



УДК 51.76+614.2

## ПРЕДСКАЗАНИЕ ПИКОВ ЭПИДЕМИЙ ГРИППА В САНКТ-ПЕТЕРБУРГЕ С ПОМОЩЬЮ ПОПУЛЯЦИОННЫХ МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ

В.Н. Леоненко<sup>а</sup>, Ю.К. Новоселова<sup>а</sup>, К.М. Онг<sup>б</sup><sup>а</sup> Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация<sup>б</sup> Школа медицины Нью-Йоркского университета, Нью-Йорк, 10016, СШААдрес для переписки: [sergei.v.ivanov@gmail.com](mailto:sergei.v.ivanov@gmail.com)**Информация о статье**

Поступила в редакцию 20.09.16, принята к печати 30.10.16

doi: 10.17586/2226-1494-2016-16-6-1145-1148

Язык статьи – русский

**Ссылка для цитирования:** Леоненко В.Н., Новоселова Ю.К., Онг К.М. Предсказание пиков эпидемий гриппа в Санкт-Петербурге с помощью популяционных математических моделей // Научно-технический вестник информационных технологий, механики и оптики. 2016. Т. 16. № 6. С. 1145–1148. doi: 10.17586/2226-1494-2016-16-6-1145-1148**Аннотация**

Исследованы методы предсказания пиков эпидемий гриппа с применением популяционных математических моделей: Барояна–Рвачева и адаптированной модели Кермака–Маккендрика, предложенной авторами статьи. Выполнен сравнительный анализ точности предсказания времени и величины пиков эпидемий на многолетних данных по заболеваемости острыми респираторными заболеваниями в Санкт-Петербурге. В основе метода сравнения лежат три критерия точности с условными названиями «квадрат», «вертикальная полоса», «горизонтальная полоса» и два варианта оценивания параметров модели. В первом варианте модель калибруется на данных первого города, охваченного эпидемией, и используется в дальнейшем для других городов, что позволяет учитывать пространственные характеристики распространения эпидемии по стране. Во втором варианте используются только ретроспективные данные, доступные на момент прогнозирования для данного города. Преимуществом предложенного подхода является отсутствие необходимости использовать дополнительные, не всегда доступные внешние данные для прогнозирования эпидемии. Результаты тестовых расчетов показали, что первый метод показывает хорошие результаты при значительных задержках между пиками эпидемий в разных городах. В случае если эпидемия в Санкт-Петербурге началась вскоре после регистрации первых эпидемических вспышек в других городах Российской Федерации, второй метод показывает сопоставимые результаты с точностью до 90%, что позволяет использовать результаты расчетов для планирования противовирусных мероприятий. Заблаговременность предсказания пиков пока остается на относительно низком уровне, что, по-видимому, связано с разнообразием шаблонов распространения вируса и постоянными изменениями транспортных связей внутри страны.

**Ключевые слова**

популяционные модели, прогнозирование, математическая эпидемиология, грипп, калибровка

**Благодарности**

Статья подготовлена по результатам работы, выполненной при поддержке Российского научного фонда (соглашение № 14–21–00137 от 15.08.2014 г.)

## PREDICTION OF FLU EPIDEMIC PEAKS IN ST. PETERSBURG THROUGH POPULATION-BASED MATHEMATICAL MODELS

V.N. Leonenko<sup>а</sup>, Yu.K. Novoselova<sup>а</sup>, K.M. Ong<sup>б</sup><sup>а</sup> ITMO University, Saint Petersburg, 197101, Russian Federation<sup>б</sup> New York University School of Medicine, New York, 10016, the USACorresponding author: [sergei.v.ivanov@gmail.com](mailto:sergei.v.ivanov@gmail.com)**Article info**

Received 27.09.16, accepted 30.10.16

doi: 10.17586/2226-1494-2016-16-6-1145-1148

Article in Russian

**For citation:** Leonenko V.N., Novoselova Yu.K., Ong K.M. Prediction of flu epidemic peaks in st. petersburg through population-based mathematical models. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2016, vol. 16, no. 6, pp. 1145–1148. doi: 10.17586/2226-1494-2016-16-6-1145-1148

### Abstract

The paper presents two methods of predicting the peak of influenza epidemics using population-based mathematical models: Baroyan-Rvachev and modified Kermack-McKendrick model, proposed by the authors. We perform the comparison of the prediction accuracy of time and the value of epidemics peaks on long-term data of ARI incidence in the city of St. Petersburg. The methodology of comparison is based on three criteria of accuracy conventionally named as "square", "vertical stripe" and "horizontal stripe", and two variants of the model parameters estimation. In the first variant we calibrate the model on the data of the first city impacted by the epidemic, and use these parameters in the future for the other cities, that allows taking into account the spatial characteristics of the epidemic in the country. In the second case, we only use historical data available at the time of the prediction for a given city. The advantage of this approach is the lack of need for additional, not always available, external data to predict the epidemic. The results of test calculations have demonstrated that the first method shows good results in the case of significant delays between the peaks of epidemics in different cities. If the outbreak in St. Petersburg started soon after the registration of the first outbreaks in the other cities of the Russian Federation, the second method shows comparable results to an accuracy of 90% to predict the peak of the epidemic. In most cases, it is sufficient for the use of the results of calculations for planning antiviral activities. The lead time of the peak prediction is still at a relatively low level, that seems to be associated with a variety of patterns of virus spread and permanent changes in transport communications within the country.

### Keywords

population-based models, prediction, mathematical epidemiology, influenza, calibration

### Acknowledgements

This paper is financially supported by the Russian Scientific Foundation, Agreement No.14-21-00137 (15.08.2014)

Острые респираторные заболевания, или ОРВИ – одни из самых известных и распространенных инфекционных заболеваний человека. Самое опасное из них, грипп, обладает способностью вызывать повторяющиеся эпидемические вспышки, во время которых суммарная заболеваемость (грипп + фоновые ОРВИ) значительно превосходит средний сезонный уровень. В мировом масштабе вспышки гриппа приводят к 3–5 млн заболеваний в тяжелой форме ежегодно. Смертность от последствий гриппа, включающих в себя, в частности, инфаркты и инсульты [1], составляет 250–500 тыс. индивидов в год [2]. Даже во время эпидемий гриппа только 15–20% случаев заболеваемости ОРВИ можно отнести собственно к гриппу [3], и поставить точный диагноз, отличив грипп от других заболеваний верхних дыхательных путей со схожими симптомами, можно только путем лабораторного обследования [4]. Для планирования мер сдерживания заболеваемости гриппом исследователями предлагаются различные методы, основанные на применении математической эпидемиологии. В конце 60-х г.г. XX века было создано несколько математических моделей для прогнозирования эпидемических вспышек гриппа. Одно из наиболее совершенных исследований по данной тематике, выполненное О.В. Барояном и Л.А. Рвачевым, заключалось в исследовании распространения вируса гриппа по городам Советского Союза [5]. Модель Барояна–Рвачева была комбинацией классической популяционной *SEIR*-модели Кермака–Маккендрика [6] и линейной модели междугородних миграционных потоков. Несмотря на то, что структура модели сама по себе не была новой, с ее помощью удалось получить достаточно точные прогнозы моментов старта эпидемического процесса и моментов его пика в городах Советского Союза в 70-х г.г. XX века. К примеру, день старта эпидемиологического процесса предсказывался без ошибок в 56,1% случаев, а с ошибкой менее одной недели – в 92,2% случаев. Аналогичные показатели для предсказания дня пика эпидемиологической вспышки составили 53% и 87,4% соответственно [7]. В начале 80-х г.г. модельный комплекс, созданный в НИИ гриппа и основанный на модели Барояна–Рвачева, продемонстрировал признаки растущего несоответствия между модельными прогнозами и реально наблюдаемой эпидемической картиной. Дальнейшие события, связанные в первую очередь с распадом СССР, привели к полной остановке работ в данном направлении. В результате в течение последних тридцати лет систематических исследований по прогнозированию гриппа в городах Российской Федерации на основе моделирования не производилось. Целью настоящей работы является моделирование эпидемических вспышек с помощью популяционных *SEIR*-моделей и оценка точности полученных предсказаний эпидемических пиков по ретроспективным данным на примере города Санкт-Петербурга. Полученные результаты позволяют оценить эффективность методов прогнозирования и предложить варианты их улучшения.

В настоящей работе авторы поставили перед собой задачу оценить эффективность предсказания ожидаемого дня эпидемического пика в Санкт-Петербурге и его величины (в количестве заболевших за единицу времени) в условиях начавшегося эпидемического сезона. При этом использование наработок по модельному комплексу Барояна–Рвачева в значительной степени затруднено по двум основным причинам:

1. отсутствие современных данных по транспортным потокам, использованных в оригинальной модели;
2. низкая детализация данных заболеваемости, имеющихся у авторов (еженедельная заболеваемость вместо ежедневной).

С учетом приведенных ограничений в работе авторами предлагаются два метода предсказания эпидемических пиков на основе математического моделирования:

1. предсказание пиков в заданном городе с использованием популяционной модели, настроенной на данных других городов, с определением начала эпидемического процесса из данных сезонной заболеваемости;
2. предсказание пиков в заданном городе с помощью калибровки популяционной модели на неполных эпидемических данных.

Данные, использованные для расчетов, были предоставлены ФГБУ НИИ гриппа. Они содержат недельную заболеваемость всеми видами ОРВИ (включая грипп) в Санкт-Петербурге с 1986 по 2014 год. Перед калибровкой моделей и построением прогнозов исходные данные были обработаны следующим образом: исправлен систематический недоучет больных в праздничные дни [5]; с помощью интерполяции из еженедельных данных получены ежедневные; из сезонной заболеваемости ОРВИ выделена эпидемическая заболеваемость с использованием регрессионного анализа [8]. В результате получено 24 набора данных эпидемической заболеваемости за указанные годы (с учетом того факта, что в некоторые годы эпидемических всплесков в Санкт-Петербурге не наблюдалось).

Для калибровки и предсказания пиков использовалась модификация классической популяционной модели Кермака–Маккендрика, заданная системой обыкновенных дифференциальных уравнений. В данной модели рассматриваемая популяция делится на четыре группы ( $S$ ,  $E$ ,  $I$ ,  $R$ ), численности которых выражаются долями исходной популяции:  $S$  – доля восприимчивых индивидов;  $E$  – доля индивидов в инкубационном периоде;  $I$  – доля инфекционных индивидов;  $R$  – доля индивидов, иммунных к гриппу. Следуя [5, 7], мы полагаем, что некоторая доля населения имеет иммунитет к гриппу изначально.

При калибровке модели варьированием параметров модели и смещением получаемой модельной кривой заболеваемости относительно эпидемических данных по вертикальной и горизонтальной оси достигается наилучшее приближение модельных данных к реальным. Качество приближения характеризуется суммой квадратов евклидовых расстояний между реальным и модельным показателями заболеваемости в каждый день эпидемии. Подробное описание алгоритма калибровки приведено авторами в [8]. Калибровочный алгоритм был реализован в виде совокупности программ на языке Python 3.x с использованием библиотек SciPy, Matplotlib и NumPy.optimize. Для оценки качества прогнозов параметров эпидемического пика использовались критерии, применявшиеся ранее в модельном комплексе Барояна–Рвачева [7], а именно:

- «квадрат». Предсказание удовлетворяет критерию точности, если  $dt \in -8..8, dh \in (0,5; 2,0)$ ;
- «вертикальная полоса».  $dt \in -7..7$ ;
- «горизонтальная полоса».  $dh \in (0,7; 1,5)$ ,

где  $dt$  – абсолютная ошибка предсказания момента пика (в днях), а  $dh$  – относительная ошибка предсказания величины пика.

Для предсказательного моделирования на основе данных других городов модель, откалиброванная на данных первого города, охваченного эпидемией, используется для других городов. Использование лишь популяционной эпидемической модели, без транспортной модели, приводит к необходимости дожидаться начала наблюдаемой эпидемии в городе, перед тем как прогнозировать параметры пика. По этой причине прогноз будет дан за  $t_{peak} - t_{start}$  дней до его наступления. Здесь  $t_{start}$  – день объявления эпидемии местными органами здравоохранения (обычно 1-я–2-я неделя со дня старта эпидемии),  $t_{peak}$  – реальный день наступления пика. Таким образом, прогноз может быть дан за 1–8 недель до пика, в зависимости от эпидемического сезона. Эксперимент показал, что 90,5% прогнозов, полученных с помощью метода, удовлетворяют критерию «горизонтальная полоса», связанного с оценкой высоты пика. При этом день наступления пика предсказывается неудовлетворительно – 31,8% прогнозов удовлетворяют критерию «вертикальная полоса». Из-за этого общая точность предсказания также остается низкой (33,3% прогнозов удовлетворяют критерию «квадрат»).

Предсказательное моделирование на основе неполных данных заболеваемости в целом аналогично процедуре калибровки популяционной модели на ретроспективных данных: также находится модельная кривая, наилучшим образом соответствующая имеющимся данным реальной заболеваемости, и оцениваются значения параметров. Главное отличие заключается в том, что вместо полных данных об эпидемической вспышке используется лишь их часть, доступная на момент прогнозирования. В связи с этим возникает проблема недостатка данных: через небольшое число реальных точек эпидемической кривой можно провести достаточно большое число модельных кривых с весьма отличающимися расположениями пиков. В результате должен искажаться компромисс между точностью прогноза и его заблаговременностью. Для получения прогнозов SEIR-модель настраивалась на неполные данные по эпидемической заболеваемости в Санкт-Петербурге разного размера, начиная с пяти доступных значений и далее, до значения, совпадающего с эпидемическим пиком. Полученная эффективность предсказаний сравнивалась с таковой, полученной предыдущим методом, и измерялось время до наступления пика, в течение которого объема эпидемических данных становилось достаточно, чтобы достичь аналогичной эффективности. Точность предсказания пиков на неполных данных достигает эффективности предсказания при калибровке модели на данных других городов только лишь при используемом объеме данных, соответствующем

шему дню  $t_{peak}-3$ , что, безусловно, слишком поздно для использования прогноза органами здравоохранения. При этом точность предсказания дня эпидемического пика по-прежнему остается неудовлетворительной. Тем не менее, метод может быть использован для оценки высоты эпидемического пика с удовлетворительной точностью, в условиях, когда предыдущий метод не применим. Например, расчетной вероятности удовлетворения критерию точности «горизонтальная полоса» в 80% можно добиться при прогнозировании за неделю до эпидемического пика.

### Литература

1. Flu and Heart Disease & Stroke [Электронный ресурс]. Режим доступа: [www.cdc.gov/flu/heartdisease](http://www.cdc.gov/flu/heartdisease), своб. (дата обращения 20.09.16).
2. WHO. Influenza (Seasonal) [Электронный ресурс]. 2014. Режим доступа: [www.who.int/mediacentre/factsheets/fs211/en](http://www.who.int/mediacentre/factsheets/fs211/en), своб. (дата обращения 20.09.16).
3. Романоха А.А., Санникова Т.Е., Дрынов И.Д. Возникновение эпидемий острых респираторных заболеваний // Вестник РАН. 2011. Т. 81. № 2. 122–126.
4. Influenza Signs and Symptoms and the Role of Laboratory Diagnostics [Электронный ресурс]. Режим доступа: [www.cdc.gov/flu/professionals/diagnosis/labrolesprocedures.htm](http://www.cdc.gov/flu/professionals/diagnosis/labrolesprocedures.htm), своб. (дата обращения 20.09.16).
5. Бароян О.В., Рвачев Л.А., Иванников Ю.Г. Моделирование и прогнозирование эпидемий гриппа для территории СССР. М.: ИЭМ им. Н.Ф.Гамалеи, 1977. 546 с.
6. Anderson R.M., May R.M., Anderson B. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, 1992. 768 p.
7. Иванников Ю.Г., Исмагулов А.Т. Эпидемиология гриппа. Алма-Ата, 1983. 204 с.
8. Leonenko V.N., Ivanov S.V., Novoselova Yu.K. A computational approach to investigate patterns of acute respiratory illness dynamics in the regions with distinct seasonal climate transitions // *Procedia Computer Science*. 2016. V. 80. P. 2402–2412. doi: 10.1016/j.procs.2016.05.538

### References

1. Flu and Heart Disease & Stroke. Available at: [www.cdc.gov/flu/heartdisease](http://www.cdc.gov/flu/heartdisease) (accessed 20.09.16).
2. WHO. Influenza (Seasonal). Available at: [www.who.int/mediacentre/factsheets/fs211/en](http://www.who.int/mediacentre/factsheets/fs211/en) (accessed 20.09.16).
3. Romanyukha A.A., Sannikova T.E., Drynov I.D. The origin of acute respiratory epidemics. *Herald of the Russian Academy of Sciences*, 2011, vol. 81, no. 2, pp. 122–126. doi: 10.1134/S1019331611010114
4. Influenza Signs and Symptoms and the Role of Laboratory Diagnostics. Available at: [www.cdc.gov/flu/professionals/diagnosis/labrolesprocedures.htm](http://www.cdc.gov/flu/professionals/diagnosis/labrolesprocedures.htm) (accessed 20.09.16).
5. Baroyan O.V., Rvachev L.A., Ivannikov Yu.G. *Modelling and Forecasting of the Influenza Epidemics for USSR Territory*. Moscow, 1977, 546 p.
6. Anderson R.M., May R.M., Anderson B. *Infectious Diseases of Humans: Dynamics and Control*. Oxford, 1992, 768 p.
7. Ivannikov Yu.G., Ismagulov A.T. *Epidemiologiya Grippa [Influenza Epidemiology]*. Alma-Ata, 1983, 204 p.
8. Leonenko V.N., Ivanov S.V., Novoselova Yu.K. A computational approach to investigate patterns of acute respiratory illness dynamics in the regions with distinct seasonal climate transitions. *Procedia Computer Science*, 2016, vol. 80, pp. 2402–2412. doi: 10.1016/j.procs.2016.05.538

### Авторы

**Леоненко Василий Николаевич** – кандидат физико-математических наук, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [VNLeonenko@yandex.ru](mailto:VNLeonenko@yandex.ru)

**Новоселова Юлия Константиновна** – студент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, [juliaeternity@gmail.com](mailto:juliaeternity@gmail.com)

**Онг Карен Мелинда** – PhD, аспирант, Школа медицины Нью-Йоркского университета, Нью-Йорк, 10016, США, [Karen.Ong@med.nyu.edu](mailto:Karen.Ong@med.nyu.edu)

### Authors

**Vasily N. Leonenko** – PhD, engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, [VNLeonenko@yandex.ru](mailto:VNLeonenko@yandex.ru)

**Yulia K. Novoselova** – student, ITMO University, Saint Petersburg, 197101, Russian Federation, [juliaeternity@gmail.com](mailto:juliaeternity@gmail.com)

**Karen Melinda Ong** – PhD, MD–PhD student, New York University School of Medicine, New York, 10016, the USA, [Karen.Ong@med.nyu.edu](mailto:Karen.Ong@med.nyu.edu)