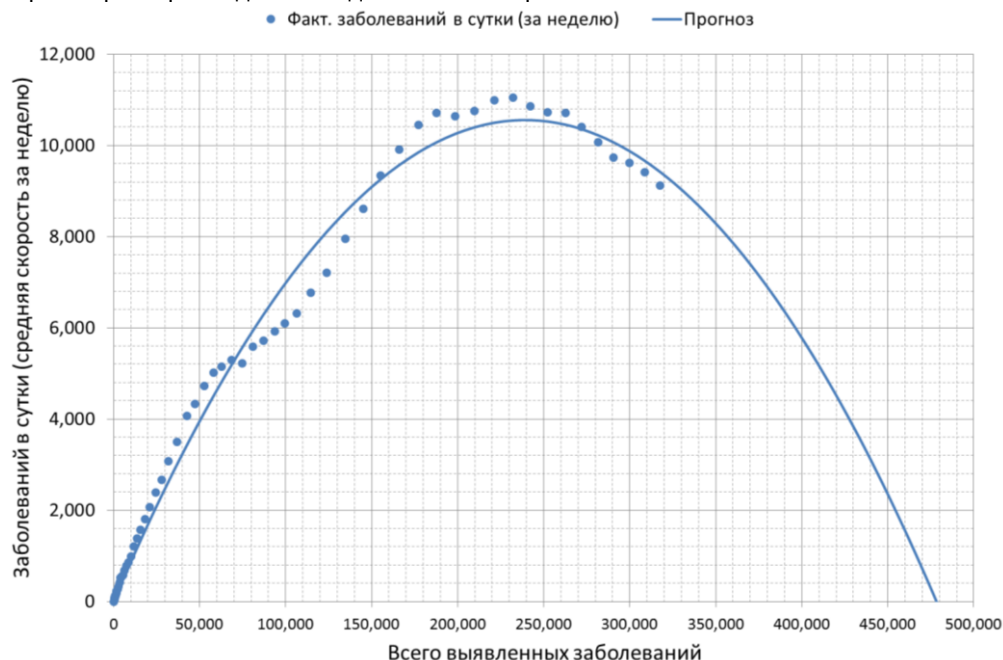


Рисунок 5. Траектория эпидемии на основе статистики 16.03.2020–21.05.2020. Круглые точки — значения (V_i, \dot{V}_i) из реальной статистики, непрерывные линии — аппроксимирующая парабола с ветвями вниз и корнями 0 и N .

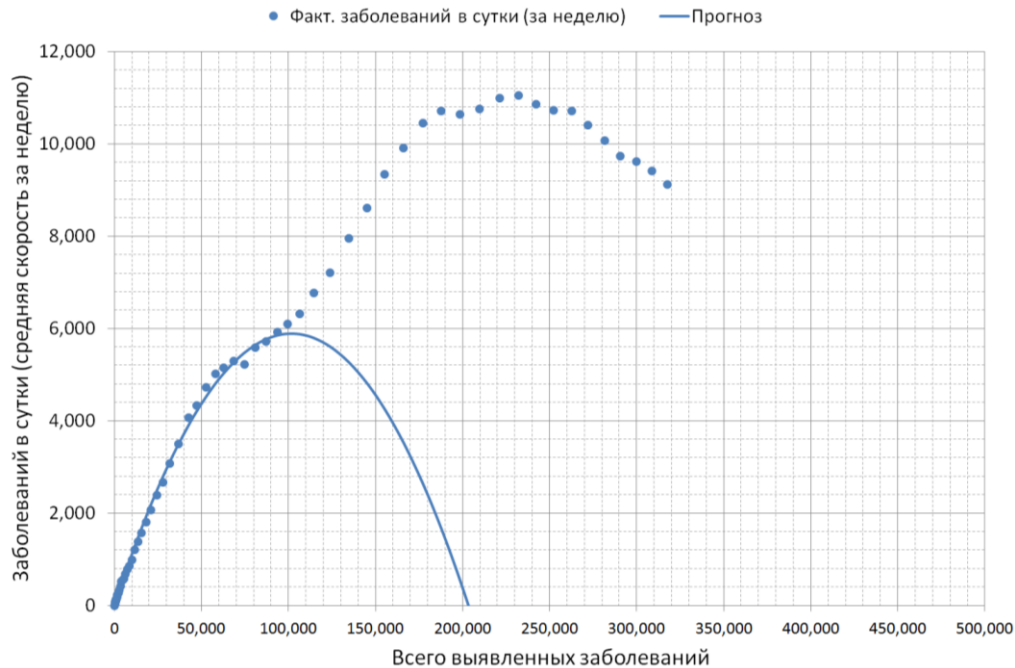


Рисунок 6. Вероятная упущенная возможность завершить эпидемию с $N \approx 203,000$ выявленных заболеваний. Траектория эпидемии на основе статистики 16.03.2020–27.04.2020. Круглые точки — значения (V_i, \hat{V}_i) из реальной статистики, непрерывные линии — аппроксимирующая парабола с ветвями вниз и корнями 0 и N .

Эффективные действия властей приводят к тому, что траектория максимально близка к параболе. В качестве таких примеров можно посмотреть на ресурсе 3 траекторию Германии, Австрии или Турции.

В неудачных случаях (Испания) можно распознать «упущенные возможности»: траектория вот уже переваливает вершину аппроксимирующей параболы и идет вниз, но ... ситуация упускается из-под контроля, и траектория уходит на другую, более широкую, параболу. Вполне вероятно (Рисунок 6), что и Россия в 20-х числах апреля потеряла контроль над эпидемией и упустила возможность завершить эпидемию с $N \approx 203,000$ выявленных заболеваний, а не с $N \approx 450,000$, как это предсказывают прогнозы сегодня (Рисунок 5).

Что такое «дрейф прогноза»?

Итак, мы провели 21.05.2020 дальний прогноз на массиве данных за отрезок времени 16.03.2020–21.05.2020 и получили 5 чисел и 6 дат, описывающих дальнейшее предвидение ключевых событий эпидемии.

Пройдет один день, и 22.05.2020 мы сможем повторить дальний прогноз на уже расширенном массиве данных — за отрезок времени 16.03.2020–22.05.2020. И мы, скорее всего, получим немного иные 5 чисел и немного иные 6 дат, описывающих новый дальний прогноз — новое предвидение ключевых событий эпидемии. Произойдет пересмотр — изменение, дрейф — этих 5 чисел и 6 дат. Будем называть это *дрейфом прогноза*. Так как новый прогноз будет опираться на расширенную статистику, то природа этого дрейфа — уточнение прогноза за счет расширения исходных данных.

Будем дальний прогноз *пересчитывать каждый день*, добавляя новую и новую статистику и получая каждый день дрейф прогноза — уточнение прогноза. Можно формально обосновать, что каждое из 5 чисел и каждая из 6 дат с течением времени *будет дрейфовать к своим окончательным и, в итоге, истинным значениям*. А это значит, что наблюдение за графиками (Рисунок 7, Рисунок 8, Рисунок 9) дрейфа прогноза позволяет с течением времени выдвигать обоснованные экспертные суждения о 5 численных значениях и о 6 датах ключевых событий в течении эпидемии.

Важное обстоятельство третьей модели. Аппроксимирующие функции $f(t)$ — формула (1),— и $g(t)$ — формула (3),— приближают реальную статистику V_i и E_i , которая *строго возрастает*. А это значит, что свежие данные по величине больше давних данных. И МНК автоматически будет свежие данные *учитывать с большим приоритетом*: при близких значениях абсолютной погрешности аппроксимации, *относительная погрешность аппроксимации свежих данных будет гораздо меньше, чем у старых данных*. Это одно из важных свойств, позволяющих обосновать, почему дальние прогнозы величин и дат в конце концов дрейфуют к своим окончательным и, в итоге, истинным значениям.

Еще одно важное замечание: видно, что иногда дрейф происходит *существенным образом и в неблагоприятную сторону* — это сигнал о потере контроля над эпидемией, произошедшей немного ранее. А вот существенный позитивный дрейф сигнализирует об удачных мерах борьбы с эпидемией.

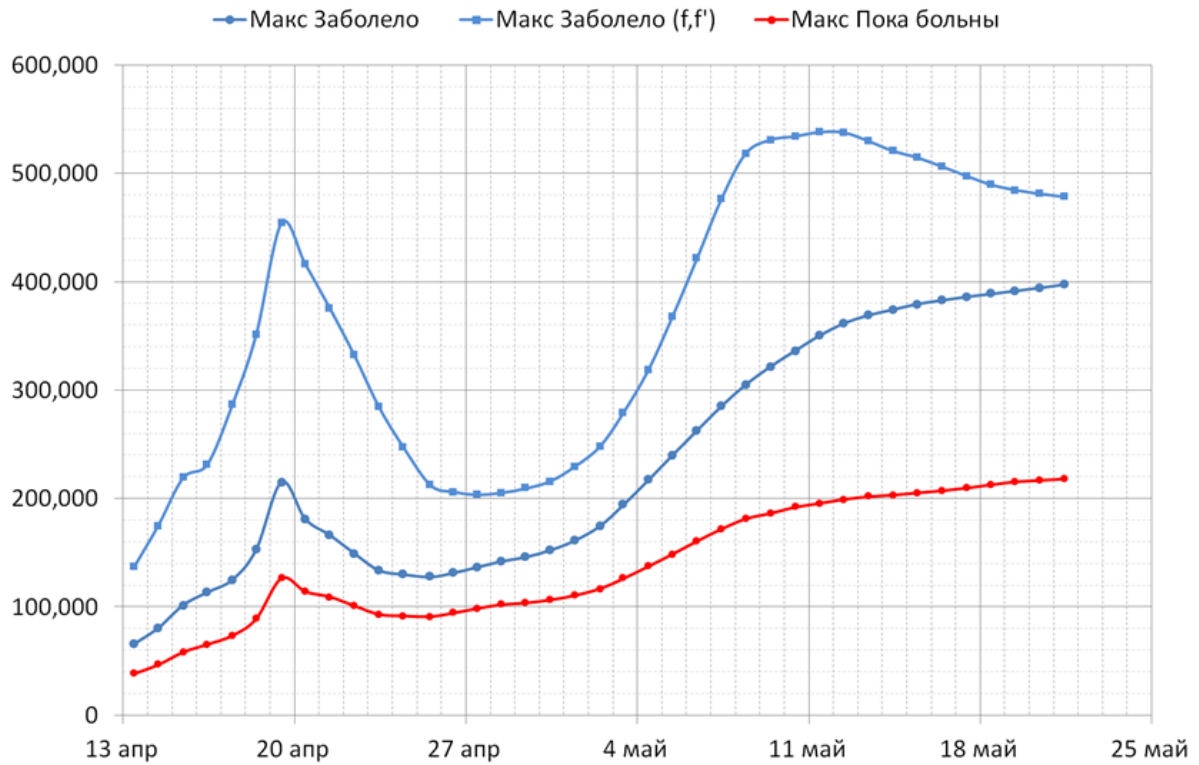


Рисунок 7. Дрейф прогноза величин «Макс Заболело» «Макс Заболело (f, f')» и «Макс Пока больны» (ось ординат — даты выполнения прогноза)

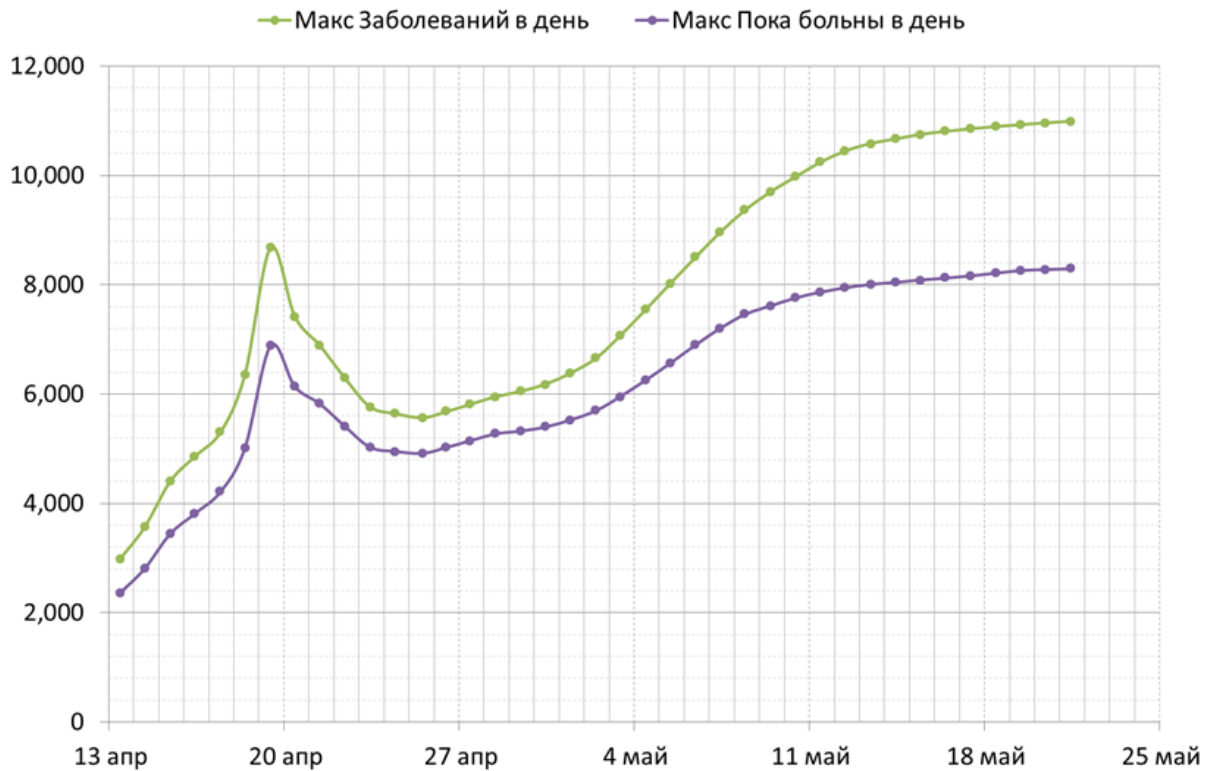


Рисунок 8. Дрейф прогноза величин «Макс Заболеваний в день» и «Макс Пока больны в день» (ось ординат — даты выполнения прогноза)

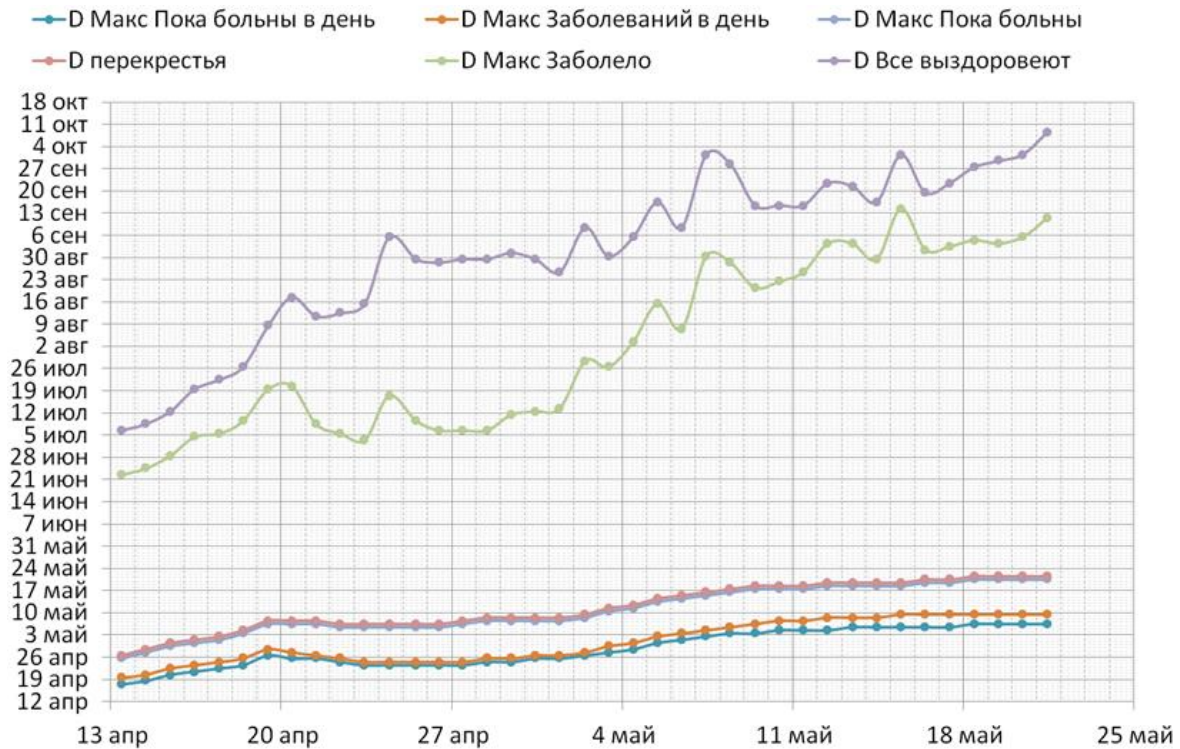


Рисунок 9. Дрейф прогноза дат наступления шести ключевых событий в протекании эпидемии (ось абсцисс — даты выполнения прогноза)

Как можно анализировать статистику, которая неточна, неполна и противоречива

Предложенная в данной работе методика позволяет опираться на анализ неполной и не совсем достоверной статистической информации и делать вполне надежные выводы. Как это, в принципе, возможно, поясним при помощи аналогии.

Наблюдая за айсбергом, мы всегда видим совсем малую его часть, большая часть не видна. Но, наблюдая только за видимой частью, мы можем *весьма достоверно* оценивать динамику процессов в масштабе всего айсберга. Например, можем достоверно оценивать растаявшую за отрезок времени долю айсберга, оценить скорость таяния, вычислить прогноз дат, когда растает 25%, 50%, 75% и 100% всего айсберга. Это сделать можно, оценки и прогнозы будут весьма достоверны.

Такие рассуждения проходят, если верно предположение, что видимая доля и невидимая доля наблюдаемого явления более-менее постоянны.

Так вот, для эпидемии коронавируса в России это предположение (постоянство долей «видно/не видно») более-менее верно *в начале процесса*. Пока система не перегружена (в первую очередь, система тестирования) и пока числа были *не очень страшные* с точки зрения критериев внутренней политики и отчетности регионов. А вот позже отношение долей «видно/не видно» начинает меняться и, вероятно, весьма сильно. И вот здесь помогает только одна методика, предложенная выше: *мониторинг дрейфа прогноза*.

Если постоянно следить за дрейфом прогноза, то это даст нам путеводную нить для понимания, в какой фазе находится процесс эпидемии и чего следует ждать впереди даже в условиях неточной, неполной и противоречивой статистики.

Моделирование и прогнозирование количества смертей в эпидемии коронавируса в России

Первоначальные идеи моделирования статистики смертей были просты. Пусть у некоторого человека выявлено заболевание covid-19. В официальной статистике это проявляется как еще одна единица в V_i . Кажется бы, должны быть predetermined (условные) вероятности:

- q — вероятность перехода этого человека (через несколько h дней) в состояние «выздоровел», и это проявляется в официальной статистике как плюс единица в H_{i+h} ;
- $p = 1 - q$ — вероятность перехода этого человека (через несколько d дней) в состояние «умер», и это проявляется в официальной статистике как единица в D_{i+d} .

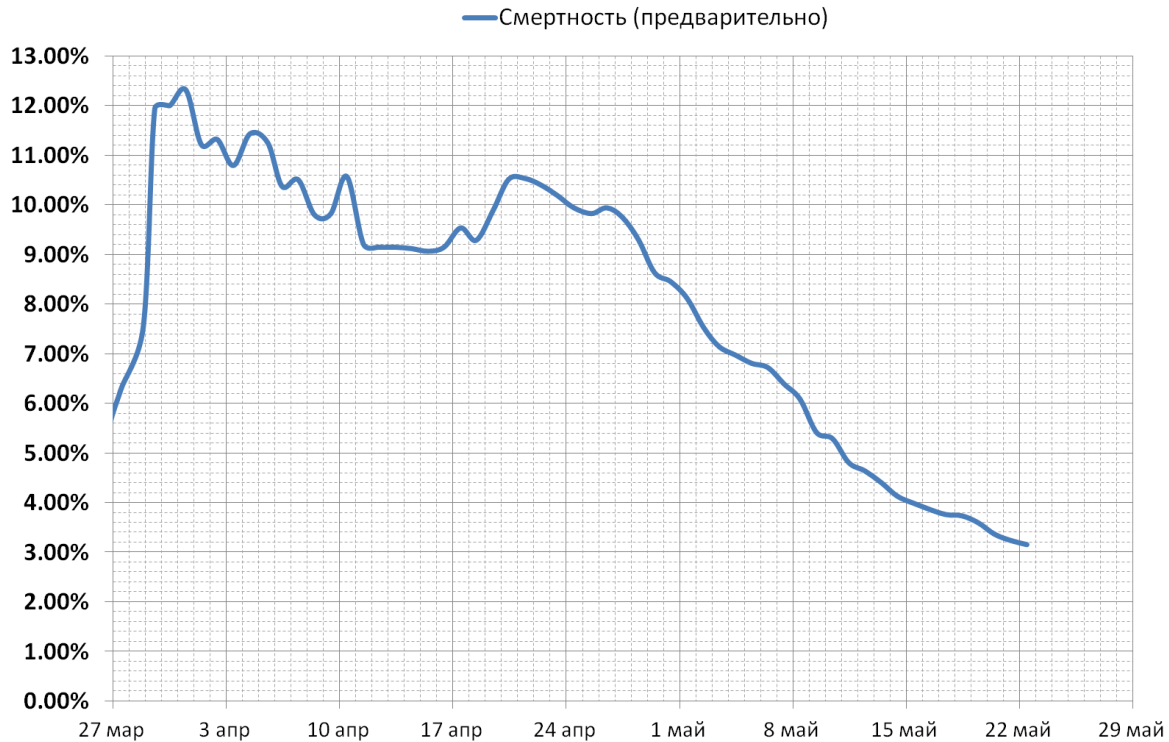


Рисунок 10. Динамика изменения величины p_i «смертность (предварительно)», статистика за период 16.03.2020–21.05.2020.

Моделирование и прогнозирование величины E_i обсуждено выше (напомним, это общее число зараженных, покинувших период болезни — в сторону выздоровления или смерти). Знание величин p и q полностью решало бы проблему моделирования и прогнозирования смертей и выздоровления: $D_i = p \cdot E_i$, $H_i = q \cdot E_i$. При этом вероятности p и q должны быть predeterminedены долгосрочными причинами, независимыми от текущих мероприятий: природой заболевания, средним уровнем здоровья и иммунной системы у населения, состоянием медицинской системы страны. Конечно, в разумных пределах вероятности p и q могут и меняться, например, в зависимости от уровня нагрузки на медицинскую систему, которая на первом этапе эпидемии растет: на этом этапе можно ожидать некоторого увеличения p и сопутствующего спада q .

Если бы регистрация выздоровлений и смертей в медицинской статистике была такой, как описано выше, то простое отношение $p_i = \frac{D_i}{E_i} = \frac{D_i}{D_i+H_i}$ должно было бы постепенно, все точнее и точнее, оценивать ту самую условную вероятность $p|CV19$ — вероятность умереть после выявления заболевания covid-19. Будем называть величину вероятности p_i — «смертность (предварительно)».

Простые соображения⁸ говорят за то, что после накопления некоторого количества (единицы тысяч) выздоровлений и смертей, p_i должно быть довольно стабильно или, возможно, немного расти. Однако, наблюдается (Рисунок 10) совершенно иная динамика изменения p_i . До 27.04.2020 p_i достаточно стабильно держалось в отрезке [9% ... 13%], и к этому дню была накоплена значительная статистика выздоровлений и смертей — 7,346 и 794. Но после 27.04.2020, как по команде (...или именно «по команде?»), p_i стала монотонно и стремительно уменьшаться.

С точки зрения логики, такого поведения быть не должно. Однако, после частной беседы со специалистом в медицинском статистическом учете (с Марией Скачек) и получения информации о том, как на самом деле устроен учет, выяснились неожиданные для авторов детали. Оказывается, сценарий, при котором состояние болезни можно покинуть либо в сторону выздоровления, либо в сторону смерти, не полон. А именно, если у некоторого человека выявлено заболевание covid-19, то, действительно, в официальной статистике это проявляется как еще одна единица в V_i . А вот дальше, с человеком, с точки зрения статистики, может произойти не две, а три возможности:

Случай «Н»: переход этого человека (через несколько h дней) в состояние «выздоровел от covid-19» — это проявляется в официальной статистике как единица в H_{i+h} ;

⁸ — На первом этапе развития эпидемии нагрузка на медицинскую систему будет нарастать и p скорее всего возрастет или, по крайней мере, — не будет убывать!

Случай «D»: переход этого человека (через несколько d дней) в состояние «умер, причем именно от covid-19» (в справке о смерти в качестве причины указано именно covid-19, а не какая-то иная причина) — это проявляется в официальной статистике как единица в D_{i+d} .

Случай «Sch»: исчезновение этого человека (через несколько s дней) из статистического учета событий эпидемии covid-19 в России. Можем считать, что это проявляется как единица в некотором счетчике Sch_{i+s} , и счетчик этот нам недоступен. И как-то достоверно оценить его из одних только данных официальной статистики представляется практически невозможным.

Совершенно различные события могут привести к дальнейшему исчезновению человека из статистического учета эпидемии covid-19 в России. Если с человеком имел место случай «Sch», то человек может быть как мертв, так и жив. Буква «Sch» взята нами из немецкой фамилии «Schrödinger». Вот только два возможных события (существуют и иные), относящиеся к случаю «Sch»:

Лечащий врач с человека *снял диагноз covid-19*. Возможны разные варианты: был установлен иной диагноз; выяснилось, что человек вовсе не был болен и др. Это не считается выздоровлением от covid-19.

Человек умер, но в справке о смерти написана иная причина, а не covid-19.

Также можно предположить, что, по мере увеличения количества проводимых тестов, должно нарастать количество людей, у которых тест дал положительный результат, но которые по тем или иным причинам не были госпитализированы, а остались на домашнем режиме, откуда с высокой вероятностью выходят именно в Шредингеровскую группу.

Таким образом, утверждение, будто « $V_i = E_i = H_i + D_i$ », оказалось не совсем точным. Если i соответствует окончанию эпидемии — в правой части не хватает слагаемого Sch_i . Но, мы надеемся, что по сравнению с величинами V_i и E_i неизвестное слагаемое Sch_i мало и серьезным образом на моделирование не влияет. Нелогичная динамика величины p_i (Рисунок 10) и полученная новая информация о статистическом учете требуют изменить подходы к регистрации статистики смертности. Но это выходит за рамки данной работы.

Заключение

Выводы. В данной работе изложены разработанные и примененные на практике методы анализа, моделирования и прогноза статистики эпидемии коронавируса в России. Эти методы позволили вычислять достаточно точные краткосрочные (недельные) прогнозы ожидаемых показателей эпидемии и дальние прогнозы для величин и дат наступления ключевых событий в эпидемии. Предложена, реализована и использована на практике методика мониторинга дрейфа дальних прогнозов, что позволяет ориентироваться в том, на какой фазе эпидемии сейчас находится процесс и в каких окончательных объемных и временных рамках следует ожидать завершения каждой фазы и эпидемии в целом.

Возможные дальнейшие работы. Работа с числовыми параметрами примененной для поздних этапов SIR-модели заставила нас обратить внимание на несколько моментов, обычно выпадающих из зоны внимания аналитиков. Упомянем только два из них.

Первый — это причина перехода из фазы экспоненциального роста в фазу насыщения. Начиная с классической работы 11, в которой впервые был использован гиперболический тангенс в качестве функции от времени, описывающей количество переболевших, основной причиной насыщения считается исчерпание ресурса неинфицированных особей, которые могли бы участвовать в процессе цепного распространения болезни.

На самом деле это не так, и в настоящее время приходится иногда выслушивать мнение о несовершенстве и даже непригодности SIR-модели к эпидемиям, которым противостоит современная медицина. Действительно, во всех странах, переживающих пандемию covid-19, затухание сводит заболеваемость на нет задолго до исчерпания субстрата для дальнейшего его распространения. В большинстве случаев количество переболевших не превышает одного процента населения страны. Предположение, будто остальные 99% переносят заболевание в латентной форме или, более того, обладают изначальным иммунитетом к инфекции, представляется избыточно искусственным.

Численным моделированием схемы SIR можно «нащупать» режимы, при которых затухание эпидемии будет достигнуто при малой доле переболевших. Для этого необходимо и достаточно, чтобы соотношение констант скорости (то есть коэффициентов в дифференциальных уравнениях модели) для каналов распространения инфекции и выздоровления было очень близко к единице, причем с небольшим перекосом в сторону распространения. Если изначально коэффициент передачи будет меньше единицы, болезнь затухнет, не успев распространиться, то есть эпидемии не будет. И только случай, когда на начальном этапе коэффициент распространения превышает единицу, но по мере выгорания субстрата уменьшается ниже нее, дает картину с сигмоидальной динамикой. Но при этом требуется уникально тонкая настройка коэффициентов модели. Даже незначительное отклонение приведет к тому, что заражение охватит всех.

Поскольку подобного в нынешней пандемии не произошло ни разу, аналитик просто обязан предположить какие-то иные механизмы устойчивости модели, помимо уникально узкого диапазона ее числовых параметров.

Вторая слабость SIR-модели напрямую связана с первой и тоже касается численных значений параметров дифференциальных уравнений. Даже на примере России мы видим, что динамика развития пандемии, например, в Урюпинске (55 случаев на дату подготовки статьи 23.05.2020) и в Москве (161

тысяча) отличаются только масштабным фактором. То есть с точностью до масштабирования и сдвига начала — совпадают. Это подразумевает, что степень «заразительности» в Москве должна быть в 300 раз меньше, в соответствии с пропорцией по населению (или в 3 000 раз по заболевшим). Ни одна модель такой вариативности коэффициентов не допускает. Поэтому напрашивается следующее дополнение к классической модели.

Реально зараженный индивидум не может общаться со всем населением страны, города и даже района. Параметр скорости заражения ограничен кругом его общения, который составляет в среднем 50–100 человек. Очевидно, что никакие меры (само)изоляции этот круг контактов затронуть практически не могут. И в пределах одного очага заражения инфекция распространяется практически бесконтрольно.

А вот контакты между разными группами общения организационными мерами пресечены быть могут. И скорость передачи инфекции от одного очага к другому — явление вполне подконтрольное. В настоящее время мы занимаемся разработкой такой двухуровневой модели.

Несмотря на ее очевидную понятность, она представляет серьезные математические сложности. В частности, от обыкновенных дифференциальных уравнений (ОДУ) необходимо переходить к уравнениям с запаздывающим аргументом. Это, казалось бы, незначительное усовершенствование дает далеко идущие последствия, например, появление второй волны эпидемии, о чем классическая модель молчит.

Также имеются идеи и предпосылки для вычисления правдоподобного распределения по дням среднего числа заражаемых людей каждым зараженным и распределения по дням (после установленного диагноза) вероятности смерти и вероятности выздоровления для зараженного.

Еще одно очевидное направление работ — применение разработанных методов к статистическим массивам различных стран, в том числе тех, где эпидемия коронавируса уже закончилась или близка к этому. За счет таких исследований можно накопить больше данных, например, о фактической относительной точности краткосрочных прогнозов на базе третьей модели (1).

Благодарности. Авторы благодарны Марии Скачек, которая, будучи специалистом, рассказала некоторые особенности сбора и учета медицинской статистики, по поводу которых авторы никогда бы не смогли предположить, что такое может быть в реальности. На работу сильно повлияли плодотворные обсуждения с Андреем и Аркадием Климовыми, которые, среди прочего, познакомили авторов с сервисом 3, обратили внимание на пользу анализа траекторий на фазовой плоскости (f, \dot{f}) и на то, что траектория идеальной SI-модели эпидемии будет параболой с корнями 0 и N , где N — максимальное (финальное) число заболевших в эпидемии.

Список литературы

1. Леоненко В.Н. **Математическая эпидемиология** // Учебно-методическое пособие по выполнению лабораторных работ. Учебно-методическое пособие. — СПб: Университет ИТМО, 2018.
2. *Herbert W. Hethcote* **Three Basic Epidemiological Models** // S. A. Levin et al. (eds.), *Applied Mathematical Ecology*, © Springer-Verlag Berlin Heidelberg 1989.
3. *M.G. Garner & S.A. Hamilton* **Principles of epidemiological modelling** // *Rev. sci. tech. Off. int. Epiz.*, 2011, 30 (2), 407–416.
4. *Min-Shi Lee & D. James Nokes* **Predicting and comparing long-term measles antibody profiles of different immunization policies** // *Bulletin of the World Health Organization*, 2001 — SciELO Public Health.
5. *Haposan, Jonathan* **Developing a method for characterizing genealogy of parameter values from epidemiological models of infectious disease** // Master Essay, University of Pittsburgh, 2019, URL: <http://d-scholarship.pitt.edu/36731/> (доступ проверен 21.05.2020).
6. *J Lennert Veerman, Jan J Barendregt, Johan P Mackenbach and Johannes Brug* **Using epidemiological models to estimate the health effects of diet behaviour change: the example of tailored fruit and vegetable promotion** // *Public Health Nutrition*: 9(4), 415–420 DOI: 10.1079/PHN2005873.
7. *Абрамов С.М., Травин С.О.* **Вопросы и ответы по мониторингу и прогнозу эпидемии коронавируса в России** // Рукопись, электронный ресурс, 2020, URL: <https://tinyurl.com/y7bplslc> (доступ проверен 21.05.2020).
8. *Абрамов С.М., Травин С.О.* **Электронная таблица мониторинга, анализа и прогноза эпидемии коронавируса в России** // Рукопись, электронный ресурс, 2020, URL: <https://tinyurl.com/yck5gucy> (доступ проверен 21.05.2020).
9. **Коронавирус COVID-19: Официальная информация о коронавирусе в России** (сайт «стопкоронавирус.рф») // Электронный ресурс в сети Интернет, URL: <https://стопкоронавирус.рф/#> (доступ проверен 21.05.2020).
10. **Trajectory of World COVID-19 Confirmed Cases** // Электронный ресурс в сети Интернет, URL: <https://aatishb.com/covidtrends> (доступ проверен 21.05.2020).
11. *Kermack, W. O., McKendrick, A. G.* **A Contribution to the Mathematical Theory of Epidemics** // *Proceedings of the Royal Society*, 1927. Vol. 115, No. A771, P.700-721.

References in Cyrillics

1. Leonenko V.N. Matematicheskaya e`pidemiologiya // Uchebno-metodicheskoe posobie po vy`polneniyu laboratorny`x rabot. Uchebno-metodicheskoe posobie. — SPb: Universitet ITMO, 2018.
2. Herbert W. Nethcote Three Basic Epidemiological Models // S. A. Levin et al. (eds.), Applied Mathematical Ecology, © Springer–Verlag Berlin Heidelberg 1989.
3. M.G. Garner & S.A. Hamilton Principles of epidemiological modelling // Rev. sci. tech. Off. int. Epiz., 2011, 30 (2), 407–416.
4. Min-Shi Lee & D. James Nokes Predicting and comparing long-term measles antibody profiles of different immunization policies // Bulletin of the World Health Organization, 2001 — SciELO Public Health.
5. Haposan, Jonathan Developing a method for characterizing genealogy of parameter values from epidemiological models of infectious disease // Master Essay, University of Pittsburgh, 2019, URL: <http://d-scholarship.pitt.edu/36731/> (dostup proveren 21.05.2020).
6. J Lennert Veerman, Jan J Barendregt, Johan P Mackenbach and Johannes Brug Using epidemiological models to estimate the health effects of diet behaviour change: the example of tailored fruit and vegetable promotion // Public Health Nutrition: 9(4), 415–420 DOI: 10.1079/PHN2005873.
7. Abramov S.M., Travin S.O. Voprosy` i otvety` po monitoringu i prognozu e`pidemii ko-ronavirusa v Rossii // Rukopis`, e`lektronny`j resurs, 2020, URL: <https://tinyurl.com/y7bplslc> (dostup proveren 21.05.2020).
8. Abramov S.M., Travin S.O. E`lektronnaya tablicza monitoringa, analiza i prognoza e`pi-demii koronavirusa v Rossii // Rukopis`, e`lektronny`j resurs, 2020, URL: <https://tinyurl.com/yck5gucy> (dostup proveren 21.05.2020).
9. Koronavirus COVID-19: Oficial`naya informaciya o koronavirusе v Rossii (sajt «stopkoronavirus.rf») // E`lektronny`j resurs v seti Internet, URL: <https://stopkoronavirus.rf/#> (dostup proveren 21.05.2020).
10. Trajectory of World COVID-19 Confirmed Cases // E`lektronny`j resurs v seti Internet, URL: <https://aatishb.com/covidtrends> (dostup proveren 21.05.2020).
11. Kermack, W. O.; McKendrick, A. G. A Contribution to the Mathematical Theory of Epidemics // Proceedings of the Royal Society, 1927. Vol. 115, No. A771, P.700-721.

*Абрамов Сергей Михайлович — ORCID: 0000-0001-6603-7971,
доктор физико-математических наук, член-корреспондент РАН,
Институт программных систем им. А.К. Айламазяна Российской академии наук.*

*Травин Сергей Олегович — ORCID: 0000-0003-2470-7855, кандидат химических наук,
профессор, Федеральный Исследовательский центр химической физики им. Н.Н. Семенова
Российской академии наук.*

Ключевые слова

Ключевые слова: коронавирус, моделирование эпидемии, прогнозирование развития эпидемии

Sergei Abramov, Sergei Travin, Modeling and Forecast of the Coronavirus Epidemic Statistics in Russia

Keywords

coronavirus, epidemic modeling, epidemic forecasting.

DOI: 10.34706/DE-2020-02-01

JEL classification C53 Forecasting and Prediction Methods • Simulation Methods, C15 Statistical Simulation Methods: General

Abstract

It became obvious in March 2020, that the coronavirus epidemic in Russia would be quite significant in scope and duration. To draw up personal plans, or plans of small institutions, it was necessary to answer somehow next simple questions at every moment of time: What awaits us in the near future? What stage of the epidemic we are at? When will this be over? In fact, it became necessary to make judgments about the dynamics of the epidemic, considering the incoming official statistics about the epidemic, modeling and predicting the behavior of these statistics. This paper is devoted to some methods and results of such analysis, modeling and forecasting.