

УДК 616-036.22

<http://dx.doi.org/10.22328/2077-9828-2022-14-1-90-99>

ПОДХОДЫ К ПРОГНОЗИРОВАНИЮ ВРЕМЕННОЙ ДИНАМИКИ ПРОЦЕССОВ И ИСПОЛЬЗОВАНИЕ ЗАКОНА ЭВОЛЮЦИИ ЭПИДЕМИЙ ФАРРА НА ПРИМЕРЕ ДИНАМИКИ НОВЫХ СЛУЧАЕВ ВИЧ-ИНФЕКЦИИ В РОССИЙСКОЙ ФЕДЕРАЦИИ

© А. Н. Барина, А. А. Лебедева, М. В. Гусаров, С. Л. Плавинский*

Северо-Западный государственный медицинский университет имени И. И. Мечникова, Санкт-Петербург, Россия

Введение. Прогнозирование течения различных процессов, включая эпидемический, является важным в общественном здоровье. Существует мнение, что в ряде случаев более простые модели дают адекватную возможность прогнозирования. *Целью данного исследования* было оценить применимость и результаты прогнозирования новых случаев ВИЧ-инфекции в Российской Федерации на основании давно известного закона Фарра.

Материалы и методы. Были использованы данные официального статистического наблюдения за числом вновь регистрируемых случаев ВИЧ-инфекции в Российской Федерации в 1999–2020 гг. Параметры для прогнозирования новых случаев на период до 2027 г. были рассчитаны в соответствии с методикой D. J. Bregman и A. D. Langmuir. Для контроля прогностической точности прогноз был оценен на периоде 2015–2020 гг. после расчетов по данным до 2014 г. Были выделены нормальные компоненты эмпирической кривой и оценен истинный характер их распределения.

Результаты. Расчеты по закону Фарра несколько занижали истинное число случаев ВИЧ-инфекции (составляя 99% (95% ДИ 92–106%) при использовании сглаженных данных и 97% (95% ДИ 89–106%) при использовании исходных данных). В целом, в особенности по сглаженным данным, отклонения были приемлемыми. Прогноз до 2027 г. показывает, что общее число зарегистрированных в 1999–2027 гг. случаев составит 1,7–2,0 млн. Анализ истинного распределения второго пика эпидемической кривой показывает, что оно является логнормальным, что предполагает возможность значительно большего числа инфицированных в средне- и долгосрочной перспективе.

Заключение. Хотя закон Фарра может использоваться для краткосрочного прогноза, не рекомендуется ослаблять профилактические программы ввиду возможного увеличения числа ВИЧ-инфицированных в сравнении с прогнозированием по закону Фарра.

Ключевые слова: прогнозирование временных процессов, ВИЧ-инфекция, закон Фарра, теория распределения экстремальных значений

*Контакт: Плавинский Святослав Леонидович, s.plavinskij@szgmu.ru

APPROACHES TO FORECASTING OF TIME DYNAMICS OF THE PROCESSES AND USE OF FARR'S EPIDEMIC EVOLUTION LAW ON EXAMPLE OF DYNAMICS OF NEW CASES OF HIV-INFECTION IN RUSSIAN FEDERATION

© A. N. Barinova, A. A. Lebedeva, M. V. Gusarov, S. L. Plavinskii*

North-Western State Medical University named after I. I. Mechnikov, St. Petersburg, Russia

Introduction. Forecasting of different processes in health, including epidemics, are important area of public health. There exists an idea that in some cases simple models can give adequate forecasts.

Goal of this study was to evaluate possible use and results of forecasting of registration of new cases of HIV infection in Russian Federation based on well-known Farr's law.

Materials and methods. The official statistical data on new HIV cases in 1999–2020 in Russian Federation were used. Parameters for forecasting new cases until 2027 were calculated according to Bregman and Langmuir. For evaluation of forecasting the calculations were done for 2015–2020 after fitting model with data until 2014. Normal components of the empirical epidemic curve were estimated and more appropriately fitted distributions were found for the data described by those components.

Results. Estimations according to the Farr's law somewhat undercount number of the new cases of HIV infection (it forecast 99% (95% CI 92–106%) cases when smoothing was used and 97% (95% CI 89–106%) when raw data were used). In general, especially when smoothing was used, fit was satisfactory. Forecast until 2027 show that total number of HIV cases in 1999–2027 will

be 1.7–2.0 mln people. Analysis of most probable distribution of the second peak of epidemic curve show that it is lognormal, which allow for much larger number of infected in medium- and long-term perspective.

Conclusion. Though Farr's law could be used for short-term forecast it is not recommended to weaken preventive programs due to possibility of large increase in number of HIV-infected in comparison with Farr's law forecast.

Key words: time-series forecasting, HIV-infection, Farr's law, extreme value distribution theory

*Contact: Plavinsky Svyatoslav Leonidovich, s.plavinskij@szgmu.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Для цитирования: Барينو А.Н., Лебедева А.А., Гусаров М.В., Плавинский С.Л. Подходы к прогнозированию временной динамики процессов и использование закона эволюции эпидемий Фарра на примере динамики новых случаев ВИЧ-инфекции в Российской Федерации // *ВИЧ-инфекция и иммуносупрессии*. 2022. Т. 14, № 1. С. 90–99, doi: <http://dx.doi.org/10.22328/2077-9828-2022-14-1-90-99>.

Conflict of interest: the authors stated that there is no potential conflict of interest.

For citation: Barinova A.N., Lebedeva A.A., Gusarov M.V., Plavinskii S.L. Approaches to forecasting of time dynamics of the processes and use of Farr's epidemic evolution law on example of dynamics of new cases of HIV-infection in Russian Federation // *HIV Infection and Immunosuppressive Disorders*. 2022. Vol. 14, No. 1. P. 90–99, doi: <http://dx.doi.org/10.22328/2077-9828-2022-14-1-90-99>.

Введение. Прогнозирование развития эпидемического процесса является одной из важнейших задач общественного здоровья, поскольку позволяет подготовить силы и средства и сконцентрировать их на тех направлениях, где ущерб может оказаться максимальным, а стоимость-эффективность применения средств будет давать наилучший результат. Моделирование и прогнозирование относятся к инструментам современного изучения эпидемического процесса, в том числе ВИЧ-инфекции [1].

Однако прогнозирование осложняется тем, что действовать приходится в условиях неопределенности и надо выбрать не только наиболее адекватную прогнозную модель, но и подобрать данные, которые будут использоваться в модели. При этом данные могут и не собираться службами общественного здоровья, что будет приводить к их нерепрезентативности (при сборе данных ad hoc), снижая точность предсказания, а иногда делая модель и вовсе неадекватной.

Строго говоря, описание эпидемического процесса может складываться из двух основных теоретических подходов: (1) понимания законов распространения заразного начала в коллективах лиц и взаимодействия людей друг с другом [2, 3]. Подобный подход называется математическим эпидемиологическим моделированием, и он стал очень популярен на заре развития эпидемии ВИЧ-инфекции/СПИД, позволяя отвечать на вопросы «а что, если...» будет предпринято то или иное вмешательство [2]. Можно относиться к временным сериям как к другим статистическим данным —

когда есть специфический механизм их генерации и исследователь может его описать, отделив от дополнительных компонентов и случайного (чаще всего «белого» шума).

Теоретическая (эпидемиологическая) модель иногда называется «контейнерной», поскольку предполагает существование гомогенных групп лиц, которые контактируют друг с другом с одинаковой частотой. Фактически все лица, находящиеся внутри «контейнера», имеют одинаковую вероятность контакта друг с другом и членами других «контейнеров». По тому, какие «контейнеры» выделяют, модели классифицируют как SIR (Susceptible-Infected-Removed: Уязвимый-Инфицированный-Удаленный (умерший или приобретший иммунитет в результате заболевания/вакцинации)). Популярным в начале пандемии COVID-19 вариантом этой модели была модель SEIR (Susceptible-Exposed-Infected-Removed: Уязвимый-Проконтактировавший с заразным-Инфицированный-Удаленный (умерший/приобретший иммунитет)). Модели могли быть модифицированы с включением мер противодействия (социального дистанцирования [4], потребностей в оборудовании и персонале и т.д.). Поскольку формирование иммунитета не является обязательным, то возможны варианты модели, в которых выздоровевший может заразиться снова (SIS: Susceptible-Infected-Susceptible: Уязвимый-Инфицированный-Уязвимый) — популярные при изучении инфекций, передающихся преимущественно половым путем. Технически модели формулировались как системы дифференциальных

уравнений, для которых существовали пакеты прикладных статистических программ, что упрощало моделирование. Кроме того, можно было делать допущения открытости популяции (в популяции появляются новые члены) или ее закрытости (число убывших равнялось числу прибывших), а также постоянности параметров уравнений (детерминистские варианты) или их случайного варьирования (стохастические). Модели могли быть на удивление гибкими, но качество предсказания напрямую зависело от качества входных параметров, и результат мог варьировать в большой степени. В особенности в начале распространения заболевания, когда важнейшие показатели: заразность, длительность заразного периода и частота контактов уязвимых с заразными (часто объединяемые в виде «репродуктивного числа», R) — не могли быть оценены с достаточной степенью точности.

Контейнерные модели по определению являются серьезным упрощением, не учитывая географию местности, где распространяется заболевание, плотности контактов между жителями разных районов и т.п. Именно поэтому были предложены так называемые агентские модели, где вначале строится математическая модель некоего сообщества — на основании изучения реальных данных — например, взаимодействия людей в определенном городе. После построения и валидации модели анализируется вопрос о том, что произойдет, если в этом обществе возникнет вспышка некоего заболевания (например, оспы) и анализируются последствия возможных профилактических мероприятий. Однако результаты, полученные на модели одной общины, не переносимы на другие (генерализация отсутствует).

Другой подход к прогнозированию — статистический, с использованием в качестве источника данных временных рядов. Для анализа временных рядов как набора данных, уже давно используется подходы с их разложением на ряд компонентов. Обычно выделяют три основных компонента временных рядов: тренд (а), сезонность (б) и циклы (в) — не считая случайной ошибки данных. Когда временной ряд разлагается на компоненты, обычно тренд и цикл объединяется в один компонент тренд-цикл (иногда для простоты называемый трендом). Таким образом, система упрощается так, что временной ряд состоит из трех компонентов: компонента цикла тренда, сезонного компонента и компонента остатка (содержащего все остальное

во временном ряду). Часто это делается для улучшения понимания временных рядов, но также может использоваться для повышения точности прогнозов.

R. J. Hyndman и G. Athanasopoulos [5] выделяют несколько типов прогнозных методов:

- а) оценочные суждения;
- б) регрессионные модели;
- в) разложение временных серий;
- г) экспоненциальное сглаживание;
- д) модели типа ARIMA;
- е) динамические регрессионные модели;
- ж) другие подходы.

Простейшими являются методы разложения/сглаживания. Классический метод декомпозиции временных рядов зародился в 1920-х годах и широко использовался до 1950-х годов. Он по-прежнему составляет основу многих методов разложения временных рядов, поэтому важно понимать, как он работает. Первым шагом в классической декомпозиции является использование метода скользящего среднего для оценки тренд-цикла.

Оценка тренд-цикла в момент времени t получается путем усреднения значений временного ряда за k периодов t . Наблюдения, близкие по времени, также, вероятно, будут иметь близкие значения. Следовательно, среднее значение устраняет некоторую случайность в данных, оставляя компонент сглаженного тренд-цикла. Обычно это обозначается сокращением m -МА (от английского *Moving Averages* — скользящие средние), что означает скользящую среднюю порядка m .

Простые скользящие средние обычно имеют нечетный порядок (например, 3, 5, 7 и т.д.). Это значит, что они симметричны: в скользящем среднем порядка $m = 2k+1$ среднее наблюдение и k наблюдений с каждой стороны усредняются. Но если бы m было четным, оно больше не было бы симметричным.

Полученные средние можно дополнительно усреднить с помощью все той же методики скользящих средних. Одна из причин для этого — сделать скользящее среднее четного порядка симметричным. Например, мы могли бы взять скользящее среднее 4-го порядка, а затем применить к результатам другое скользящее среднее 2-го порядка.

Когда 2-МА следует за скользящей средней четного порядка (например, 4), это называется «центрированной скользящей средней четвертого порядка». Это потому, что теперь результаты симметричны.

Возможны и другие комбинации скользящих средних. Например, часто используется 3×3 -МА, которая состоит из скользящей средней 3-го порядка, за которой следует еще одна скользящая средняя 3-го порядка. В общем, за МА четного порядка должна следовать МА четного порядка, чтобы сделать ее симметричной. Точно так же за МА нечетного порядка должна следовать МА нечетного порядка.

Чаще всего центрированные скользящие средние используются для оценки тренд-цикла на основе сезонных данных. Применительно к квартальным данным каждому кварталу года придается равный вес, поскольку первый и последний коэффициенты применяются к одному и тому же кварталу в последовательные годы. Следовательно, сезонные колебания будут усреднены, и в результирующих значениях. Они, таким образом, устраняются. Аналогичный эффект может быть получен при использовании 2×8 -МА или 2×12 -МА для квартальных данных.

В общем, $2 \times m$ -МА эквивалентен взвешенному скользящему среднему порядка $m+1$, где все наблюдения принимают вес $1/m$, за исключением первого и последнего членов, которые принимают веса $1/(2m)$. Итак, если сезонный период четный и порядка m , используется $2 \times m$ -МА для оценки тренд-цикла. Если сезонный период нечетный и порядка m , используется m -МА для оценки тренд-цикла. Например, 2×12 МА может использоваться для оценки цикла тренда месячных данных, а 7-МА может использоваться для оценки цикла тренда дневных данных с недельной сезонностью.

Другой выбор порядка МА обычно приводит к загрязнению оценок цикла тренда из-за сезонности данных.

Иногда требуется ответить на простой вопрос: все ли повторяется как всегда или наступили какие-то необычные изменения? Часто такой вид анализ называется «анализом точки изменения» (Change point analysis), хотя могут использоваться и лингвистически иные термины, и математически другие подходы. Простейший вариант можно проиллюстрировать следующим образом. Анализируется заболеваемость ОРВИ на 30 неделе года. Для определения границ недели можно использовать функцию НОМНЕДЕЛИ системы MS EXCEL. Данные за дни недели усредняются.

После этого в качестве референтных берутся данные за 5 лет для 30 недели и усредняются (сравнение недели с неделями устраняет/сглаживает сезонный компонент). Наличие двух соседних недель позволяет вычислить среднее значение для недели 30 и стандартное отклонение для недель 29–31 (т.е. используя уже 105 точек). Установить, является ли значение «типичным» для недели 30, можно, сравнив расчетную величину данного года с усредненным значением за 5 лет и диапазоном, равным среднему за 5 лет $+1,96 \times$ стандартное отклонение по данным 105 точек (верхняя граница интервала) и равным среднему за 5 лет $-1,96 \times$ стандартное отклонение по данным 105 точек (нижняя граница интервала). Для повышения специфичности можно использовать не стандартное отклонение, а ошибку среднего. При использовании такого подхода прогнозирование не проводится на большую глубину, а скорее используется для ответа на вопрос о том, не отличаются ли данные для точки t в этом году от данных точки t в предшествующие годы. Таким образом, можно создать систему раннего предупреждения о биотеррористической атаке или появлении новой, необычной пневмонии. Для этих целей очень часто используются математические инструменты набора специалистов по контролю за качеством (контрольные графики). Например, для нахождения точки изменения используется методика кумулятивных сумм или CUSUM. Известно, что самым активным убийцей-одиночкой был английский семейный врач Гарольд Шипман. Поскольку он имел право выписывать свидетельства о смерти, то ничто не мешало ему убить, в основном инъекциями героина, 250 своих пациентов. После проведенного ретроспективного анализа количества умерших на участке, где работал Шипман, D. Spiegelhalter и соавт. [6] писали, что «статистический мониторинг мог бы привести к появлению требующего исследования сигнала в конце 1996 года, когда число избыточных смертей у женщин старше 65 составило 67 в сравнении со 119 в 1998 году»¹.

Однако вернемся к определению формы эпидемической кривой. Хотя за вторую половину XX в. человечество привыкло к термину «вакциноуправляемые заболевания», понимая под ними то, что есть вакцина, которая повлияет на ход эпидемического процесса, либо приведя к его эрадикации

¹ Шипман попался в 1998 г., когда подделал завещание своей пациентки в свою пользу и ввел ей героин. Жалоба со стороны дочери умершей и запустила расследование.

(оспа), либо сильно ослабив последствия (краснуха). Но ведь так было не всегда, эпидемии приходили, поражали определенную часть общества и отступали. Так было во время «русского гриппа» 1889–1890 гг., когда погибло около миллиона человек (это некоторыми авторами рассматривается как первое столкновение с коронавирусной инфекцией (ОС43) [7, 8]), от гриппа «испанки» также не было вакцины, пандемия гриппа 1957–1958 гг. достигла пика до того, как вакцина стала доступной в значительных количествах. Это означает, что если после перенесенного заболевания остается иммунитет (SIR-подобный механизм распространения), то заболевание должно следовать некоему общему закону. М. Santillana и соавт. [9] показали, что SIR модель аппроксимируется «законом Фарра», названным так в честь эпидемиолога, жившего на рубеже XIX–XX веков и обратившего в 1840 г. внимание на то, что эпидемии, достигая своего расцвета, повторяли одну и ту же симметричную форму с достижением пика и медленным снижением, что позволяло — на коротком промежутке времени — делать прогноз относительно дальнейшего развития эпидемического процесса (Фарр анализировал вспышку натуральной оспы в Англии и Уэльсе).

D. J. Bregman и A. D. Langmuir в статье 1990 г., посвященной применению закона Фарра для оценки числа случаев СПИД в США, дают следующую его формулировку: «В случае возникновения большой эпидемии количество новых случаев имеет тенденцию нарастать к некоторому максимуму и затем снижаться так, что вся кривая напоминает кривую нормального распределения» [10]. Сам закон был популяризирован Brownlee, а Фарр использовал его для предсказания времени наступления пика и общего числа потерь от чумы крупного рогатого скота в 1866 г. В настоящее время интерес к закону Фарра не угас. Его использовали для предсказания числа умерших от передозировок наркотиков в США [11], а также описания поведения «волн» COVID-19 [12, 13]. Целью данного исследования было попытаться применить закон Фарра к кривой новых случаев ВИЧ-инфекции в Российской Федерации с целью ответа на вопрос о прохождении максимума случаев заражения.

Материалы и методы. Для анализа были взяты данные официального статистического наблюде-

ния за выявлением новых случаев ВИЧ-инфекции в 1999–2020 годах¹ [14, 15].

Расчеты были выполнены на основании методики, представленной D. J. Bregman и A. D. Langmuir [10]. Оценивались отношения R_1 (число новых случаев в данном году, деленное на количество в предшествующем) и R_2 (отношение R_1 в данном году к R_1 в предшествующем). Для прогнозирования числа новых случаев последние три наблюдавшихся отношения R_2 были усреднены и использованы как оценка отношений R_2 в прогнозируемом периоде. Для проверки модели период использования данных ограничивался 2014 годом, чтобы на протяжении 2015–2020 гг. можно было провести сравнение наблюдаемого числа новых случаев и ожидаемого в соответствии с законом Фарра. Сравнение проводилось расчетом процента ошибки (отношения истинного значения к предсказанному, выраженное в процентах) и его суммарных показателей (средняя абсолютная процентная ошибка, MAPE [5]). Расчеты были выполнены как для исходных данных, так и для сглаженных по методике простого движущегося среднего с временным окном в 3 года [16]. Для определения точности прогноза были рассчитаны доверительные интервалы для MAPE с помощью методики откорректированного на смещение бутстрапа (BCa). Если доверительные интервалы пересекали 100%, то можно было принять нулевую гипотезу о том, что прогнозные и реальные значения не отличаются на уровне ошибки I типа 0,05. Кроме того, было выполнено прогнозирование по закону Фарра на основе всей совокупности данных. Однако ввиду того, что данные за 2019–2020 гг. различались очень значительно (что могло быть связано с ограничением контактов лиц с рискованным поведением с системой здравоохранения в 2020 г. ввиду начала пандемии COVID-19 и уменьшением их вероятности тестирования (на $12,7\%^2$), для прогнозирования по полным данным использовались простые движущиеся средние числа новых случаев ВИЧ-инфекции. Для оценки параметров закона Фарра были взяты данные 2016–2018 гг. Оценен прогноз числа случаев ВИЧ-инфекции в сочетании с уже выявленными случаями в диапазоне с 1999 по 2027 г.

Автоматическое извлечение параметров нормального распределения было выполнено с помо-

¹ <http://www.hivrussia.info/dannye-po-vich-infektsii-v-rossii>. Дата обращения 07.09.2021.>

² <http://www.hivrussia.info/wp-content/uploads/2021/03/VICH-infektsiya-v-Rossijskoj-Federatsii-na-31.12.2020-.pdf>. Дата обращения 09.09.2021.

щью пакета mixtools [17] системы R версии 3.5 (R Foundation for Statistical Computing, Vienna, Austria). Проверка наиболее вероятного распределения данных в диапазоне периодов, извлеченных при разложении данных на нормальные распределения, было выполнено при помощи процедуры SEVERITY системы SAS (SAS On Demand for Academics, Cary, NC, USA) с учетом факта «усеченности» распределений.

Более сложные подходы к прогнозированию временных рядов (например, ARIMA) не применялись ввиду короткой длительности временной серии и отсутствие в данных сезонного компонента, требуемого многими пакетами для применения техники ARIMA.

Результаты и их обсуждение. Результаты моделирования распределения количества новых случаев по закону Фарра представлены на рис. 1. Моделирование осуществлялось на основании временного ряда до 2014 г.

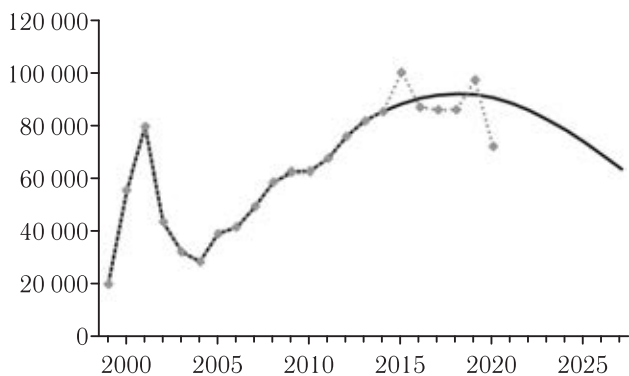


Рис. 1. Количество вновь выявляемых случаев ВИЧ-инфекции согласно данным официального статистического наблюдения и согласно закону Фарра с прогнозом до 2027 г.
Fig. 1. Number of newly diagnosed HIV infections according to official statistical surveillance data and according to Farr's law with a forecast up to 2027

Видно, что закон Фарра предсказывает прохождение максимума числа заражений в 2017–2019 гг. и затем плавное снижение до горизонта моделирования (2027 г.). Видно, что данные 2020 г. достаточно резко отличаются от прогнозных и от предшествующих, по всей вероятности ввиду снижения числа обращений потенциально заразившихся лиц в систему здравоохранения ввиду переориентировки ее на противодействие новой коронавирусной инфекции. Сравнение наблюдаемых и предсказанных значений в диапазоне 2015–2020 гг. (которые в данной модели для прогнозирования не использовались) приведено в табл. 1.

Как видно из таблицы, отношения наблюдаемых значений к прогнозным были близки вне зависимо-

Таблица 1
Процент ошибок для наблюдаемых данных вновь выявленных случаев ВИЧ-инфекции в Российской Федерации в сравнении с предсказанными по закону Фарра
Table 1
Percentage error for observed data of newly diagnosed HIV cases in the Russian Federation compared to those predicted by Farr's law

Год	Количество ошибок, %	
	со сглаживанием	без сглаживания
2015	113	114
2016	96	96
2017	94	94
2018	100	93
2019	108	106
2020	85	80
Суммарно	99	97
95% ДИ	92–106	89–106

сти от того, используются ли исходные данные или к ним, как принято для временных серий, была применена процедура сглаживания, хотя при сравнении со сглаженными данными прогноз и реальные данные совпадали лучше. Если на исходной серии колебания составляли от 80 до 114%, то на сглаженной — от 85 до 113%. Однако суммарно все равно было видно, что предсказанные по закону Фарра значения были меньше, чем наблюдаемые, — 97% в случае исходных данных и 99% в случае использования сглаженных данных. Тем не менее можно заметить, что предсказание числа случаев ВИЧ инфекции на основании очень простого закона Фарра на горизонте в 6 лет оказываются достаточно приемлемыми, поскольку доверительные интервалы включали 100% (равенство предсказанных и реальных значений) и не были очень широкими.

На рис. 2 показано соотношение наблюдаемых значений новых случаев ВИЧ-инфекции и предсказанных на основании сглаженных данных и закона Фарра с экстраполяцией до 2027 г. (с исключением из данных для применения закона Фарра резкого подъема числа случаев в 2019 г. и снижения в 2020 г., но учетом их при сглаживании).

Как видно из рис. 2, сглаживание приводит к тому, что наблюдаемые значения оказываются немного выше сглаженных, хотя в общем естественно повторяют форму эпидемической кривой. Интересно, что даже некоторый подъем в 2019 г. предсказан по закону Фарра на сглаженных данных, хотя и не такой выраженный, как зарегистрирован в данных официального статистического наблюдения.

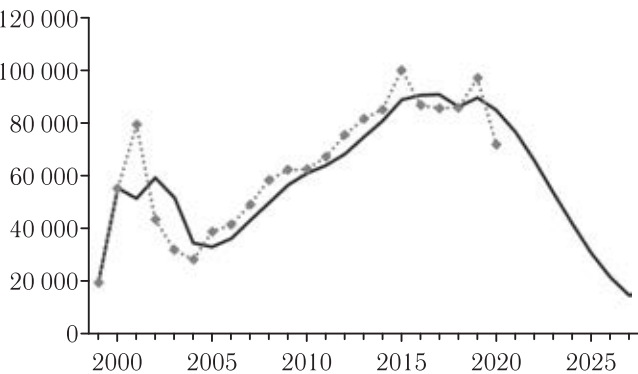


Рис. 2. Количество вновь выявленных случаев ВИЧ-инфекции согласно данным официального статистического наблюдения (сглаженные движущиеся средние) и согласно закону Фарра с прогнозом до 2027 г.

Fig. 2. Number of newly diagnosed HIV infections according to official statistical surveillance data (smoothed moving averages) and according to Farr's law with a forecast up to 2027

Как уже отмечалось выше, закон Фарра базируется на допущении, что эпидемическая кривая имеет форму, близкую к форме нормального распределения (при этом не исключается несколько «волн» эпидемии). Поскольку на графиках видна явная

и будут выявлены в Российской Федерации в период с 1999 по 2027 г. Эти результаты вместе со сравнительными данными за 1999–2020 гг. приведены в табл. 2.

Как видно из табл. 2, в зависимости от источника данных, оценки общего числа случаев ВИЧ-инфекции с 1999 по 2027 г. разнятся. Если за основу брать данные до 2014 г., то прогнозируется почти 2 млн случаев. Если берутся сглаженные данные полного периода, то количество случаев снижается до 1,7 млн. Из них, соответственно, 1,42 млн прогнозируются до 2020 г. при использовании укороченного ряда и 1,37 — при использовании полного сглаженного ряда. Расхождения связаны с тем, что коэффициенты закона Фарра на укороченном ряду оказываются таковыми, что снижения числа новых случаев в 2020–2027 гг. прогнозируется значительно более плавным в сравнении с полным сглаженным рядом. Сглаживание, как было отмечено выше, занижает число новых случаев, что видно на примере сравнения не сглаженных наблюдаемых данных за период

Таблица 2

Прогнозируемое число вновь выявленных случаев ВИЧ-инфекции с 1999 до 2027 г. в сравнении с данными официального статистического наблюдения до 2020 г. (тыс. чел.)

Table 2

Projected number of newly diagnosed cases of HIV infection from 1999 to 2027 compared with official statistical data until 2020 (thousand people)

Показатель Источник данных, период	Общее число случаев, период	
	1999–2027 гг.	1999–2020 гг.
Сглаженные значения, 1999–2020 гг.	1679	1373
Исходные значения, 1999–2014 гг.	1964	1426
Наблюдаемые данные, 1999–2020 гг. (не сглаженные)	н/п	1410
Наблюдаемые данные, 2001–2020 гг. (сглаженные)	н/п	1298
Наблюдаемые данные, 2001–2020 гг. (не сглаженные)	н/п	1334

Примечание: н/п — не применимо.

бимодальность распределения с двумя «волнами» заражений, была предпринята попытка разложить видимую последовательность на два компонента с нормальным распределением. Установлено, что действительно можно выделить две компоненты с нормальным распределением, одна из них с пиком в 2000–2001 гг. и длительностью до 2004 г., вторая — с пиком в 2015 г. и началом с началом наблюдений. Точкой наилучшего разделения этих двух компонент были 2004 (относимый к первой компоненте) и 2005 (относимый ко второй компоненте) годы.

Приближенность предсказанных значений к наблюдаемым позволяла оценить общее количество случаев ВИЧ инфекции, которые были

1999–2020 гг. — 1,41 млн, что очень близко к данным, использованным с укороченным рядом (до 2014 г.) и дальнейшей экстраполяцией до 2020 г. по закону Фарра. Различия сглаженных и не сглаженных данных можно проследить к началу временной серии, поскольку различия данных 2001–2020 гг. значительно меньше — 1,30 млн в случае сглаженных данных и 1,33 млн в случае не сглаженных — разность около 30 тыс. случаев.

Таким образом, закон Фарра, несмотря на свою простоту, позволяет достаточно эффективно экстраполировать форму эпидемической кривой и оценивать число случаев заболевания (и смертности, хотя в данном исследовании эта гипотеза не проверялась).

Вместе с тем интерпретировать данные, полученные с помощью закона Фарра, надо с осторожностью в случае использования оценок на среднесрочную и долгосрочную перспективу. Дело в том, что нормальное распределение (на использовании которого он построен) может плохо аппроксимировать вероятность наступления событий, находящихся на значительном удалении от среднего. Это происходит в связи с тем, что в реальности распределение имеет «тяжелые» хвосты, т.е. вероятность наступления редкого события выше, чем можно предсказать при использовании нормального распределения. На это указывал N. Taleb [18] в 2020 г. в контексте обсуждения реакции на эпидемию COVID-19, отмечая, что в случае распределения с «тяжелыми» хвостами обычные методы моделирования могут не дать реального заключения о количестве заболевших и умерших, поэтому для управления рисками надо использовать сценарий наихудшего развития событий. Также в 2020 г. P. Crillo и N. Taleb [19] опубликовали анализ смертности от вспышек инфекционных заболеваний и показали, что распределение потерь человеческих жизней является распределением с «тяжелыми» хвостами и анализировать данные надо с применением теории экстремальных значений. В нашем исследовании попытка анализа характера распределения числа заразившихся ВИЧ-инфекцией в 2005–2020 гг. (вторая компонента бимодального распределения) с учетом того, что данные показывают снижение, но все равно еще достаточно велики, показывает, что наиболее адекватным распределением, описывающим эти данные будет логнормальное. Оно относится к семейству распределений с «тяжелыми» хвостами и, соответственно, вероятности сохранения высокого уровня заболеваемости в среднесрочной и долгосрочной перспективе должны анализироваться с помощью инструментов теории экстремальных значений, однако для ее применения данных такой временной серии недостаточно. Именно поэтому, перенося позицию N. Taleba на заболеваемость ВИЧ-инфекцией нельзя исклю-

чить новый выраженный подъем заболеваемости и поэтому, с точки зрения борьбы с ВИЧ-инфекцией следует не снижать работу по ее профилактике, в особенности в группах населения повышенного риска (потребители инъекционных наркотиков, секс-работники и мужчин, занимающихся сексом с мужчинами), тем более что биоповеденческое исследование 2017 г. [20] показало, что пораженность ВИЧ в этих группах позволяет дополнительное распространение заболевания. Как отмечает N. Taleb [21], в ряде случаев важна не точность прогноза, а готовность к ответу на вызовы. При этом, возможно, надо расширять профилактический инструментарий, поскольку, как было показано в исследовании А. Е. Grulich и соавт. [22], ежедневная доконтактная пероральная профилактика ВИЧ-инфекции тенофовиром в сочетании с эмитрицитабином в группе МСМ высокого риска снижала в долгосрочной перспективе (3 года) риск заражения ВИЧ-инфекцией почти в 10 раз.

Заключение. Закон Фарра позволяет оценивать течение эпидемии на краткосрочную перспективу на этапе ее близости к эпидемическому «пику». Применение закона Фарра к данным официального статистического наблюдения позволяет заключить, что общее число зарегистрированных случаев ВИЧ-инфекции в 1999–2027 гг. достигнет 1,7–2,0 млн. Важными для применения закона Фарра являются эффекты сглаживания временных рядов, а также полнота и точность регистрации случаев заболевания на близких к моменту начала экстраполяции временных периодах.

Судя по результатам анализа, реальное распределение случаев не является нормальным, как того требует закон Фарра, а имеет «тяжелые» хвосты, распределение значений в которых подчиняется теории экстремальных значений.

Возможность появления значительно большего числа инфицированных в среднесрочной и долгосрочной перспективе требует активной работы по профилактике ВИЧ-инфекции, в особенности в группах населения повышенного риска.

ЛИТЕРАТУРА/REFERENCES

1. Беляков Н.А., Рассохин В.В., Розенталь В.В., Огурцова С.В., Степанова Е.В., Мельникова Т.Н. и др. Эпидемиология. ВИЧ-инфекции. Место мониторинга, научных и дозорных наблюдений, моделирования и прогнозирования обстановки // *ВИЧ-инфекция и иммуносупрессии*. 2019. Т. 11, № 2. С. 7–26. [Belyakov N.A., Rassohin V.V., Rozental V.V., Ogurcova S.V., Stepanova E.V., Melnikova T.N. et al. Epidemiology of HIV infection. Monitoring, scientific and surveillance studies, modeling and forecasting of situation. *HIV Infection and Immunosuppressive Disorders*, 2019, Vol. 11, No. 2, pp. 7–26 (In Russ.).]

2. Розенталь В.В., Беляков Н.А., Пантелеева О.В. Подходы к прогнозированию эпидемии ВИЧ-инфекции. *ВИЧ-инфекция и иммуносупрессии*. 2010. Vol. 2, No. 3. P. 7–15. [Rozental V.V., Belyakov N.A., Panteleeva O.V. Approaches to forecasting HIV-infection. *HIV Infection and Immunosuppressive Disorders*, 2010, Vol. 2, No. 3, pp. 7–15 (In Russ.)].
3. Розенталь В.В., Виноградова Т.Н., Беляков Н.А., Пантелеева О.В. Динамическая модель для описания и прогнозирования течения эпидемии ВИЧ-инфекции // *ВИЧ-инфекция и иммуносупрессии*. 2012. Т. 12, № 1. С. 95–102. [Rozental V.V., Vinogradova T.N., Belyakov N.A., Panteleeva O.V. Dynamic model for description and forecasting HIV epidemics. *HIV Infection and Immunosuppressive Disorders*, 2012, Vol. 12, No. 1, pp. 95–102 (In Russ.)].
4. Mwalili S., Kimathi M., Ojiambo V., Gathungu D., Mbogo R. SEIR model for COVID-19 dynamics incorporating the environment and social distancing // *BMC Res Notes*. 2020. Vol. 13, No. 1. P. 352.
5. Hyndman R., Athanasopoulos G. *Forecasting: Principles and Practice*. 2nd ed. Melbourne: OTexts, 2018.
6. Spiegelhalter D., Grigg O., Kinsman R., Treasure T. Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery // *Int. J. Qual Health Care*. 2003. Vol. 15, No. 1. P. 7–13.
7. King A. An uncommon cold // *New Sci*. 2020. Vol. 246, No. 3280. P. 32–35.
8. Brussow H., Brussow L. Clinical evidence that the pandemic from 1889 to 1891 commonly called the Russian flu might have been an earlier coronavirus pandemic // *Microb. Biotechnol*. 2021. Jul.
9. Santillana M., Tuite A., Nasserie T., Fine P., Champredon D., Chindelevitch L. et al. Relatedness of the incidence decay with exponential adjustment (IDEA) model, «Farr's law» and SIR compartmental difference equation models // *Infect Dis Model*. 2018. Vol. 3. P. 1–12.
10. Bregman D.J., Langmuir A.D. Farr's law applied to AIDS projections // *JAMA*. 1990. Vol. 263, No. 11. P. 1522–1525.
11. Darakjy S., Brady J.E., DiMaggio C.J., Li G. Applying Farr's Law to project the drug overdose mortality epidemic in the United States // *Inj. Epidemiol*. 2014. Vol. 1, No. 1. P. 31.
12. Bhattacharya S., Islam M.M., De A. Search for the trend of COVID-19 infection following Farr's law, IDEA model and power law // *medRxiv*. 2020.
13. Pacheco-Barrios K., Cardenas-Rojas A., Giannoni-Luza S., Fregni F. COVID-19 pandemic and Farr's law: A global comparison and prediction of outbreak acceleration and deceleration rates // *PLoS One*. 2020. Vol. 15, No. 9. P. e0239175.
14. Министерство здравоохранения Российской Федерации. *Социально значимые заболевания населения России в 2014 году*. Москва: Министерство здравоохранения Российской Федерации. Департамент мониторинга, анализа и стратегического развития здравоохранения. ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, 2015. 72 с. [Ministry of Health of the Russian Federation. *Socially important diseases in Russian population in 2014*. Moscow: Ministry of the health of Russian Federation. Department of monitoring, analysis and strategic development of health care. FSBI «Central scientific research of health care organization and informatization» of Ministry of health of the Russian Federation, 2015. 72 p. (In Russ.)].
15. Министерство здравоохранения Российской Федерации. *Социально значимые заболевания населения России в 2018 году*. Москва: Министерство здравоохранения Российской Федерации. Департамент мониторинга, анализа и стратегического развития здравоохранения. ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, 2019. 73 с. [Ministry of Health of the Russian Federation. *Socially important diseases in Russian population in 2018*. Moscow: Ministry of the health of Russian Federation. Department of monitoring, analysis and strategic development of health care. FSBI «Central scientific research of health care organization and informatization» of Ministry of health of the Russian Federation, 2019. 73 p. (In Russ.)].
16. Boslaugh S., Watters P.A. *Statistics in a Nutshell*. Sebastopol: O'Reilly & Associates, Inc., 2008. 476 p.
17. Benaglia T., Chauveau D., Hunter D.R., Young D. mixtools: An R Package for Analyzing Finite Mixture Models // *Journal of Statistical Software*. 2009. Vol. 32, No. 6. P. 1–29.
18. Taleb N.N., Bar-Yam Y., Cirillo P. On single point forecasts for fat-tailed variables // *Int. J. Forecast*. 2020. Oct. P. 1–10.
19. Cirillo P., Taleb N.N. Tail risk of contagious diseases // *Nature Physics*. 2020. Vol. 16, No. 6. P. 606–613.
20. Плавинский С.Л., Ладная Н.Н., Зайцева Е.Е., Баранова А.Н. Пораженность ВИЧ-инфекцией среди уязвимых групп населения в России — результаты интегрированного биоповеденческого исследования в 2017 г. // *Журнал микробиологии, эпидемиологии и иммунобиологии*. 2018. № 6. С. 10–18. [Plavinski S.L., Ladnaya N.N., Zaytseva E.E. Prevalence of HIV-infection among vulnerable population groups in Russian Federation, results of the bio-behavioral study of 2017. *Journal of microbiology, epidemiology and immunobiology*, 2018, No. 6, pp. 10–18 (In Russ.)].
21. Taleb N.N. *The Black Swan: The Impact of the Highly Improbable*. 2nd ed. London: Random House, 2010.
22. Grulich A.E., Jin F., Bavinton B.R., Yeung B., Hammoud M.A., Amin J. et al. Long-term protection from HIV infection with oral HIV pre-exposure prophylaxis in gay and bisexual men: findings from the expanded and extended EPIC-NSW prospective implementation study // *Lancet HIV*. 2021. Vol. 8, No. 8. P. e486–e494.

Авторство:

Вклад в концепцию и план исследования — *А. Н. Барина*ва, *А. А. Лебеде*ва, *С. Л. Плавин*ский. Вклад в сбор данных — *А. А. Лебеде*ва, *М. В. Гусар*ов. Вклад в анализ данных и выводы — *А. А. Лебеде*ва, *А. Н. Барина*ва, *М. В. Гусар*ов, *С. Л. Плавин*ский. Вклад в подготовку рукописи — *А. А. Лебеде*ва, *А. Н. Барина*ва, *М. В. Гусар*ов.

Сведения об авторах:

*Барина*ва Анна Николаевна — профессор кафедры общественного здоровья и управления здравоохранением ФГБОУ ВО «Северо-Западный государственный медицинский университет имени И. И. Мечникова» МЗ РФ; 191015, Санкт-Петербург, Кирочная ул., д. 41; e-mail: anna_n_barinova@mail.ru; ORCID 0000–0002–8180–9340; SPIN-код 2010–4354;

*Лебеде*ва Анна Александровна — очный аспирант кафедры общественного здоровья и управления здравоохранением ФГБОУ ВО «Северо-Западный государственный медицинский университет имени И. И. Мечникова» МЗ РФ; 191015, Санкт-Петербург, Кирочная ул., д. 41; e-mail: anna.lebedeva@mail.ru;

*Гусар*ов Михаил Вячеславович — очный аспирант кафедры общественного здоровья и управления здравоохранением ФГБОУ ВО «Северо-Западный государственный медицинский университет имени И. И. Мечникова» МЗ РФ; 191015, Санкт-Петербург, Кирочная ул., д. 41; e-mail: mvgusarov@gmail.com; ORCID 0000–0003–2848–1726; SPIN-код 9447–5958;

*Плавин*ский Святослав Леонидович — профессор кафедры педагогики, философии и права ФГБОУ ВО «Северо-Западный государственный медицинский университет имени И. И. Мечникова» МЗ РФ; 191015, Санкт-Петербург, Кирочная ул., д. 41; e-mail: s.plavinskij@szgmu.ru; ORCID 0000–0001–9159–6177; SPIN-код 5660–4661.