

Применение математических моделей динамики заболеваемости эпидемическими ОРВИ для увеличения эффективности эпидемиологического надзора

Леоненко В.Н.*^{1,2}, Корзин А.И.¹, Даниленко Д.М.²

¹ *Университет ИТМО, Санкт-Петербург, Россия*

² *ФГБУ «НИИ гриппа им. А.А. Смородиной» МЗ РФ, Санкт-Петербург, Россия*

Аннотация. Неопределённость в расчётах прогнозов распространения эпидемических острых респираторных инфекций, получаемых с помощью математических моделей, связанная с погрешностью данных и неопределённостью выбора модели, а также отсутствие верификации результатов моделирования междисциплинарными коллективами, включающими специалистов по эпидемиологии, препятствуют корректному прогнозу эффективности противоэпидемических мероприятий. В настоящей работе предлагается вариант решения указанных проблем путём использования программного комплекса, состоящий из семейства эпидемических моделей, методов оценки погрешности выходных данных в зависимости от погрешности исходных данных заболеваемости, а также графического интерфейса с возможностью ручной коррекции результатов автоматической калибровки и генерации эпидемических бюллетеней. Новизной представленного исследования является методика интеграции эпидемических моделей в программные средства, используемые надзорными органами, что позволяет дополнять еженедельные бюллетени и ежегодные эпидемиологические отчёты в полуавтоматическом режиме с количественной интервальной оценкой погрешности рассчитываемых показателей. Конечной целью является предоставление органам надзора информативных и оперативно получаемых расчётных данных для принятия решений в сфере контроля распространения инфекций.

Ключевые слова: *эпидемиология, прогнозы заболеваемости, ОРВИ, модель Барояна-Рвачева, модель Кермака-Маккендрика, программный комплекс.*

ВВЕДЕНИЕ

1. Постановка проблемы

Принятие решений по сдерживанию и контролю заболеваемости в условиях эпидемий – сложная задача, требующая максимально точной и выверенной количественной информации. Для оценки ключевых эпидемических показателей и краткосрочного прогнозирования течения эпидемии надзорными органами осуществляется сбор данных и их анализ с помощью различных математических методов. Характерным в современной практике специалистов–эпидемиологов, отвечающих за подготовку отчётов по динамике

*vnleonenko@yandex.ru

эпидемических процессов, является использование для ретроспективного анализа простых и апробированных статистических методов, в то время как исследователи с изначальной специализацией в прикладной математике и программировании зачастую предпочитают использование объясняющих математических моделей различной степени сложности. Модельный подход по сравнению с чисто статистическим имеет ряд заметных преимуществ. В частности, с помощью моделей становится возможным более точно прогнозировать динамику заболеваемости, а также рассчитывать значения эпидемических индикаторов, таких как R_0 и R_t , и производить отработку сценариев типа 'что-если' (what-if scenarios), оценивая, например, эффективность карантинных мер и вакцинации. С другой стороны, недостаточно обоснованное использование моделей в рамках эпидемиологической практики несёт ряд рисков. Среди них можно выделить следующие:

- Погрешности в исходных данных, использованных для калибровки моделей, а также неэффективная работа калибровочного алгоритма (например, попадание в локальные оптимумы), могут приводить к соответствующим погрешностям в результатах модельных оценок. Во многих модельных работах это не берётся в расчёт и в качестве выходных данных приводятся точечные оценки эпидемических показателей без учёта влияния погрешностей вычислений.
- Вариация структуры модели и степени ее детализации может приводить к значительному изменению количественных характеристик эпидемических сценариев. Как следствие, результаты прогнозирования и расчёта эффективности противоэпидемических мероприятий зависят от выбора конкретной модельной структуры, что также зачастую не обсуждается исследователями.

Как следствие, неопределённость в расчётах прогнозов и эпидемических индикаторов, вычисляемых с помощью модельного подхода, может быть велика, и отсутствие её учёта, характерное для большого количества работ по моделированию распространения эпидемических ОРВИ, затрудняет корректный прогноз эффективности противоэпидемических мероприятий. Кроме того, проведение эпидемиологического анализа специалистами по моделированию на основе непроверенных моделей и без консультаций с эпидемиологами повышает недоверие к математическому моделированию в эпидемиологии в целом и в долгосрочном периоде усложняет проведение междисциплинарных исследований в этой области. Наконец, критерии оптимальности калибровки модели и выбора наиболее эффективных мер сдерживания, основанные только на количественных измерениях, могут быть ограничены с точки зрения эпидемиологов и лиц, принимающих решения, поскольку не берут в расчёт знания экспертов предметной области.

В настоящей работе предлагается программный инструмент, состоящий из семейства эпидемических моделей, методов оценки погрешности выходных данных в зависимости от погрешности исходных данных заболеваемости, а также графического интерфейса с возможностью ручной коррекции результатов автоматической калибровки и генерации эпидемических бюллетеней. Новизной представленного исследования является методика интеграции эпидемических моделей в программные средства, используемые надзорными органами, что позволяет дополнять еженедельные бюллетени и ежегодные эпидемиологические отчёты в полуавтоматическом режиме с количественной интервальной оценкой погрешности рассчитываемых показателей. Конечной целью разработки указанного инструмента является эффективное использование разрабатываемых авторами методов и алгоритмов [1], [2] в практике

эпидемиологического надзора и, как результат, предоставление органам надзора информативных и оперативно получаемых расчётных данных для принятия решений в сфере контроля распространения инфекций.

2. Существующие решения

На протяжении большого исторического периода основную часть работ по моделированию эпидемических ОРВИ составляли работы по моделированию динамики распространения гриппа – в особенности, его пандемических штаммов. С начала 2020 года, в связи с начавшейся пандемией COVID-19 и сопутствующим падением заболеваемости гриппом, наибольшую актуальность стала иметь проблема прогнозирования и контроля распространения коронавирусной инфекции. Подавляющее большинство работ по оценке динамики гриппа и COVID-19 базируется на двух основных методологиях:

- Применение классических популяционных SIR-моделей, в том числе с переменными коэффициентами, отражающими изменение особенностей протекания болезни в связи с принятыми контрольными мерами, например, [3], [4], [5], [6].
- Мультиагентное моделирование эпидемий гриппа и COVID-19 в отдельных районах (город, округ) и на ограниченных территориях (например, отдельные предприятия, круизные лайнеры, а также медицинские учреждения) [7], [8], [9], [10].

В связи с развитием технологий обработки больших данных популярность приобрела оценка эффективности противоэпидемических мероприятий в реальном времени на основе машинного обучения [11]. Этот подход был применён для прогнозирования COVID-19, например, см. [12].

Среди наиболее популярных подходов к применению инструментов анализа динамики инфекции, можно привести следующие два, наиболее близкие к представляемому в работе подходу:

- Средства поддержки принятия решений, основанные на моделях распространения инфекционных заболеваний. Модели калибруются на реальные данные, результаты моделирования (оценки эпидемических параметров, прогнозы и результаты проверки сценариев «что-если») предлагаются в качестве готовых результатов без возможности их оперативной верификации и пересчёта. Для некоторых подобных моделей исследователями-разработчиками предлагается графический интерфейс, позволяющий запускать автоматическую калибровку моделей на данные с заданием ограниченного ряда параметров (регион, численность населения, пр.) и визуализацией результатов [13], [14].
- «Эпидемические калькуляторы». Предлагают возможность изменять значения параметров эпидемических моделей посредством графического интерфейса и строить таким образом прогнозные кривые. При этом реальные данные заболеваемости либо не используются вовсе [15], либо приводятся исключительно для визуальной оценки результатов ручной калибровки [16].

Предлагаемый авторами программный комплекс позволяет одновременно получать значения параметров на основе эпидемических данных путем автоматической калибровки (аналогично системам поддержки принятия решений) и изменять эти значения в интерактивном режиме (аналогично эпидемическим калькуляторам). Важной

особенностью комплекса является возможность его интеграции в систему надзора, что в перспективе позволит производить оперативное уточнение прогнозов с помощью усвоения данных и генерацию отчётов для надзорных органов в реальном времени.

2. МЕТОДЫ

1. Структура программного комплекса

Программный комплекс состоит из следующих модулей:

- Модуль загрузки и преобразования данных.
- Семейство моделей динамики инфекционных заболеваний с различной степенью детализации.
- Модуль автоматической калибровки моделей на данные заболеваемости.
- Модуль расчёта неопределённости эпидемических показателей (прогнозов заболеваемости и значений эпидемических индикаторов) на основе неопределённости исходных данных.
- Графический интерфейс для ручной калибровки моделей и визуализации выходных данных.
- Модуль генерации отчётных материалов на основе результатов моделирования.

Модульная структура обеспечивает гибкость использования инструмента, поскольку позволяет изменять отдельные компоненты без ущерба для функциональности.

Последовательность операций для работы с комплексом с точки зрения пользователя приведена на рис. 1.

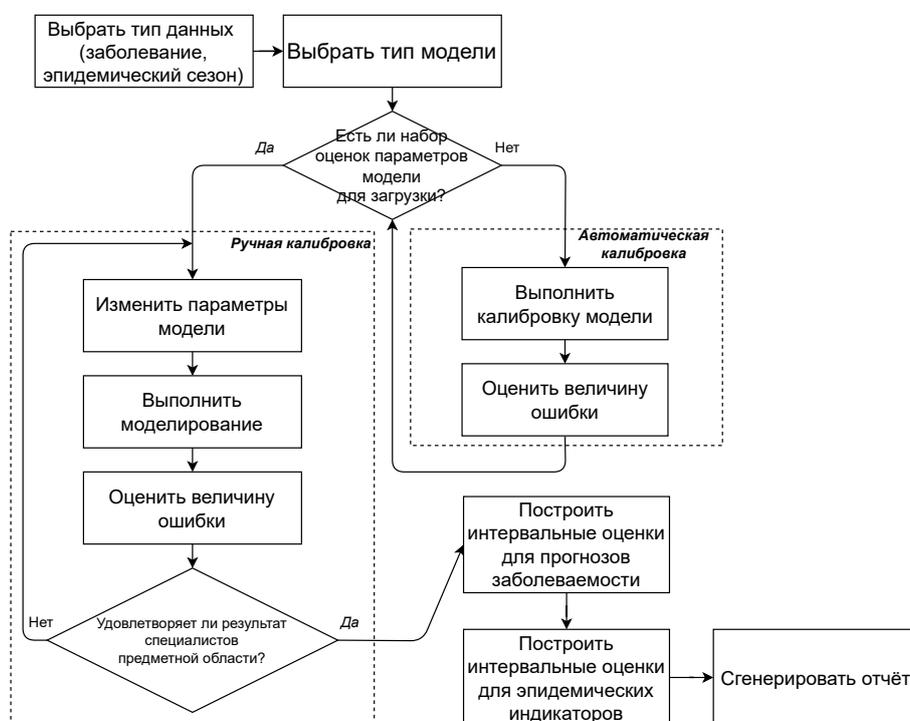


Рис. 1. Алгоритм операций при использовании программного комплекса

Далее каждая из компонент будет описана более подробно.

2. Данные

Для тестирования программной реализации модельного комплекса использовались следующие данные.

2.1. Грипп

В модельном комплексе в качестве исходных используются данные, предоставляемые НИИ гриппа, а именно:

- Данные о еженедельно регистрируемых клинически регистрируемых случаях ОРВИ и гриппа по отдельным возрастным группам (0–2, 3–6, 7–14 и 15+).
- Данные лабораторных анализов по распределению случаев ОРВИ, отнесенных к различным штаммам гриппа (A(H1N1)pdm09, A(H3N2), B).

На основе реализованного алгоритма сопряжения неоднородных данных по заболеваемости [1] модельный комплекс генерирует данные суммарной еженедельной заболеваемости с разбивкой по отдельным штаммам и возрастным группам. Возрастные группы от 0 до 14 лет могут быть агрегированы и использованы в исследовании как одна группа. Результирующее деление на две группы (0-14 и 15+ лет) позволяет добиться компромисса между детальностью получаемых модельных результатов (возрастное деление необходимо для учёта разницы в заболеваемости детей и взрослых) и уровнем неопределенности в оценке параметров моделей (деление на большое количество возрастных групп вызывает падение качества калибровки моделей на данные).

2.2. COVID-19

Данные для тестовых запусков модельного комплекса были взяты из репозитория [17], который содержит ежедневную динамику COVID-19 в Санкт-Петербурге. Рассматриваемая подвыборка содержит данные о заболеваемости за период со 02 марта 2020 года по 15 декабря 2022 года. Для анализа были сформированы следующие наборы данных:

- Ежедневная заболеваемость, т.е. количество активных случаев — взято непосредственно из исходных данных;
- Общее совокупное число зарегистрированных случаев — рассчитано по исходным данным.

Весь набор данных, соответствующий нескольким волнам эпидемии, был разделен на шесть отдельных вспышек COVID-19 согласно информации, доступной из открытых источников.

После внедрения комплекса в практику работы НИИ гриппа предполагается загружать данные заболеваемости непосредственно из базы данных, что позволит ускорить и упростить процесс их усвоения в модели.

3. Математические модели

Для максимально объективного расчёта эффективности контрольных мер модельный комплекс включает в себя возможность использования различных моделей применительно к одним и тем же данным. Пример сравнения авторами настоящей работы разных моделей на основе индикатора AIC и выбора модели с оптимальной структурной сложностью приведён в работах [18], [1]. Необходимо отметить, что модели не являются полностью взаимозаменяемыми, поскольку некоторые важные показатели эпидемии,

которые могут быть посчитаны одним типом модели, недоступны в модели другого типа. Например, логистические модели не могут предоставить значение базового числа репродукции R_0 [19], поскольку процесс выздоровления в таких моделях не учитывается в явном виде. Этот фактор может быть решающим для эпидемиологов при выборе модели для исследований, поскольку параметр R_0 и сходный динамический показатель R_t являются наиболее информативными при изучении эпидемических вспышек.

Ниже приведён набор моделей, которые включены в состав комплекса и могут быть использованы для моделирования гриппа и COVID-19. Приведенные модели численно реализованы авторами настоящего исследования и верифицированы на реальных данных с подтверждением их адекватности реальным эпидемическим процессам.

3.1. Дискретная модель Барояна-Рвачева

Модель Барояна-Рвачева [20] представлена системой разностных уравнений с шагом по времени, равным одному дню. Пусть x_t — доля восприимчивых в популяции, y_t — число вновь инфицированных в момент времени t , \bar{y}_t — общее число инфекционных индивидов к моменту времени t . Тогда система уравнений записывается в следующем виде:

$$\bar{y}_t = \sum_{\tau=0}^T y_{t-\tau} g_\tau, \tag{1}$$

$$y_{t+1} = \frac{\beta}{\rho} x_t \bar{y}_t,$$

$$x_{t+1} = x_t - y_{t+1},$$

$$x_0 = \alpha \rho. \tag{2}$$

Кусочно-постоянная функция g_τ отражает долю заразных особей в группе лиц, инфицированных за τ дней до текущего момента t . Функция отражает изменение индивидуальной заразности с течением времени от момента заражения гриппом. Предполагается, что существует некоторый момент \bar{t} : $\forall t \geq \bar{t} \ g_\tau = 0$, который отражает момент выздоровления.

Номенклатура модельных и калибровочных параметров приведена в Таблице 1.

3.2. Модель Барояна-Рвачева с возрастной структурой

Модель получена авторами из [20] разделением популяции на l возрастных групп. Пусть $x_t^{(j)}$ — доля восприимчивых возраста j , $y_t^{(j)}$ — число вновь инфицированных возраста j в момент времени t , $\bar{y}_t^{(j)}$ — общее число инфекционных индивидов возраста j к моменту времени t . Тогда система уравнений записывается в следующем виде [21]:

$$\bar{y}_t^{(j)} = \sum_{\tau=0}^T y_{t-\tau}^{(j)} g_\tau, j \in \overline{0, l}, \tag{3}$$

$$y_{t+1}^{(j)} = \sum_i \frac{\beta_{ij}}{\rho} x_t^{(j)} \bar{y}_t^{(i)}, i, j \in \overline{0, l}$$

$$x_{t+1}^{(j)} = x_t^{(j)} - y_{t+1}^{(j)}, j \in \overline{0, l}$$

$$x_0^{(j)} = \alpha_j \rho_j, j \in \overline{0, l}. \tag{4}$$

Таблица 1. Параметры калибровочного алгоритма для модели Барояна-Рвачёва

Переменная	Описание	Значение	Единица измерения
Модельные параметры			
α	Начальная доля восприимчивых индивидов в популяции	Варьирует	-
β	Интенсивность эффективных контактов	Варьирует	-
I_0	Начальная доля инфицированных в популяции	Варьирует	-
T	Отрезок моделирования	Фиксировано	День
g_τ	Доля заразных среди заразившихся τ дней до текущего момента	Фиксировано	-
ρ	Размер популяции	Фиксировано	Чел.
Параметры для калибровочного алгоритма ретроспективного анализа			
Δ_p	Абсолютное горизонтальное смещение положения пика смоделированной кривой по сравнению с данными	Варьирует	День

Интенсивность эффективных контактов между возрастными группами записывается как

$$\beta_{ij} = \lambda \cdot \delta_{ij}$$

где λ — вирулентность текущего штамма гриппа, а δ_{ij} — ежедневное количество контактов между возрастными группами i и j , $i, j \in \overline{1, \dots}$. Значения δ_{ij} составляют контактную матрицу $C = C(i, j)$.

3.3. Непрерывная SEIR–модель распространения гриппа

Непрерывная SEIR–модель распространения гриппа представляет собой стандартную дифференциальную модель типа Кермака-Маккендрика [22]. Население рассматриваемой городской территории представлено совокупностью четырех групп лиц: восприимчивые индивиды (уязвимые к гриппозной инфекции), индивиды в инкубационном периоде (бессимптомные и неинфекционные), инфекционные и выздоровевшие (не восприимчивые к гриппу). Размеры групп измеряются в долях общей популяции N : пусть S будет долей восприимчивых индивидов, E — долей индивидов в инкубационном периоде, I — долей инфекционных индивидов, и R — доля выздоровевших индивидов. Определенная часть населения в каждом рассматриваемом городе не подвержена гриппу. В эту подгруппу входят люди с иммунитетом, полученным от предыдущих инфекций, и те, кто не имеет иммунитета сам по себе, но защищен коллективным иммунитетом. Доля населения, уязвимая к циркулирующему в настоящее время штамму вируса гриппа, обозначается $\alpha \in [0; 1]$, а оставшаяся часть населения считается невосприимчивой к инфекции. Динамика размеров групп во времени задается следующими уравнениями:

$$\begin{aligned} \frac{dS}{dt} &= -\beta SI, \\ \frac{dE}{dt} &= \beta SI - \gamma E, \\ \frac{dI}{dt} &= \gamma E - \delta I, \\ \frac{dR}{dt} &= \delta I, \end{aligned} \tag{5}$$

$$\begin{aligned} S(t_0) = S_0 \geq 0, E(t_0) = E_0 \geq 0, I(t_0) = I_0 \geq 0, \\ S_0 + E_0 + I_0 = \alpha, \\ R(t_0) = 1 - \alpha. \end{aligned} \tag{6}$$

Коэффициенты интенсивности переходов между группами β , γ и δ считаются неотрицательными. Поскольку продолжительность эпидемического процесса относительно невелика, влияние процессов рождения и миграции на динамику заболевания считается пренебрежимо малым и не учитывается в модели. Номенклатура модельных и калибровочных параметров приведена в Таблице 2.

Таблица 2. Параметры калибровочного алгоритма для непрерывной SEIR-модели

Переменная	Описание	Значение	Единица измерения
Эпидемиологические параметры			
α	Исходное соотношение восприимчивых индивидов в популяции	Расчетное	-
β	Интенсивность заражения	Расчетное	1/(чел. · день)
γ	Интенсивность перехода в инфекционную форму болезни	0,39	1/сут
δ	Интенсивность выздоровления	0,133	1/день
I_0	Начальное соотношение зараженных	0,0001	-
Параметры позиционирования кривой			
k_{inc}	Относительное смещение смоделированной заболеваемости	вертикальное положение кривой	Расчетное -
Δ_s	Абсолютное смещение кривой относительно начала эпидемии	горизонтальное смоделированной заболеваемости начального дня по сравнению с данными	Оценка день

3.4. Мультиштаммовая модель Барояна-Рвачева с учётом предыстории заболеваемости

Мультиштаммовая компартментальная модель динамики гриппа основана на детерминированной системе разностных уравнений с шагом по времени, равным одному дню. Разработана авторами на основе стандартной модели Барояна-Рвачева [20] для моделирования многолетней динамики циркуляции вирусов гриппа с учётом формируемого иммунитета [23]. Подобные модели на основе стандартной дифференциальной SEIR-модели Кермака-Маккендрика приводятся, в частности в работе [24]. В этой версии модели мы рассматриваем совместную циркуляцию трех штаммов гриппа, A(H1N1)pdm09, A(H3N2) и B, поэтому мы принимаем $n_s = 3$, где n_s — общее количество рассматриваемых штаммов. Различные штаммы гриппа типа B не различаются, и в течение каждого эпидемического сезона рассматривается доминирующий штамм типа B. Пусть $x_t^{(h)}$ — доля восприимчивых лиц в популяции с историей воздействия $h \in \overline{1, n_s + 1}$, $y_t^{(m)}$ — количество лиц, вновь инфицированных в момент t штаммом вируса m и $\bar{y}_t^{(m)}$ — совокупное число заразившихся к моменту t передачи вируса деформации m , $m \in \overline{1, n_s}$. Возможность коинфицирования несколькими штаммами в течение одного сезона не рассматривается, поэтому лица, переболевшие гриппом, вызванным любым из циркулирующих штаммов, считаются невосприимчивыми. Однако этот предполагаемый перекрестный иммунитет между штаммами вируса не переносится на следующий эпидемический сезон.

Восприимчивые люди разделены на подгруппы в зависимости от их истории воздействия h , $h \in \overline{1, n_s + 1}$. Группу восприимчивых лиц со статусом экспозиции $h \in \overline{1, n_s}$ составляют лица, подвергшиеся заражению штаммом m в предыдущем эпидемическом сезоне, а группа с историей экспозиции $h = n_s + 1$ считается невосприимчивой к инфекции, вызванной любым штаммом. Переменная $\mu \in [0; 1)$ отражает долю населения, не участвующего в передаче инфекции. Мы делаем упрощающее допущение, что на уровень иммунитета людей влияет исключительно предшествующий эпидемический сезон, тем самым пренебрегая эффектами более ранней заболеваемости гриппом. Из-за ослабления иммунитета лица, имевшие в анамнезе контакт с некоторым штаммом гриппа в предыдущем сезоне, могут потерять иммунитет к тому же штамму в следующем эпидемическом сезоне. Мы предполагаем, что доля a этих лиц, $a \in (0; 1)$, становится восприимчивой, тогда как $1 - a$ индивидов сохраняют свой иммунитет в течение моделируемого эпидемического сезона. В результате в модель вводится функция $f(h, m)$, определяющая долю лиц с историей заражения h , восприимчивых к штамму вируса m :

$$f(h, m) = \begin{cases} a, & m = h, \\ 1, & m \neq h. \end{cases} \quad (7)$$

Уравнения модели выглядят следующим образом:

$$x_{t+1}^{(h)} = \max \left\{ 0, \left(1 - \sum_{m=1}^{n_s} \frac{\beta^{(m)}}{\rho} \bar{y}_t^{(m)} f(h, m) \right) x_t^{(h)} \right\}, h \in \overline{1, n_s + 1}, \quad (8)$$

$$y_{t+1}^{(m)} = \frac{\beta^{(m)}}{\rho} \bar{y}_t^{(m)} \sum_{h=1}^{n_s+1} f(h, m) x_t^{(h)}, m \in \overline{1, n_s},$$

$$\bar{y}_t^{(m)} = \sum_{\tau=0}^T y_{t-\tau}^{(m)} g_{\tau}^{(m)}, m \in \overline{1, n_s},$$

$$x_0^{(h)} = \alpha^{(h)} \left((1 - \mu) \rho - \sum_{m=1}^{n_s} y_0^{(m)} \right) \geq 0, h \in \overline{1, n_s + 1},$$

$$y_0^{(m)} = \varphi_0^{(m)} \geq 0, m \in \overline{1, n_s}. \quad (9)$$

Функция g_{τ} задаётся как в описанной ранее исходной модели Барояна–Рвачёва [20].

Считаем, что $\beta^{(m)}$, интенсивность эффективных контактов (т.е. контактов с последующей передачей инфекции), относящихся к штамму вируса m , определяется следующим выражением:

$$\beta^{(m)} = \lambda^{(m)} \delta,$$

где $\lambda^{(m)}$ — вирулентность штамма m , δ — среднее число контактов в популяции. Параметры модели приведены в Таблице 3.

Таблица 3. Параметры мультиштаммовой модели

Переменная	Описание	Значения
ρ	Численность населения, человек	Варьирует
$\alpha^{(h)}$	Доля населения, подвергшегося воздействию штамма m в предшествующий эпидемический сезон, $h \in \overline{1, m^*}$	Варьирует
$\lambda^{(m)}$	Вирулентность штамма вируса m	Варьирует
a	Доля людей, утративших иммунитет после контакта со штаммом вируса в предыдущем эпидемическом сезоне	Варьирует
δ	Среднесуточное количество контактов в популяции для фиксированного человека	6,528
$I_0^{(m)}$	Начальное число лиц, инфицированных штаммом m , человек	1
σ	Параметр для вычисления весов w_d точек данных	Варьирует

* Доля $\alpha^{(m+1)}$ 'наивной' популяции была рассчитана как $\alpha^{(m+1)} = 1 - \sum_{h=1}^m \alpha^{(h)}$

3.5. Мультиштаммовая модель Барояна-Рвачева с учётом предыстории заболеваемости и возрастной структуры населения

Получена совмещением двух ранее описанных моделей Барояна-Рвачева — мультиштаммовой модели и модели с возрастной структурой населения. Разработана авторами специально для практического применения в рамках эпидемического надзора,

где важно учитывать особенности динамики и регистрации заболеваемости в разных возрастных группах, а также корректно оценивать уровни иммунитета. Технические подробности модели описаны в работе [1].

3.6. Непрерывная SEIR-модель распространения COVID-19

Модель разработана авторами на основе ранее приведённой непрерывной модели распространения гриппа (см. раздел 3.3). Аналогично ей, население представлено совокупностью четырех групп лиц: восприимчивые, индивиды в инкубационном периоде, инфекционные и выздоровевшие. Размеры групп измеряются в долях общей популяции N : пусть S будет долей восприимчивых индивидов, E — долей индивидов в инкубационном периоде, I — долей инфекционных индивидов, и R — доля выздоровевших индивидов. Определенная часть населения в каждом рассматриваемом городе не подвержена COVID-19. В эту подгруппу входят люди с иммунитетом, полученным от перенесения COVID-19, и те, кто не имеет иммунитета сам по себе, но защищен коллективным иммунитетом. Восприимчивая доля населения обозначается $\alpha \in [0; 1]$, а оставшаяся часть населения считается невосприимчивой к инфекции. Динамика размеров групп во времени задается следующими уравнениями [18]:

$$\begin{aligned} \frac{dS}{dt} &= -\beta(t)SI, \\ \frac{dE}{dt} &= \beta(t)SI - \gamma E, \\ \frac{dI}{dt} &= \gamma E - \delta I, \\ \frac{dR}{dt} &= \delta I, \\ S(t_0) &= S_0 \geq 0, E(t_0) = E_0 \geq 0, I(t_0) = I_0 \geq 0, \\ S_0 + E_0 + I_0 &= \alpha, \\ R(t_0) &= 1 - \alpha. \end{aligned} \tag{10}$$

$$\tag{11}$$

Ключевым отличием данной модели от модели распространения гриппа состоит в использовании переменной величины интенсивности контактов $\beta = \beta(t)$. На динамику COVID-19 в значительной степени влияет изменение поведения людей (уменьшение социальной активности, дистанцирование, применение СИЗ) и введение мер ограничения контактов со стороны государства. По этой причине стандартные SIR-модели с постоянной интенсивностью контактов β могут завышать оценку числа заболевших в средних и поздних стадиях эпидемии. Переменная интенсивность контактов задаётся с помощью кусочно-дифференцируемой функции. Для моделирования первой волны COVID-19 использовалась функция

$$\beta(t) = \begin{cases} b_0, & t \leq t^*, \\ b_0 \cdot e^{-\mu(t-t^*)}, & t > t^*, \end{cases} \tag{12}$$

где t^* — момент введения ограничений, а μ отражает скорость воздействия ограничений на интенсивность контактов.

3.7. Непрерывная модель распространения COVID-19 с учётом бессимптомных носителей инфекции

Модификация предыдущей модели с введением подгрупп выявленных и невыявленных больных. Уравнения модели принимают следующий вид:

$$\frac{dS}{dt} = -\beta(t, \mu)S(I_0 + I_1), \quad (13)$$

$$\frac{dE}{dt} = \beta(t, \mu)S(I_0 + I_1) - \gamma E,$$

$$\frac{dI_0}{dt} = \theta\gamma E - \delta I_0,$$

$$\frac{dI_1}{dt} = (1 - \theta)\gamma E - \delta I_1,$$

$$\frac{dR}{dt} = \delta(I_0 + I_1)$$

$$S(t_0) = S^{(0)} \geq 0, E(t_0) = E^{(0)} \geq 0, I(t_0) = I_0^{(0)} \geq 0, I_1(t_0) = I_1^{(0)} \geq 0, \\ S^{(0)} + E^{(0)} + I_0^{(0)} + I_1^{(0)} = \alpha, R(t_0) = 1 - \alpha. \quad (14)$$

Здесь $S(t)$ - доля восприимчивых индивидов в популяции в момент времени t , $E(t)$ - доля индивидов в инкубационном периоде в момент времени t , $I_0(t)$ - доля невыявленных инфицированных индивидов в момент времени t , $I_1(t)$ - доля инфицированных индивидов в момент времени t , $R(t)$ - доля переболевших индивидов в момент времени t , α - начальная доля восприимчивых индивидов в популяции, δ - коэффициент интенсивности выздоровления инфицированных индивидов, γ - интенсивность перехода в стадию инфицированных индивидов, β - коэффициент интенсивности эффективных контактов индивидов. По сравнению с ранее использованной моделью, был введён новый параметр θ , отвечающий за долю невыявленных индивидов среди всех вновь заболевших. В настоящей модели было решено использовать упрощённую интерпретацию процесса выявления – количество индивидов, перешедших в категорию выявленных/невыявленных, определяется сразу после перехода заболевших в латентной форме в активную фазу заболевания и доля переходов является постоянной.

3.8. Непрерывная модель распространения COVID-19 с учётом вакцинации

Разработанная авторами популяционная модель описывает процесс изменения численности групп популяции, включающих в себя восприимчивых, латентно инфицированных, инфицированных и выздоровевших и умерших индивидов. Каждая из этих групп делится на подгруппы вакцинированных и невакцинированных. Изменение долей указанных групп в популяции в непрерывном времени задаётся следующей системой обыкновенных дифференциальных уравнений:

$$\begin{aligned}
 \frac{dS_N}{dt} &= -\beta(t)S_N(I_N + I_V) - \mu(t)S_N, \\
 \frac{dS_V}{dt} &= \mu(t)S_N - k\beta(t)S_V(I_N + I_V), \\
 \frac{dE_N}{dt} &= \beta(t)S(I_N + I_V) - \gamma E_N, \\
 \frac{dE_V}{dt} &= k\beta(t)S(I_N + I_V) - \gamma E_V, \\
 \frac{dI_N}{dt} &= \gamma E_N - (\delta + \theta)I_N, \\
 \frac{dI_V}{dt} &= \gamma E_V - (l\delta + m\theta)I_V, \\
 \frac{dR_N}{dt} &= \delta I_N, \\
 \frac{dR_V}{dt} &= l\delta I_V, \\
 \frac{dD_N}{dt} &= \sigma I_N, \\
 \frac{dD_V}{dt} &= m\sigma I_V,
 \end{aligned}
 \tag{15}$$

$$\begin{aligned}
 I_{exp}(t) &= \nu(t)(I_N + I_V), \\
 S_N(t) + S_V(t) + E_N(t) + E_V(t) + I_N(t) + I_V(t) + \\
 &+ R_N(t) + R_V(t) + D_N(t) + D_V(t) = 1, \\
 S_N(t_0) = S_N^{(0)} \geq 0, E_N(t_0) = E_N^{(0)} \geq 0, I_N(t_0) = I_N^{(0)} \geq 0, \\
 S_N^{(0)} + E_N^{(0)} + I_N^{(0)} &= \alpha \in [0; 1], R_N(t_0) = 1 - \alpha, \\
 S_V(t_0) = E_V(t_0) = I_V(t_0) = R_V(t_0) = D_N(t_0) = D_V(t_0) &= 0.
 \end{aligned}
 \tag{16}$$

Интерпретация параметров приведена в Таблице 4.

Для всех указанных моделей выполнены программные реализации на языке программирования Python, которые производят численное решение систем уравнений с заданными параметрами с помощью библиотеки `scipy`.

4. Графический интерфейс

Графический интерфейс, входящий в состав модельного комплекса, реализован на основе библиотеки `dash` языка программирования Python. Вид отдельных панелей управления представлен на рис. 2, 3 и 4. Код, отвечающий за отрисовку интерфейса, запускается на удалённом сервере, а сам интерфейс доступен через окно интернет-браузера. В настоящий момент реализован единственный режим работы, – с полным доступом к панели управления, – к которому после внедрения программного комплекса будет добавлено ещё два режима ограниченного доступа – для представителей дружественных организаций и для сторонних неавторизованных пользователей.

5. Калибровка моделей

Основным методом поиска параметров моделей по входным данным заболеваемости является модуль автоматической калибровки, который предполагает применение двух алгоритмов оптимизации: имитации отжига и L-BFGS-B. Погрешность калибровки модели вычисляется как евклидово расстояние между данными заболеваемости $Z^{(dat)}$ и

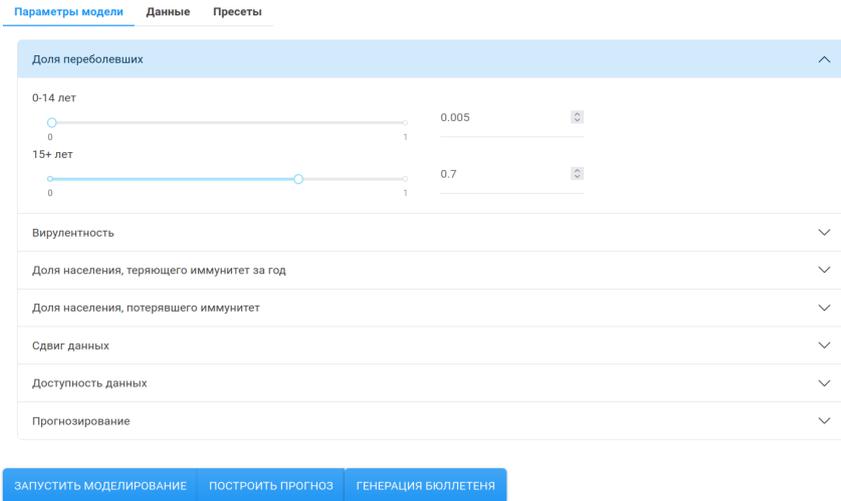


Рис. 2. Вид основной панели параметров графического интерфейса

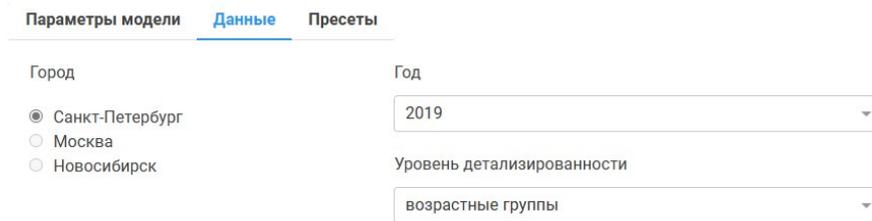


Рис. 3. Вид панели выбора данных

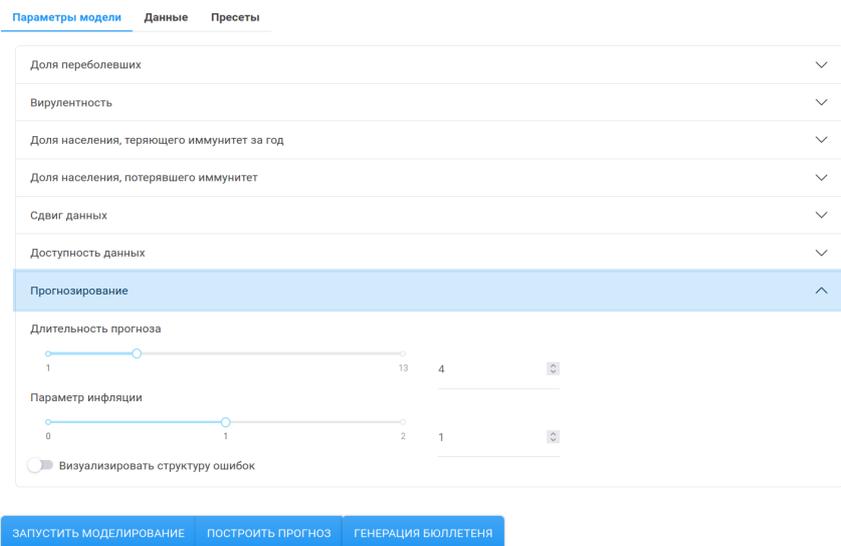


Рис. 4. Вид панели выбора параметров модельного прогноза

Таблица 4. Параметры модели с учётом вакцинации

Переменная	Описание
$S_N(t)$	Доля восприимчивых невакцинированных индивидов в популяции в момент времени t
$S_V(t)$	Доля восприимчивых вакцинированных индивидов в популяции в момент времени t
$E_N(t)$	Доля невакцинированных индивидов в инкубационном периоде в момент времени t
$E_V(t)$	Доля вакцинированных индивидов в инкубационном периоде в момент времени t
$I_N(t)$	Доля невакцинированных инфицированных индивидов в момент времени t
$I_V(t)$	Доля вакцинированных инфицированных индивидов в момент времени t
$R_N(t)$	Доля невакцинированных переболевших индивидов в момент времени t
$R_V(t)$	Доля вакцинированных переболевших индивидов в момент времени t
$D_N(t)$	Доля невакцинированных умерших индивидов в момент времени t
$D_V(t)$	Доля вакцинированных умерших индивидов в момент времени t
$I_{exp}(t)$	Доля выявленных инфицированных индивидов в момент времени t
α	Начальная доля восприимчивых индивидов в популяции
δ	Коэффициент интенсивности выздоровления инфицированных индивидов
γ	Интенсивность перехода в стадию инфицированных индивидов
$\beta(t)$	Коэффициент интенсивности эффективных контактов индивидов (контактов с последующим инфицированием)
k	Поправочный коэффициент интенсивности эффективных контактов вакцинированных индивидов
σ	Коэффициент интенсивности гибели инфицированных индивидов
m	Поправочный коэффициент интенсивности гибели вакцинированных инфицированных индивидов
l	Поправочный коэффициент интенсивности выздоровления вакцинированных инфицированных индивидов
$\nu(t)$	Коэффициент интенсивности выявления инфицированных индивидов в момент времени t
$\mu(t)$	Коэффициент интенсивности вакцинации восприимчивых индивидов в момент времени t

смоделированными значениями $Z^{(mod)}$. В формуле для метода наименьших квадратов используется вариант с весами w_i , который является обобщением стандартного МНК. Весовые коэффициенты, включенные в формулу для минимизируемой ошибки, используются для отражения «важности» подгонки кривой модели к конкретным точкам данных. В случае сезонного гриппа данные, близкие к пику, имеют большую достоверность, поскольку во время подъёма заболеваемости выборки

пациентов, у которых определяют тип вирусного агента лабораторным путём, имеют тенденцию увеличиваться и, таким образом, доля различных штаммов вируса в общей заболеваемости ОРВИ рассчитывается более точно. В случае COVID-19 использование калибровки с весами так же даёт преимущество, потому что качественная подгонка модельного пика к пику по данным является важным условием правильной оценки значимых параметров эпидемических вспышки. Пример откалиброванной модели приведён на рис. 5.



Рис. 5. Отображение результатов моделирования в графическом интерфейсе в режиме ретроспективного моделирования полного эпидемического сезона

Для оценки качества калибровки модели применяется коэффициент детерминации R^2 , который вычисляется на основе значения евклидова расстояния между модельной кривой и данными и поэтому тождественен ему, но является более простым в интерпретации.

В силу недостаточной формализованности «правдоподобного» соответствия модельной кривой реальному эпидемическому процессу количественные методы оценки не всегда могут адекватно отразить качество калибровки модели, что приводит к необходимости обсуждения результатов моделирования со специалистами предметной области и опоры на их субъективную оценку. Например, несмотря на регулировку приоритизации различных точек данных заболеваемости путём изменения весов w_i , результирующие модельные кривые зачастую сильно завышают или занижают пиковую заболеваемость, что является негативным фактором с точки зрения практического

применения моделей. По этой причине для программного комплекса авторами разработан графический пользовательский интерфейс для уточнения параметров модели путём ручной интерактивной калибровки, которая направлена на повышение адекватности работы модели с помощью экспертных знаний. Функционал работы интерфейса, связанный с интерактивной калибровкой, представляет собой развитие более ранних работ авторов по полуавтоматическому ретроспективному анализу данных заболеваемости [25]. Начальными значениями параметров для ручной настройки являются результаты оценок параметров, полученные в результате описанной ранее процедуры оптимизации. Эта особенность работы комплекса даёт ему преимущество по сравнению с «эпидемическими калькуляторами», поскольку последние обычно имеют лишь один вариант поиска параметров модели — с автоматической калибровкой на данные либо с ручным выставлением значений параметров в графическом интерфейсе.

Помимо возможности тонкой настройки, панель изменения параметров модели позволяет использовать программный комплекс в качестве системы поддержки принятия решений. Корректируя коэффициенты интенсивности эффективных контактов, которые регулируют вероятность передачи инфекции между восприимчивыми и инфицированными людьми, исследователи могут моделировать эффекты действия различных внешних и внутренних факторов на заболеваемость (введение ограничительных мер, приход новых штаммов эпидемических ОРВИ, последствия массовых мероприятий, пр.). Аналогичным образом изменение доли иммунных индивидов, отражающее либо результаты предшествующего заражения, либо влияние вакцинации, позволяет исследовать пороговые значения коллективного иммунитета и потенциальную эффективность различных стратегий иммунизации.

6. Количественная оценка неопределённости

Чтобы оценить неопределённость выходных показателей модели, связанной с неопределённостью в данных, был выбран метод генерации ошибки в данных с фиксированным вероятностным распределением, выбранным таким образом, чтобы результирующие модельные кривые целиком покрывали всю вариацию точек данных вокруг гипотетической средней величины заболеваемости. В соответствии с методом параметрического бутстрэпа, описанного более подробно в [2], смоделированная ошибка добавляется к модельной кривой заболеваемости $Z^{(\text{mod})}(\hat{\theta})$, которая получена из модели, откалиброванной на исходных данных заболеваемости $Z^{(\text{dat})}$. Генерация ошибки производится методом Монте-Карло в соответствии с выбранным распределением и повторяется N раз, таким образом, давая N новых массивов данных. Эти массивы данных могут далее использоваться для оценки неопределённости при ретроспективном анализе заболеваемости и построении интервальных прогнозов, а также позволяют получать интервальные оценки эпидемических параметров, таких как R_0 и R_t . В последнем случае для получения выборок значений параметров размерности N требуется перекалибровка используемой модели N раз на новых данных, что является трудоёмким процессом. По этой причине соответствующие вычисления не осуществляются в интерактивном режиме, а графики с интервальными оценками не демонстрируются посредством графического интерфейса. Эти графики используются при генерации итоговых отчётов, информация про которые приведена в следующей главе. Пример прогнозного графика приведен на рис. 6.

7. Еженедельные бюллетени и годовые отчёты

Одной из основных функций описываемого в работе программного инструмента, помимо точечной и интервальной оценки эпидемических параметров, является

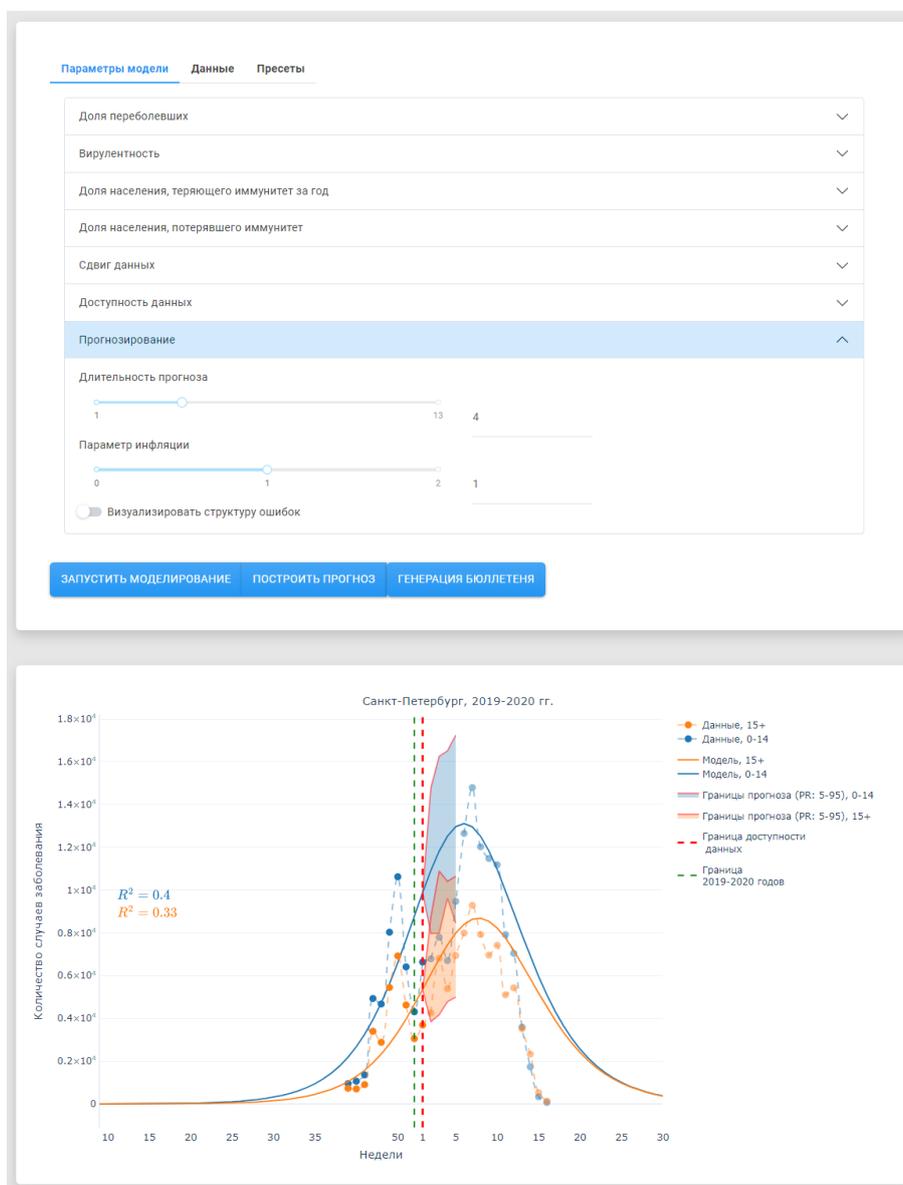


Рис. 6. Отображение результатов моделирования в графическом интерфейсе в режиме прогнозирования по неполным данным заболеваемости

использование оценок этих параметров для генерации отчётных материалов. В настоящее время предусмотрена генерация двух вариантов материалов:

- Отчёт по прогнозу динамики заболеваемости для текущего эпидемического сезона и оценка динамических параметров по неполным данным заболеваемости, таких как R_t и динамика доли иммунных индивидов. Отчёт призван дополнять информацию еженедельных бюллетеней заболеваемости, подготавливаемых НИИ гриппа для онлайн-размещения [26].
- Итоговый отчёт для прошедшего эпидемического сезона, содержащий вдобавок к интервальным оценкам динамических параметров распределения скалярных параметров, таких как R_0 и финальная оценка уровня иммунитета на момент окончания эпидемической вспышки. Отчёт призван дополнять информацию ежегодных публикаций НИИ гриппа и отчётов Министерства здравоохранения со статистикой по прошедшим сезонам гриппа и других эпидемических ОРВИ (см., например, [27]).

В текущей версии программного модельного комплекса генерация отчётов производится в формате pdf с помощью библиотеки pdf_latex языка программирования Python. За запуск соответствующей процедуры отвечает кнопка «Генерация бюллетеня» графического интерфейса. Содержимое отчётов формируется из модельных графиков, табличных оценок эпидемических показателей и фиксированного сопровождающего текста, в который добавляются полученные значения параметров. Видится перспективным разработка алгоритма для более гибкой генерации текстовых описаний на базе больших языковых моделей открытого доступа. Это позволит больше приблизить результат к текущей форме написанного вручную бюллетеня и освободить работников органов надзора от наиболее рутинных операций.

Пример применения программного комплекса для генерации отчёта по конкретному эпидемическому сезону приведен в следующем разделе.

3. ВЫЧИСЛИТЕЛЬНЫЙ ЭКСПЕРИМЕНТ

Демонстрация работы программного комплекса осуществлялась со следующими условиями вычислительного эксперимента.

Данные: Эпидемический сезон гриппа 2019–2020 гг. в г. Санкт-Петербург

Модели: дискретные модели Барояна-Рвачёва в вариантах без разбиения на возрастные группы (см. раздел 3.1) разбиением по возрастам (раздел 3.2), по штаммам гриппа (раздел 3.4), одновременно по возрастам и штаммам (раздел 3.5)

Тип отчёта: еженедельный

Момент отсчёта модельного времени: 1 неделя (30.12-05.01) 2020 года

Горизонт прогнозирования: 6 недель

Выбор города и сезона не принципиален для демонстрации работы программного комплекса, разница может заключаться лишь в картине эпидемической динамики, которая влияет на качество калибровки модели (некоторые сезоны в некоторых городах демонстрируют близкую к теоретической кривую заболеваемости, в то время как в другие сезоны неопределённость данных и их отклонение от классической колоколообразной кривой возрастает). Город Санкт-Петербург был выбран для демонстрации в силу того, что это место работы авторов исследования, а сезон 2019–2020 гг. примечателен тем, что является последним доковидным сезоном циркуляции гриппа.

Этапы выполнения работы, начиная с загрузки исходных данных из баз НИИ гриппа им. Смородинцева, выглядят следующим образом:

- В панели «Данные» графического интерфейса выбирается город Санкт-Петербург и эпидемический сезон 2019-2020 гг.
- Для каждого выбранного в панели графическом интерфейсе типа детализации данных (агрегированные данные, разбивка по возрастам, разбивка по штаммам, разбивка по возрастам и штаммам):
 - На основе массива суммарных данных заболеваемости всеми ОРВИ и массива данных вирусных исследований генерируются данные заболеваемости с разбивкой по возрастным группам и штаммам гриппа (как альтернатива, они могут быть посчитаны предварительно для всех интересующих городов и эпидсезонов при условии, что рассматриваются уже завершившиеся эпидемические вспышки, т.е. нет необходимости в усвоении данных моделями в реальном времени).
 - В зависимости от выбранного типа данных, осуществляется автоматическая калибровка соответствующей модели на неполных данных заболеваемости (от начала эпидемической вспышки до 1 недели 2020 года), оцениваются параметры модели и вычисляются точечные оценки эпидемических индикаторов.
 - С помощью процедуры бутстрэппинга генерируется $N = 200$ массивов данных с ошибкой, распределённой по негативному биномиальному закону.
 - В графическом интерфейсе отрисовывается модельная кривая на данных заболеваемости с интервальной оценкой прогноза.
 - Совместно с коллегами из НИИ гриппа им. Смородинцева обсуждается корректность описания данных полученной модельной кривой, а также правдоподобность полученных эпидемических индикаторов (R_0 , R_t) с точки зрения предметной области. По результату обсуждений осуществляется ручная корректировка параметров модели в графическом интерфейсе с перерисовкой модельной прогнозной кривой.
 - Зафиксированные значения параметров модели и подготовленные графики сохраняются на сервере для последующей генерации бюллетеня.
- Нажимается кнопка «Генерация бюллетеня».
- С сервера загружаются результаты моделирования для четырёх вариантов детализации данных.
- Начинается формирование файла бюллетеня путём дополнения шаблона текста в формате TeX значениями параметров моделей и вставкой графиков с прогнозами.

- В фоновом режиме начинается расчёт интервальных оценок эпидемических индикаторов (R_t , доля приобретших иммунитет).
- После окончания расчётов в текст отчёта вставляются интервальные оценки эпидемических индикаторов в виде таблиц и графиков.
- Происходит компиляция итогового файла отчёта в формате pdf.

Вид первой страницы сгенерированного отчёта приведён на рис. 7.

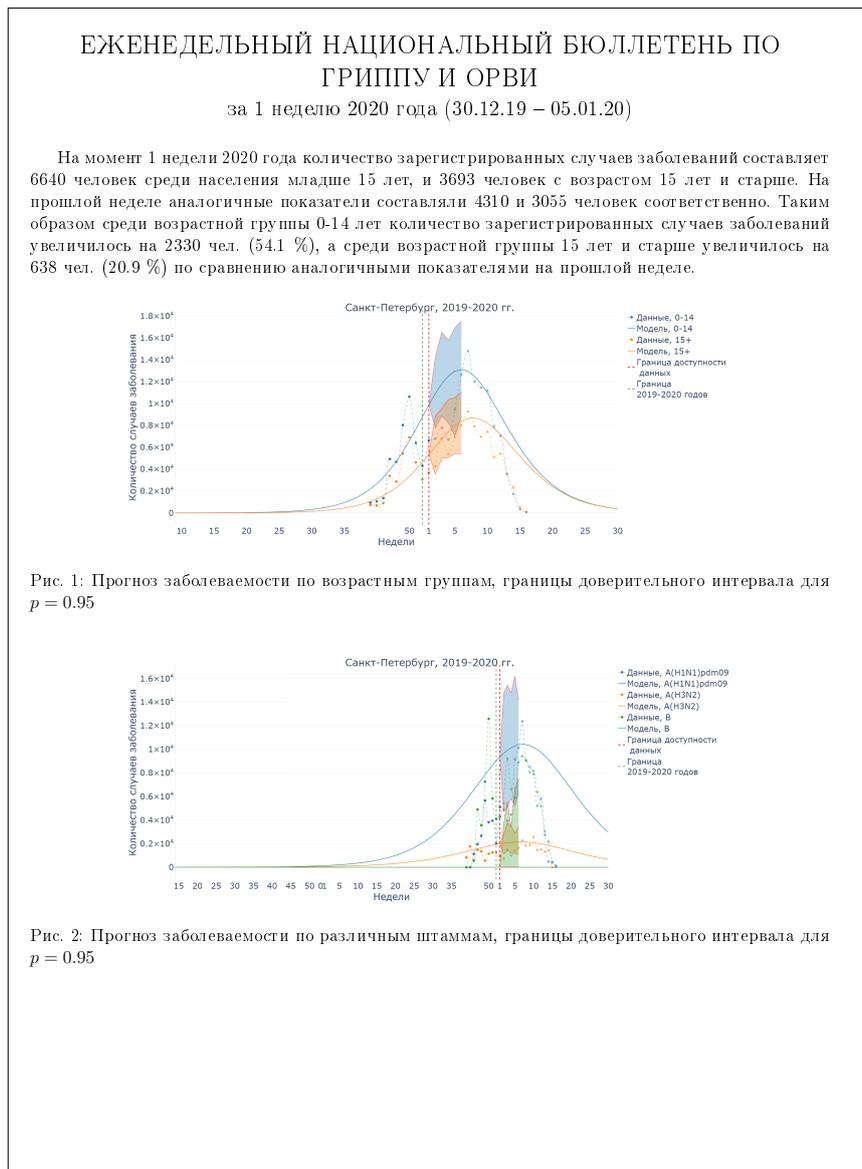


Рис. 7. Пример сгенерированного еженедельного бюллетеня

4. ЗАКЛЮЧЕНИЕ

В данной работе авторами предложен программный комплекс на основе реализации математических моделей динамики эпидемических ОРВИ (гриппа и COVID-19), который в совокупности с реализованным графическим интерфейсом и модулем

для генерации отчётов позволяет производить извлечение знаний непосредственно из данных заболеваемости, регистрируемых надзорными органами. Получаемые с помощью комплекса интервальные оценки эпидемических индикаторов и краткосрочные прогнозы дают дополнительную значимую информацию по динамике эпидемического процесса, которая не может быть получена простыми статистическими подходами. Ключевой особенностью представленного комплекса является его интеграция в программные средства, используемые надзорными органами, что позволяет производить оперативное усвоение данных заболеваемости с генерацией отчётных документов, проводя, таким образом, анализ заболеваемости эпидемическими ОРВИ в реальном времени и увеличивая эффективность работы эпидемиологического надзора. Опыт участия авторов в прогнозировании динамики COVID-19 для Межведомственного городского координационного совета по противодействию распространению в Санкт-Петербурге новой коронавирусной инфекции [28] показал, что сами по себе прогнозирующие модели не являются системами поддержки принятия решений как таковыми, поскольку требуют дополнительных времязатрат на преобразование данных и генерацию стандартизованных выходных результатов. Внедрение в практику предлагаемого в данной работе комплекса ускорит получение информации, необходимой для принятия решений экспертами.

К дополнительным возможностям комплекса, на которых авторы не стали заострять внимание в данном материале из-за недостатка места, является использование индикатора АИС вдобавок к коэффициенту детерминации для сравнения эффективности применения разных классов моделей к разным массивам эпидемических данных [18], [1], а также оценка влияния введения тех или иных мер контроля заболеваемости на прогнозы по динамике инфекции, т.е. проигрыш возможных гипотетических сценариев развития эпидемического процесса.

Среди перспективных направлений дальнейшего улучшения эффективности программного комплекса можно выделить следующие:

- На данный момент в графическом интерфейсе отсутствует возможность независимого выбора детализации модели и детализации входных данных (выбор детализации модели осуществляется автоматически на основе данных). Между тем, калибровка более детальных моделей на менее детальные данные, а также наоборот, возможна [1] и в ряде случаев даже необходима (например, при отсутствии данных достаточной детализации для определенных городов и эпидсезонов). Также наличие отдельного пункта выбора структурной сложности модели непосредственно в интерфейсе позволит пользователям самостоятельно сравнивать качество разных моделей с помощью индикатора АИС. Представляется перспективным добавить новый пункт меню на панель приборов и выводить результаты расчёта АИС на графиках.
- Для повышения качества воспроизведения эпидемической динамики моделями возможна реализация и сравнение альтернативных алгоритмов автоматической калибровки модели на данные. Конкретной целевой задачей является уменьшение числа случаев попадания оптимизационного алгоритма в локальные оптимумы. Авторами предполагается тестирование методов калибровки на базе генетических алгоритмов, методов Монте-Карло с марковскими цепями (MCMC), а также аналитических методов, опирающихся на вид уравнения моделей.
- Отдельные эпидемические сезоны характеризуются массивами данных заболеваемости, которые неудовлетворительно описываются моделями типа SIR. Эти сезоны нуждаются в дополнительном скрупулезном изучении совместно

со специалистами предметной области. Ряд таких случаев может быть объяснён ошибками в самих данных, а также излишней упрощенностью подходов на базе SIR-моделей по сравнению с реальной эпидемической динамикой. Авторами предполагается расширение списка используемых в комплексе моделей, в частности, добавление стохастических и мультиагентных моделей, а также разработку новых методов калибровки и поиска дополнительных источников данных в связи с увеличенной сложностью подобных моделей.

- Интересным представляется включение в состав комплекса гибридных моделей, совмещающих классические объясняющие модели с моделями машинного обучения. В настоящее время исследуются возможности применения для прогнозирования подхода PINN (Physics-informed neural networks) [29], который позволяет автоматически учитывать зависимость неопределенности прогнозов от разброса входных данных без прибегания к методам параметрического бутстрэппинга. Также многообещающим подходом может стать использование подхода AI-augmented SEIR («SEIR-модели, дополненные методами искусственного интеллекта»), который, по заявлению их авторов [30], способен значительно повышать качество прогнозирования SEIR-моделями и делать их менее зависимыми от шума и ошибок в данных.

Поскольку комплекс позволяет гибко изменять номенклатуру применяемых для прогнозирования моделей, это даёт возможность использовать его в качестве платформы для ансамблевого моделирования и прогнозирования с привлечением нескольких научных коллективов, аналогично инициативе FluSight Challenge Центра по контролю и профилактике заболеваний США (CDC), которая проводится начиная с сезона гриппа 2013–2014 гг. Более общая «Инициатива по прогнозированию эпидемий» (EPI), которая включает в себя, помимо гриппа, предсказание COVID-19 и других опасных инфекций, «направлена на повышение научности и удобства использования прогнозов эпидемий путем содействия открытым проектам прогнозирования, преследующим конкретные цели общественного здравоохранения» [31]. Исследователями, участвующими в инициативе, отмечается, что проект активизировал сообщество ученых, заинтересованных в прогнозировании, создав испытательный полигон для улучшения как технического понимания того, как работают различные модели прогнозирования, так и интеграции этих моделей в процесс принятия решений. Технические средства проекта FluSight обеспечивают возможность стандартизированного сравнения моделей в одном приложении, что значительно улучшает понимание того, какие модели работают лучше всего в определенных условиях, как результаты могут лучше всего распространяться и использоваться лицами, принимающими решения, и каковы узкие места с точки зрения улучшения прогнозов [32].

Авторам представляется, что в свете растущего признания в мире важности системной интеграции методов прогнозирования в работу органов здравоохранения целесообразно унифицировать модельные комплексы для прогнозирования ОРВИ в РФ и расширять широкомасштабное сотрудничество для разработки общих стандартов данных и облегчения сравнения нескольких моделей, реализуя аналог вышеописанной инициативы CDC. Наличие множества научных коллективов в РФ, разрабатывающих прогнозные модели динамики ОРВИ, позволяет перейти к их интеграции в рамках единого сервиса и перекрестному сравнению, что в конечном итоге позволит оперативнее отвечать на угрозы пандемий вирусов нового типа.

Исследование выполнено при поддержке гранта НЦМУ «Передовые цифровые технологии»

СПИСОК ЛИТЕРАТУРЫ

1. Huaman I., Plesovskaya E. P., Leonenko V. N. Matching model complexity with data detail: influenza propagation modeling as a case study. In: *2022 IEEE International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON)*. IEEE, 2022. P. 650–654. doi: [10.1109/SIBIRCON56155.2022.10017084](https://doi.org/10.1109/SIBIRCON56155.2022.10017084)
2. Sahatova K., Kharlunin A., Huaman I., Leonenko V. Accounting for Data Uncertainty in Modeling Acute Respiratory Infections: Influenza in Saint Petersburg as a Case Study. In: *International Conference on Computational Science*. Cham : Springer Nature Switzerland, 2023. P. 286–299. doi: [10.1007/978-3-031-36024-4_23](https://doi.org/10.1007/978-3-031-36024-4_23)
3. Cauchemez, S., Valleron A.-J., Boelle P.-Y., Flahault A., Ferguson N.M. Estimating the impact of school closure on influenza transmission from Sentinel data. *Nature*. 2008. V. 452. N. 7188. P. 750–754. doi: [10.1038/nature06732](https://doi.org/10.1038/nature06732)
4. Bajardi P., Poletto C., Ramasco J.J., Tizzoni M., Colizza V., Vespignani A. Human mobility networks, travel restrictions, and the global spread of 2009 H1N1 pandemic. *PloS one*. 2011. V. 6. N. 1. Article No. e16591. doi: [10.1371/journal.pone.0016591](https://doi.org/10.1371/journal.pone.0016591)
5. Read J.M. Bridgen J.R.E. Cummings D.A.T., Ho A., Jewell C.P. Novel coronavirus 2019-nCoV (COVID-19): early estimation of epidemiological parameters and epidemic size estimates. *Philosophical Transactions of the Royal Society B*. 2021. V. 376. N. 1829. Article No. 20200265.
6. Maier B. F., Brockmann D. Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China. *Science*. 2020. V. 368. N. 6492. P. 742–746. doi: [10.1126/science.abb4557](https://doi.org/10.1126/science.abb4557)
7. Kumar S., Grefenstette J.J., Galloway D., Albert S.M., Burke D.S. Policies to reduce influenza in the workplace: impact assessments using an agent-based model. *American journal of public health*. 2013. V. 103. N. 8. P. 1406–1411. doi: [10.2105/AJPH.2013.301269](https://doi.org/10.2105/AJPH.2013.301269)
8. Lukens S., DePasse J., Rosenfeld R., Ghedin E., Mochan E., Brown S.T., Grefenstette J., Burke D.S., Swigon D., Clermont G. A large-scale immuno-epidemiological simulation of influenza A epidemics. *BMC public health*. 2014. V. 14. P. 1–15. doi: [10.1186/1471-2458-14-1019](https://doi.org/10.1186/1471-2458-14-1019)
9. Guo D., Li K.C., Peters T.R., Snively B.M., Poehling K. A., Zhou X. Multi-scale modeling for the transmission of influenza and the evaluation of interventions toward it. *Scientific reports*. 2015. V. 5. N. 1. Article No. 8980. doi: [10.1038/srep08980](https://doi.org/10.1038/srep08980)
10. Hoertel N., Blachier M., Blanco C., Olfson M., Massetti M., Limosin F., Leleu H. Facing the COVID-19 epidemic in NYC: a stochastic agent-based model of various intervention strategies. *MedRxiv*. 2020. doi: [10.1101/2020.04.23.20076885](https://doi.org/10.1101/2020.04.23.20076885)
11. Viboud C., Vespignani A. The future of influenza forecasts. *Proceedings of the National Academy of Sciences*. 2019. V. 116. N. 8. P. 2802–2804. doi: [10.1073/pnas.1822167116](https://doi.org/10.1073/pnas.1822167116)
12. Arora P., Kumar H., Panigrahi B. K. Prediction and analysis of COVID-19 positive cases using deep learning models: A descriptive case study of India. *Chaos, solitons & fractals*. 2020. V. 139. P. 110017. doi: [10.1109/SIBIRCON56155.2022.10017084](https://doi.org/10.1109/SIBIRCON56155.2022.10017084) DOI: [10.1016/j.chaos.2020.11001](https://doi.org/10.1016/j.chaos.2020.11001)
13. *FRED. U.S. Measles Simulator* URL: <https://fred.publichealth.pitt.edu/measles> (дата обращения: 29.11.2023).
14. *Математическое моделирование распространения инфекционных заболеваний*. URL: <https://covid19-modeling.ru/modeling> (дата обращения: 29.11.2023).
15. *OpenAI. Epidemic Calculator* URL: <https://gabgoh.github.io/COVID/> (дата обращения: 29.11.2023).

16. Левашкин, Р. П. Информационная система для визуализации данных КОВИД-19 и изучения контрольных параметров модели распространения пандемии в России и мире. *Информационное общество*. 2023. № 3. С. 55–68. doi: [10.52605/16059921_2023_03_55](https://doi.org/10.52605/16059921_2023_03_55)
17. Kouprianov, A. *Monitoring COVID-19 epidemic in St. Petersburg, Russia: Data and scripts*. URL: <https://github.com/alexei-kouprianov/COVID-19.SPb.monitoring> (дата обращения: 29.11.2023).
18. Huaman I., Leonenko V. Does Complex Mean Accurate: Comparing COVID-19 Propagation Models with Different Structural Complexity. *International Conference on Computational Science*. Cham : Springer Nature Switzerland, 2023. P. 270–277. doi: [10.1007/978-3-031-36024-4_21](https://doi.org/10.1007/978-3-031-36024-4_21)
19. Harko T., Lobo F.S.N., Mak M.K. Exact analytical solutions of the Susceptible-Infected-Recovered (SIR) epidemic model and of the SIR model with equal death and birth rates. *Applied Mathematics and Computation*. 2014. V. 236. P. 184–194. doi: [10.1016/j.amc.2014.03.030](https://doi.org/10.1016/j.amc.2014.03.030)
20. Бароян О.В., Рвачев Л.А., Иванников Ю.Г. *Моделирование и прогнозирование эпидемий группа для территории СССР*. М.: Ин-т эпидемиологии и микробиологии им. Н.Ф. Гамалеи АМН СССР, 1977. 546 с.
21. Leonenko V., Bobashev G. Analyzing influenza outbreaks in Russia using an age-structured dynamic transmission model. *Epidemics*. 2019. V. 29. Article No. 100358. doi: [10.1016/j.epidem.2019.100358](https://doi.org/10.1016/j.epidem.2019.100358)
22. Anderson R.M., May R.M. *Infectious diseases of humans: dynamics and control*. Oxford University Press, 1991. 757 p.
23. Leonenko V. N. Herd immunity levels and multi-strain influenza epidemics in Russia: a modelling study. *Russian Journal of Numerical Analysis and Mathematical Modelling*. 2021. V. 36. N. 5. P. 279–291. doi: [10.1515/rnam-2021-00234](https://doi.org/10.1515/rnam-2021-00234)
24. Hill E. M., Petrou S., Forster H., de Lusignan S., Yonova I., Keeling, M.J. Seasonal influenza: Modelling approaches to capture immunity propagation. *PLoS computational biology*. 2019. V. 15. N. 10. Article No. e1007096. doi: [0.1371/journal.pcbi.1007096](https://doi.org/10.1371/journal.pcbi.1007096)
25. K. Sahatova, A. Kharlunin and V. Leonenko. A Novel Approach to Modeling and Visualisation of Epidemic Outbreaks: Combining Manual and Automatic Calibration. *2023 IEEE Ural-Siberian Conference on Computational Technologies in Cognitive Science, Genomics and Biomedicine (CSGB)*. Novosibirsk, Russian Federation, 2023, P. 221-226, doi: [10.1109/CSGB60362.2023.10329850](https://doi.org/10.1109/CSGB60362.2023.10329850)
26. ФГБУ НИИ гриппа имени А.А. Смородинцева Минздрава РФ. *Еженедельный бюллетень по группе*. URL: www.influenza.spb.ru/system/epidemic_situation/laboratory_diagnostics/ (дата обращения: 29.11.2023).
27. Селькова Е. П., Гренкова Т.А., Гудова Н.В., Оганесян А.С. Итоги эпидсезона 2017/18 гг. по гриппу и острой респираторной вирусной инфекции. Особенности этиотропной терапии. *Русский медицинский журнал. Медицинское обозрение*. 2018. V. 2. N. 11. P. 49–53.
28. Экспертная группа по мерам борьбы с пандемией COVID-19. *Аналитическая записка к Правительству Санкт-Петербурга*. 26.03.2020. URL: <https://static.ngs.ru/news/2020/99/images/e707764127224490f11a985fed7762b2.pdf> (дата обращения: 29.11.2023).
29. Grimm V., Heinlein A., Klawonn A., Lanser M., Weber J. Estimating the time-dependent contact rate of SIR and SEIR models in mathematical epidemiology using physics-informed neural networks. *Electronic Transactions on Numerical Analysis*. 2022. V. 56. P. 1–27.
30. Arik S.Ö., Shor J., Sinha R., Yoon J., Ledsam J.R., Le L.T., Dusenberry M.W., Yoder N.C.,

Popendorf K., Epshteyn A. et al. A prospective evaluation of AI-augmented epidemiology to forecast COVID-19 in the USA and Japan. *NPJ digital medicine*. 2021. V. 4. N. 1. Article No. 146. doi: [10.1038/s41746-021-00511-7](https://doi.org/10.1038/s41746-021-00511-7)

31. PhiResearchLab. *Epidemic prediction initiative* URL: <https://predict.phiresearchlab.org/> (дата обращения: 29.11.2023).
32. Reich N.G., Brooks L.C., Fox S.J., Kandula S., McGowan C.J., Moore E., Osthus D., Ray E.L., Tushar A., Yamana T.K. et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academy of Sciences*. 2019. V. 116. N. 8. P. 3146–3154. doi: [10.1073/pnas.1812594116](https://doi.org/10.1073/pnas.1812594116)

Рукопись поступила в редакцию 30.11.2023.

Переработанный вариант поступил 07.12.2023.

Дата опубликования 13.12.2023.

Mathematical Biology and Bioinformatics

doi: [10.17537/2023.18.517](https://doi.org/10.17537/2023.18.517)

===== INFORMATION AND COMPUTER =====
===== TECHNOLOGIES IN BIOLOGY AND MEDICINE =====

Application of Mathematical Models of the Dynamics of the Epidemic Acute Respiratory Viral Infections to Increase the Efficiency of Epidemiological Surveillance

Leonenko V.N.^{1,2}, Korzin A.I.¹, Danilenko D.M.²

¹*ITMO University, St Petersburg, Russian Federation*

²*Smorodintsev Research Institute for Influenza, St Petersburg, Russian Federation*

Abstract. Uncertainty in the calculations of forecasts of the spread of epidemic acute respiratory infections obtained using mathematical models, associated with data error and uncertainty in the choice of a model, as well as the lack of verification of modeling results by interdisciplinary teams including epidemiology specialists, prevent the correct prediction of the effectiveness of disease control measures. In this paper, we propose a solution to these problems by using a software package consisting of a family of epidemic models, methods for estimating the error of output data depending on the error of the initial morbidity data, as well as a graphical interface with the possibility of manual correction of the results of automatic calibration and generation of epidemic bulletins. The novelty of the presented study is the methodology for integrating epidemic models into software tools used by supervisory authorities, which allows to supplement weekly bulletins and annual epidemiological reports in semi-automatic mode with a quantitative interval estimation of the error of calculated indicators. The ultimate goal is to provide the supervisory authorities with informative and promptly obtained calculated data for decision-making in the field of infection control.

Key words: *epidemiology, morbidity forecasts, acute respiratory viral infections, Baroyan-Rvachev model, Kermack-McKendrick model, software package.*