PROBLEMS OF VIROLOGY (VOPROSY VIRUSOLOGII). 2023; 68(3)
https://doi.org/10.36233/0507-4088-179
TO HELP THE VIROLOGIST

ORIGINAL STUDY ARTICLE
DOI: https://doi.org/10.36233/0507-4088-179

Mathematical model for assessing the level of cross-immunity between strains of influenza virus subtype H_3N_2

Marina N. Asatryan[1], Boris I. Timofeev[1], Ilya S. Shmyr[1], Karlen R. Khachatryan[2], Dmitrii N. Shcherbinin[1], Tatiana A. Timofeeva[1], Elita R. Gerasimuk[3], Vaagn G. Agasaryan[1], Ivan F. Ershov[1], Tatyana I. Shashkova[4], Nikita V. Ivanisenko[4], Olga L. Kardymon[4], Tatyana A. Semenenko[1], Boris S. Naroditsky[1], Denis Yu. Logunov[1], Aleksander L. Gintsburg[1]

[1]National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, 123098, Moscow, Russia;
[2]National Research University Higher School of Economics, 123458, Moscow, Russia;
[3]State University "Dubna", 141982, Dubna, Russia;
[4]Artificial Intelligence Research Institute, 121170, Moscow, Russia

Introduction. The WHO regularly updates influenza vaccine recommendations to maximize their match with circulating strains. Nevertheless, the effectiveness of the influenza A vaccine, specifically its H_3N_2 component, has been low for several seasons.
The aim of the study is to develop a mathematical model of cross-immunity based on the array of published WHO hemagglutination inhibition assay (HAI) data.
Materials and methods. In this study, a mathematical model was proposed, based on finding, using regression analysis, the dependence of HAI titers on substitutions in antigenic sites of sequences. The computer program we developed can process data (GISAID, NCBI, etc.) and create "real-time" databases according to the set tasks.
Results. Based on our research, an additional antigenic site F was identified. The difference in 1.6 times the adjusted R^2, on subsets of viruses grown in cell culture and grown in chicken embryos, demonstrates the validity of our decision to divide the original data array by passage histories. We have introduced the concept of a degree of homology between two arbitrary strains, which takes the value of a function depending on the Hamming distance, and it has been shown that the regression results significantly depend on the choice of function. The provided analysis showed that the most significant antigenic sites are A, B, and E. The obtained results on predicted HAI titers showed a good enough result, comparable to similar work by our colleagues.
Conclusion. The proposed method could serve as a useful tool for future forecasts, with further study to confirm its sustainability.

Keywords: *influenza virus; subtype H_3N_2; HAI titers; cross-immunity; antigenic distance; antigenic site; Hamming distance; regression analysis; epidemiological model; immune landscape; vaccine strain*

For citation: Asatryan M.N., Timofeev B.I., Shmyr I.S., Khachatryan K.R., Shcherbinin D.N., Timofeeva T.A., Gerasimuk E.R., Agasaryan V.G., Ershov I.F., Shashkova T.I., Ivanisenko N.V., Kardymon O.L., Semenenko T.A., Naroditsky B.S., Logunov D.Yu., Gintsburg A.L. Mathematical model for assessing the level of cross-immunity between strains of influenza virus subtype H_3N_2. *Problems of Virology* (*Voprosy Virusologii*). 2023; 68(3): 252-264. DOI: https://doi.org/10.36233/0507-4088-179 EDN: https://elibrary.ru/rexvea

For correspondence: Marina N. Asatryan, PhD (Med.), Senior Researcher, Epidemiological Cybernetics Group of the Epidemiology Department, National Research Center for Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya, 123098, Moscow, Russia. E-mail: masatryan@gamaleya.org

Information about the authors:
Marina N. Asatryan, https://orcid.org/0000-0001-6273-8615
Boris I. Timofeev, https://orcid.org/0000-0001-7425-0457
Ilya S. Shmyr, https://orcid.org/0000-0002-8514-5174
Karlen R. Khachatryan, https://orcid.org/0000-0002-1934-532X
Dmitrii N. Shcherbinin, https://orcid.org/0000-0002-8518-1669
Tatiana A. Timofeeva, https://orcid.org/0000-0002-8991-8525
Elita R. Gerasimuk, https://orcid.org/0000-0002-7364-163X
Vaagn G. Agasaryan, https://orcid.org/0009-0009-3824-7061
Ivan F. Ershov, https://orcid.org/0000-0002-3333-5347.
Tatyana I. Shashkova, https://orcid.org/0000-0002-8754-8727
Nikita V. Ivanisenko, https://orcid.org/0000-0002-0333-8117
Olga L. Kardymon, https://orcid.org/0000-0002-4827-8891
Tatyana A. Semenenko, https://orcid.org/0000-0002-6686-9011
Boris S. Naroditsky, https://orcid.org/0000-0001-5522-8238
Denis Yu. Logunov, https://orcid.org/0000-0003-4035-6581
Aleksander L. Gintsburg, https://orcid.org/0000-0003-1769-5059

Contribution: the authors contributed equally to this article.
Conflict of interest. The authors declare no conflicts of interest.
Received 12 May 2023
Accepted 22 June 2023
Published 30 June 2023

ОРИГИНАЛЬНОЕ ИССЛЕДОВАНИЕ

DOI: https://doi.org/10.36233/0507-4088-179

Математическая модель для оценки уровня перекрёстного иммунитета между штаммами вируса гриппа подтипа H_3N_2

Асатрян М.Н.[1], Тимофеев Б.И.[1], Шмыр И.С.[1], Хачатрян К.Р.[2], Щербинин Д.Н.[1], Тимофеева Т.А.[1], Герасимук Э.Р.[3], Агасарян В.Г.[1], Ершов И.Ф.[1], Шашкова Т.И.[4], Иванисенко Н.В.[4], Кардымон О.Л.[4], Семененко Т.А.[1], Народицкий Б.С.[1], Логунов Д.Ю.[1], Гинцбург А.Л.[1]

[1]ФГБУ «Национальный исследовательский центр эпидемиологии и микробиологии имени почетного академика Н.Ф. Гамалеи», 123098, г. Москва, Россия;
[2]ФГАОУ ВО Национальный исследовательский университет «Высшая школа экономики», 123458, г. Москва, Россия;
[3]Государственный университет «Дубна», 141982, г. Дубна, Россия;
[4]Институт искусственного интеллекта, 121170, г. Москва, Россия

Введение. Всемирная организация здравоохранения (ВОЗ) регулярно обновляет рекомендации по вакцинам против гриппа с целью достижения их максимального соответствия очередным циркулирующим штаммам. Тем не менее на протяжении нескольких сезонов эффективность вакцины против гриппа А, а именно её компоненты H_3N_2, определялась как низкая.
Цель исследования – разработка математической модели перекрёстного иммунитета на основании имеющегося массива опубликованных ВОЗ данных реакции торможения гемагглютинации (РТГА).
Материалы и методы. В настоящей работе представлена математическая модель, основанная на нахождении с помощью регрессионного анализа зависимости титров РТГА от замен в антигенных сайтах последовательностей. Разработанная нами компьютерная программа имеет возможность обрабатывать данные (GISAID, NCBI и др.) и формировать в режиме реального времени базы данных согласно поставленным задачам.
Результаты. На основе наших исследований был вычленен дополнительный антигенный сайт F. Разница в 1,6 раза скорректированного R^2 на подмножествах вирусов, выращенных в культуре клеток и культивируемых в куриных эмбрионах, демонстрирует обоснованность нашего решения о разделении первоначального массива данных по пассажным историям. Нами введено понятие степени гомологичности между двумя произвольными штаммами, которая принимает значение функции, зависящей от дистанции Хэмминга, и показано, что результаты регрессии существенно зависят от выбора функции. Проведённый анализ показал, что наиболее значимыми антигенными сайтами являются A, B и E. Полученные результаты прогноза титров РТГА показали достаточно хороший результат, сопоставимый с аналогичными работами наших коллег.
Заключение. Предложенный метод может послужить хорошим инструментом для будущих прогнозов с дальнейшим изучением для подтверждения его устойчивости.

Ключевые слова: *вирус гриппа; подтип H_3N_2; титры РТГА; перекрёстный иммунитет; антигенное расстояние; антигенный сайт; дистанция Хэмминга; регрессионный анализ; эпидемиологическая модель; иммунный ландшафт; вакцинный штамм*

Для цитирования: Асатрян М.Н., Тимофеев Б.И., Шмыр И.С., Хачатрян К.Р., Щербинин Д.Н., Тимофеева Т.А., Герасимук Э.Р., Агасарян В.Г., Ершов И.Ф., Шашкова Т.И., Иванисенко Н.В., Кардымон О.Л., Семененко Т.А., Народицкий Б.С., Логунов Д.Ю., Гинцбург А.Л. Математическая модель для оценки уровня перекрёстного иммунитета между штаммами вируса гриппа подтипа H_3N_2. *Вопросы вирусологии*. 2023; 68(3): 252-264. DOI: https://doi.org/10.36233/0507-4088-179 EDN: https://elibrary.ru/rexvea

Для корреспонденции: Асатрян Марина Норайровна, канд. мед. наук, старший научный сотрудник группы эпидемиологической кибернетики отдела эпидемиологии ФГБУ «Национальный исследовательский центр эпидемиологии и микробиологии имени почетного академика Н.Ф. Гамалеи», 123098, г. Москва, Россия. E-mail: masatryan@gamaleya.org

Участие авторов: все авторы сделали эквивалентный вклад в подготовку публикации.
Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила 12.05.2023
Принята в печать 22.06.2023
Опубликована 30.06.2023

Introduction

The World Health Organization (WHO) Global Influenza Surveillance and Response Network (GISRS) monitors and analyzes the evolution and epidemiology of influenza viruses with the main goal of selecting a vaccine strain and improving this process through research aimed at a better understanding of evolutionary variability, distribution factors in combination with the immunological landscape of the population and cross-immunity [1]. At the same time, WHO regularly updates recommendations for influenza vaccines in order to achieve their maximum compliance with the next circulating strains.

Nevertheless, over several seasons, the effectiveness of the influenza A vaccine, namely its H_3N_2 component, was determined to be low compared to other strains [2–5]. The reason for the low efficiency can be several factors. For example, characteristic adaptive changes during passaging of the recommended strain in chicken embryos during the production of vaccines [6]. Since the development, large-scale production and distribution of a vaccine takes many months, eventually the prevalence of circulating strains in some seasons will change significantly by the coming season [7]. But even if the recommended vaccine is well matched against circulating strains, its effectiveness could be adversely affected by the existing immune landscape [8]. Therefore, prediction of the evolutionary variability of the influenza virus is still of great interest for public health [9–12].

The most promising direction in this area is the construction of computer models that can be used to combine various modeling approaches, use multiple data sources with the ability to interpret the results for recommendations when choosing a vaccine strain. This requires close cooperation between scientists from different fields and directions, working at all levels of epidemiological surveillance and selection of vaccine strains, as well as model developers, epidemiologists and clinicians [10, 13, 14].

The team of the National Research Center of Epidemiology and Microbiology named after Honorary Academician N.F. Gamaleya in 2020 developed and successfully registered (certificate of registration No. 2020617965 dated July 15, 2020) the ***computer program*** **Influenza IDE** – an epidemiological model (EM) across continents with a simplified model of cross-immunity and a constantly updated database (of various types and subtypes of the influenza virus) **Influenza DB.** An important feature of EM is the possibility of forming the zero immune landscape of the population and then, after step-by-step simulation of the spread of the influenza virus, obtaining an immune landscape on the first day of the next season. The epidemiological model of the spread of the influenza virus among the world population over several seasons, developed using the agent approach, is presented in the form of implemented models: population behavior model, model of infectious process and infection model (based on the immune response in the body of an individual agent (person), taking into account the immune memory and the cross-immunity model). The computer program is designed to integrate various cross-immunity models.

A significant number of works on the study of cross-immunity is based on the discovery of a relationship between hemagglutination-inhibition antibody (HAI) titers and differences in the genetic sequences of viruses. To find this relationship, researchers use various mathematical methods, including regression analysis [15–18]. At the same time, various functions from HAI titers are used as a measure of cross-immunity. So, in the work of F.M. Burnet and D. Lush [19] introduced a function designated as an indicator of vaccine effectiveness $R_{ij} = c_{ij} / c_{ii}$, where c_{ij} is the concentration (dilution) value of serum to virus i in HAI assay with virus j, c_{ii} is the concentration (dilution) value of serum i in HAI assay with original virus, where the concentration (dilution) of serum in HAI assay is the reciprocal of the titer of HAI assay. Researchers I. Archetti and F.L. Horsfall [20] introduced the geometric mean of the above ratios $(R_{ij}R_{ji})^{1/2}$ as a measure of antigenic variability in their work. In later works, which laid the foundation for the currently widely used antigenic cartography, A. Lapedes and R. Farber [21], D.J. Smith et al. [22] demonstrated that it is possible to construct a low-dimensional form space in which antibodies and antigens are considered as points, and the distance between them is denoted as antigenic distance. As a measure of distance D_{ij}, the logarithm of base 2 RTGA titers and the $log_2(R_{ij})$, $0{,}5 \times log_2(R_{ij}R_{ji})$ values were used [15–17].

As a measure of the difference between sequences, either a simple Hamming distance or a function depending on it is usually used. The most frequently used model in the considered works is the consolidation of amino acid positions into antigenic sites [23–25]. In such cases, the Hamming distance between antigenic sites is considered [16–18]. But there are also complex cases when physicochemical factors of differences in amino acid positions, for example, glycosylation, act as a measure of the difference between antigenic sites [18, 26]. As materials for building models, limited datasets were used, including only reference strains, which reduces objectivity. In this paper, we present a cross-immunity regression model based on the entire available array of WHO-published HAI data, which will improve the accuracy of the model, increase its objectivity and testability.

The use of an adequate epidemiological model with a verified cross-immunity model will improve the process of selecting the necessary vaccine strains for a more successful fight against the influenza virus.

The study purpose is to develop a mathematical model of cross-immunity based on the available array of seasonal serological testing data published by WHO (HAI assay).

Materials and methods

When forming the **Influenza DB** data array, information was used from WHO's published seasonal data on the results of HAI testing and data (sequences + accompanying information) from the GISAID (Global

Initiative on Sharing All Influenza Data) platform. The computer program **Influenza IDE** [27] also provides an additional module **Influenza DP**, which is a universal database processor (GISAID, NCBI, etc.).

To study patterns in the cross-immunity model, multiple regression analysis (linear regression) was used using the least squares method (LSM) and the non-negative LSM method [28] to estimate the regression parameters, represented by formula (1):

$$\textbf{Modified\_titre} = c0 + c1 \times As1 + c2 \times As2 + c3 \times As3 + c4 \times As4 + c5 \times As5 + c6 \times As6, \quad (1)$$

where the modified HAI titers are represented as the function value, as arguments As1, As2, As3, As3, As4, As5, As6 are the values of the degree of homology of antigenic sites (As) A, B, C, D, E, F, respectively, and c0, c1, c2, c3, c4, c5, c6 – parameters (coefficients) of the model.

For regression arguments, we relied on our own method for determining antigenic sites. The term "antigenic site" was introduced by Gerhard and Webster in 1978 to describe specific monoclonal antibodies. Antibodies that competed with each other were considered to bind the same antigenic site. Each antigenic site may contain one or more epitopes – different sets of amino acids on the antigen that come into contact with the amino acids of the antibodies. Competition between antibodies that bind the same antigenic site suggests that the epitopes at a given site physically overlap, but may be different, and one antibody molecule shields the entire antigenic site.

Also, to account for differences between strains by antigenic sites, we introduced the concept of **homology degree** and defined it for two randomly selected strains as a decreasing function of the Hamming distance, taking values from 1 to 0. The justification and selection of the functions used are described in detail in the section "Results". The Stats package version 4.0.3 of the R programming language was used to train the regression model. The Pandas package version 1.4.2 of the Python programming language was used for data preprocessing. The stability of the predictive ability of the cross-immunity model was analyzed on retrospective data. As a measure of the adequacy of the model, we used the coefficient of determination R_2 [29], and as a measure of the accuracy of the forecast – the index of reproducibility of titles for one (± 1) and (or) two (±2) dilutions. We relied on numerous studies in which, when comparing the results of HAI assay for each sample within one laboratory or between several, the titers were considered equivalent if they differed by no more than one dilution (i.e. by 2 times) [30–32].

Building an Influenza DB Dataset to Study a Cross-Immunity Model

Preparation of experimental HAI data

WHO reports on influenza have been published every season since 2005 and are in the public domain [33]. The presented antibody titers in HAI reveal the antigenic properties of reference and test strains based on the cross-reactivity of ferret antibodies against reference strains that inhibit the agglutination of guinea pig erythrocytes by the tested influenza viruses. Due to the fact that WHO in 2008-2009 found a noticeable effect of neuraminidase and various sources of erythrocytes (turkey, human, guinea pig) on the results of HAI assay, in further studies we used only HAI assay tables from 2009 to 2022 performed on guinea pig erythrocytes with the addition of 20nM Oseltamivir, which is used to exclude the influence of neuraminidase. Despite the proposed international coding of viruses, researchers from all over the world encounter inaccurate annotations when depositing sequences in public databases [34]. Therefore, a separate task in our work was the procedure for processing and preparing published HAI assay tables for further use. Also, as a result of the preparation, each sequence was assigned a unique identifier.

Modified HAI titers act as a measure of cross-immunity in our model. To do this, the values of the published HAI titers are converted to the logarithm to the base 2. Next, the data is averaged according to the algorithm: subsets are combined in accordance with similar pairs of strain identifiers (reference and test); in each obtained subset, the arithmetic mean of the modified titers is calculated for the same pairs of strain identifiers.

It should be noted that the cultivation (passage) of viruses performed in chicken embryos may cause characteristic adaptive substitutions that change the receptor specificity of viruses and the nature of glycosylation [6, 35] and, as a result, affect cross-reactivity in HAI data. Therefore, as a result of the analysis of the primary data, we decided to single out several subsets in the initial HAI data array:

– with passage history in chick embryos (Egg-Egg 4226);

– with passage history in cell culture (Cell-Cell 28 621);

– with mixed passage history (Egg_s-Cell_t 16 463): reference viruses (to obtain control antisera) cultured in chick embryos in combination with test viruses grown in cell culture;

– with mixed passage history (Cell_s-Egg_t 5032): reference viruses (to obtain control antisera) grown in cell culture in combination with test viruses cultured in chick embryos.

Our solution was based on numerous comparative studies of antigenic mutations of the H3N2 influenza virus during cultivation in chicken embryos and cell passages [36–38]. Considering that the amount of data with a passage history of Cell-Cell and Egg_s-Cell_t is several times larger than the dimension of subsets with passage histories of Egg-Egg and Cell_s-Egg_t, we performed regression analysis on the data of the first two subsets.

Preparing data from the GISAID platform

The GISAID international platform was launched in 2008 and has since offered a reliable mechanism for the exchange of all genetic and influenza-related data for researchers, scientists and healthcare professionals [39]. As in the case of HAI data, when depositing sequences (when manually entering data, etc.), the format of the downloaded data is often distorted and misclassified.

This, in turn, makes it difficult to analyze and process the available information to improve approaches (methods) to the selection of vaccine strains. Therefore, for further research using the computer program **Influenza IDE** and the universal processor **Influenza DP** built into it, we carried out a thorough cleaning and reconciliation of the available data. After the completion of the data verification process, the amino acid sequences were aligned to the corresponding reference sequence: H_3N_2: A/Aichi/2/1968, 566 AA, including the signal peptide, using the original fast lexical algorithm (maximum amino acid match between the reference sequence and the sample with minimization of the number of positions deletions and insertions) or the Smith–Waterman method [40]. Samples with differences of more than 20% from the reference were discarded. Brief statistical characteristics of the data are presented in **Table 1**.

Results

Determination of antigenic sites of H_3N_2 influenza virus hemagglutinin

The combination of amino acid sequences into antigenic sites was carried out according to our own method with the inclusion of an additional antigenic site F. To determine which antigenic site of the influenza virus a particular position belongs to, we primarily focused on experimental studies [24]. Subsequently, we expanded the number of positions in antigenic sites by examining the variability of each hemagglutinin position starting from 1968. In the array of sequences obtained at the first stage, the occurrence of each amino acid in each position was counted. In parallel, an analysis was carried out for the ability of one or another amino acid of hemagglutinin to interact with antibodies, determining the immersion or, conversely, the exposure of amino acids on the surface of hemagglutinin. During the work, the GETAREA software [41] was used, the input of which was the tertiary structures of hemagglutinin molecules from the PDB (Protein Data Bank) database [42]. As a result, based on the variability of positions, their exposure in the tertiary structure of the protein on the surface, and also taking into account their maximum proximity to existing antigenic sites, variable exposed positions were determined, which were assigned to existing antigenic sites. Moreover, in addition to the well-known antigenic sites A, B, C, D, and E, we identified another antigenic site, respectively, named F. This antigenic site consists of many similar epitopes in the stem part of the hemagglutinin molecule. The database of tertiary structures contains more than 20 hemagglutinin complexes with monoclonal antibodies to this site. This site is described in detail in a study conducted by D.N. Shcherbinin et al. [43].

Thus, 6 antigenic sites were identified for hemagglutinin subtype H3. Below is a list of amino acid positions that make up these antigenic sites, a total of 109 positions (**Table 2**). Numbering is given by mature hemagglutinin H_3 [44]. It should be noted that the antigenic site F, unlike sites A–E, is located mainly in the HA_2 subunit of hemagglutinin.

The information obtained was uploaded to the **Influenza DB** database and served as a template for combining the existing array with aligned amino acid sequences into antigenic sites. Further, the array with antigenic sequences was expanded by adding information about the content of the number of amino acid substitutions between any two strains in each of the 6 antigenic sites. As a result of the above actions, we obtained for each of the 6 antigenic sites for randomly selected or separate (each) pair of strains both the value of HAI titers and the number of amino acid substitutions (Hamming distance). At the same time, substitutions in antigenic sites, regardless of the type and specific position, are accepted to be equivalent. In our model, sequences that do not have substitutions in antigenic sites, but have changes in other positions, were considered antigenically identical.

Development of a mathematical model of cross-immunity of the influenza virus

Analysis and selection of a function to assess the degree of homology

The subsets with the largest number of observations were selected for regression analysis: Cell-Cell 28,621 and Egg_s-Cell_t 16,463. Also, instead of a simple Hamming distance, we decided to use a function from it, thereby introducing the concept of the degree of homology between two arbitrary antigenic sequences, indicating how close the antigenic sites of two different strains are to each other in terms of antigenic properties, and taking the value of this function. It was necessary to choose a function that could meet the requirements and would not contradict biological processes: in a situation of complete (absolute) homology of two strains (in the absence of substitutions in antigenic sites), the function takes the value 1, and in the absence of homology (changes were noted in each amino acid position included in the antigenic site) the value of the function must be equal to 0 (or close to 0). Obviously, the function should decrease with increasing Hamming distance.

In addition, as shown in [45], the function that depends on the Hamming distance and describes the cross-reactivity of antigens is concave. Based on this, we introduced an additional restriction for the function we are considering, namely, that it is concave or at least not convex, i.e. each subsequent substitution contributes no more to the decrease in homology than the previous one.

According to the results of studies by colleagues [16, 45], which showed that when the number of amino acid substitutions in the antigenic site is equal to or greater than 7, cross-reactivity of recognizing antibodies between strains is practically not observed, we introduced an additional requirement for the desired function: for 7 substitutions, the value function should take such a small value that it can be neglected. We also used the function $1 - x / 8$ as the top-limiting function under consideration (for the values $x = \{1; 7\}$), and we took the value $1 - 7 / 8 = 0.125$ as the significance threshold for 7 substitutions.

Taking into account the specified requirements, the following functions were considered **(Figure)**:

Table 1. Data statistical characteristic
Таблица 1. Статистическая характеристика данных

Parameter Показатель		Value Значение	
		Cell-Cell	Egg_s-Cell_t
Total number of observations Общее количество наблюдений		28 621	16 463
Total number of pairs of strains Общее количество пар штаммов		21 394	12 444
Number of observations in the reporting period Количество наблюдений в отчетном периоде	2009	21	35
	2010	46	94
	2011	132	461
	2012	2142	1761
	2013	1738	2447
	2014	2891	2201
	2015	1907	1520
	2016	1209	1003
	2017	1948	1279
	2018	659	164
	2019	2236	950
	2020	2496	1064
	2021	902	281
	2022	10294	3203
Titer value, Me [Q_1; Q_3] Значение титра		160 [40; 320]	160 [40; 320]
Distribution by titer values Распределение по значениям титра	20	3404 (11,9%)	1858 (11,3%)
	40	4154 (14,5%)	2514 (15,3%)
	80	4232 (14,8%)	3852 (23,4%)
	160	6328 (22,1%)	3794 (23%)
	320	6170 (21,6%)	2446 (14,9%)
	640	3491 (12,2%)	1410 (8,6%)
	1280	772 (2,7%)	511 (3,1%)
	2560	60 (0,2%)	75 (0,5%)
	5120	10 (< 0,1%)	3 (< 0,1%)
Hamming distance across antigenic sites Дистанция Хэмминга по антигенным сайтам	A	2 [0; 4]	2 [0; 3]
	B	2 [1; 4]	3 [2; 4]
	C	1 [0; 2]	1 [0; 2]
	D	1 [0; 1]	1 [1; 2]
	E	1 [0; 2]	1 [0; 1]
	F	1 [0; 2]	1 [0; 2]

- $1 – x / 8$;
- $\exp(–a \times x)$, where a = 1/3; 1/2; 1;
- $1 / (a \times x + 1)$, where a = 1; 2; 3.

The values of HAI titers indicated in the original WHO tables as < (<40) in our model were replaced by a value of 20. We take a fixed value of the smallest value of the modified titer as the c0 coefficient, i.e. $\log2(20) \approx 4.322$.

The left part of **Table 3** presents the results for a subset with a passage history on cell culture (Cell-Cell); in the right part of the table – with a mixed passage history (Egg_s-Cell_t).

Consideration of the selected functions was carried out by cross-validation: the entire amount of data was randomly divided 5 times in the ratio of 80% to 20%, where the first part was taken as a training sample, and the second – for a test one. The results were evaluated by the highest average value of the adjusted R^2 obtained for the test samples. We also ranked the results for all functions.

Table 2. Amino acid positions at antigenic sites
Таблица 2. Аминокислотные позиции в антигенных сайтах

Site Сайт	Amino acid position numbers Номера аминокислотных позиций
A	121, 122, 124, 126, 131, 133, 135, 137, 138, 140, 142, 143, 144, 145, 146
B	128, 155, 156, 157, 158, 159, 160, 163, 186, 188, 189, 190, 192, 193, 194, 196, 197, 198, 199
C	45, 47, 48, 49, 50, 53, 54, 271, 273, 275, 276, 278, 280, 304, 311, 312
D	171, 172, 173, 201, 202, 207, 208, 212, 213, 214, 217, 219, 220, 221, 222, 223, 225, 226, 227, 229, 230, 242, 244, 246, 248
E	57, 58, 62, 63, 75, 78, 82, 83, 88, 92, 94, 96, 260, 261, 262
F	25, 31, 33, 326, 347, 361, 375, 384, 386, 450, 452, 453, 476, 479, 484, 487, 489, 501, 505

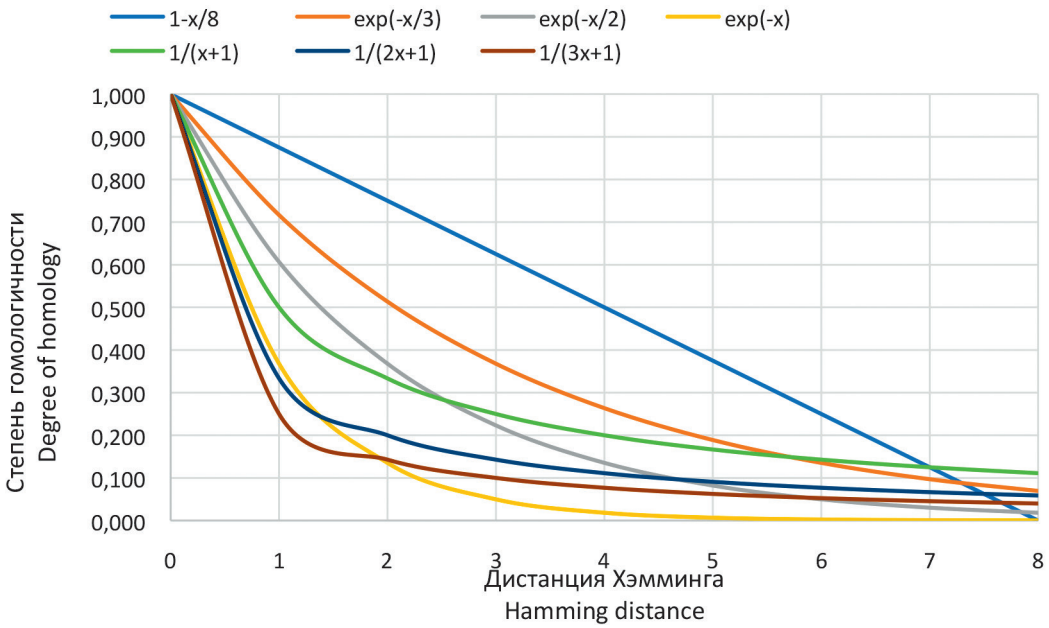

Figure. Functions for evaluating the degree of homology.
Рисунок. Функции для оценки степени гомологичности.

When comparing the results between subsets of Cell-Cell and Egg_s-Cell_t, almost complete correspondence of ranks (Spearman's rank correlation coefficient ρ = 0.96) of functions was observed in relation to subsets. Meanwhile, when comparing the values of the coefficients of determination for two subsets, the corresponding highest values were noted for the Cell-Cell subset (by 1.6 times), which confirms the expediency of separating the available data by passages, and also confirms that passages on chicken embryos contain characteristic adaptive substitutions, changing the receptor specificity of viruses and the nature of glycosylation [6, 35] and, as a result, affect cross-reactivity in HAI assay. Therefore, we carried out further calculations on the Cell-Cell subset. As a function describing the degree of homology, exp **(–x / 3)** was chosen, which showed the best result for the Cell-Cell subset and the best result for the Egg_s-Cell_t subset.

The results of regression analysis with the Cell-Cell subset (21,580 observations after averaging over pairs of strains) using the selected function with the requirement that the coefficients be non-negative are presented in **Table. 4**. In the first column of the table, in addition to the corrected R^2 already mentioned above, the designations of the corresponding antigenic sites are presented, then in the Estimate column the values of the desired C1–6 regression coefficients are given, in the third column, Standard error, the standard deviations of these coefficients are listed. The line Standard deviation represents the standard deviation of the dependent variable (modified titers). The contribution of each antigenic site can be judged by the value of its coefficient. The greater the value of the coefficient at a certain antigenic site, the greater its contribution to cross-immunity. After determining the coefficients of the regression model, the statistical hypothesis is tested about the equality to zero of the true values of the coefficients according to Student's criterion with a significance level of 0.05. The results of the test are presented in the fifth column, Pr(> |t|) p-value, the probability that the t-statistic is greater than the t-value modulus (ratio of coefficient values to their standard deviation). If this value is less than the confidence level of 0.05, the hypothesis is rejected and the parameter is considered significant. In

Table 3. R^2 values by passage histories for the considered functions

Таблица 3. Значения R^2 по пассажным историям для рассматриваемых функций

Function Функция	Cell-Cell (21 580 pairs of strains 2009–2022) (21 580 пар штаммов 2009–2022)						Egg_s-Cell_t (12 444 pairs of strains 2009–2022) (12 444 пары штаммов 2009–2022)					
	1_20%	2_20%	3_20%	4_20%	5_20%	ср._20%	1_20%	2_20%	3_20%	4_20%	5_20%	ср._20%
1 – x / 8	0,457	0,483	0,471	0,486	0,490	0,477	0,273	0,252	0,277	0,282	0,255	0,268
exp(–x / 3)	0,494	0,519	0,504	0,520	0,530	0,513	0,326	0,293	0,317	0,336	0,295	0,313
exp(–x / 2)	0,476	0,498	0,478	0,494	0,509	0,491	0,331	0,294	0,313	0,328	0,293	0,312
exp(–x)	0,366	0,380	0,346	0,363	0,393	0,370	0,253	0,205	0,200	0,242	0,207	0,222
1 / (x + 1)	0,441	0,461	0,438	0,453	0,471	0,453	0,312	0,272	0,283	0,315	0,275	0,292
1 / (2x + 1)	0,376	0,391	0,358	0,375	0,399	0,380	0,268	0,220	0,216	0,258	0,224	0,237
1 / (3x + 1)	0,328	0,340	0,301	0,320	0,348	0,327	0,213	0,162	0,149	0,196	0,168	0,178

Table 4. Regression results for the exp(–x / 3) function

Таблица 4. Результаты регрессии для функции exp(–x / 3)

Antigenic sites Антигенные сайты	Cell-Cell_2009–2022 (21 580 pairs of strains) (21 580 пар штаммов)		
	estimate значение	standard error стандартная ошибка	P-r(> |t|) p-value
A	1,233	0,034	<0,001
B	2,506	0,034	<0,001
C	0,000		
D	0,083	0,027	0,002
E	0,622	0,039	<0,001
F	0,000		
CR^2	0,514		
Standard deviation Стандартное отклонение	1,161		

our case, all antigenic sites, except for C and F, were recognized as significant. The coefficients at the antigenic site C and F took the value 0, and this is equivalent to insignificance.

Analysis of the stability of the predictive ability of the cross-immunity model based on retrospective data

For a more detailed understanding of the influence of the time factor and the number of observations involved in tuning the model on the accuracy of the forecast, the next step is to divide the subset of Cell-Cell_2009–2022 into different time periods (1-, 2-, 3-,4-, 5-year) adjustable periods with a forecast for each subsequent year. The full results of the regression analysis are presented in **Appendix**.

Based on presented in **table 5.1** and **5.2** of the results, we can state that the adjusted R2, which takes a value not higher than 0.3 in adjustable periods, begins to increase significantly from 2019, changing from 0.32 to 0.64 (the more years in the adjustable interval, the lower the value). 2021 is out of this trend due to low incidence and little data due to the COVID-19 pandemic. The dynamics of the value of the adjusted R^2 is also directly reflected in the forecast periods, taking values no higher than 0.4 (including negative ones) and further increasing from 2020 (from 0.47 to 0.6).

We also evaluated the predictive ability of the results presented in Table. 5.1 and 5.2, according to the reproducibility of titers for one (±1) and two (±2) dilutions. When comparing the calculated titles with those published in the forecast periods, with the exception of 2016 and 2021 (a small number of observations – 841 and 478, respectively) and 2015 (when the recommended vaccine strains failed to provide effective protection), the reproducibility of titers per (±1) dilution was demonstrated in more than 50% of cases, and at times reached 67% of cases. As for the reproducibility of titers at two (±2) dilutions, approximately 90% of cases and above are consistently observed over all time periods.

The results of regression analysis over adjustable (1-5year time intervals) demonstrated the significance of the contribution of each antigenic site to the immune response. The coefficients at antigenic sites A and B have high values in all forecast periods. This can be seen especially clearly on 3–5-year customizable periods. Antigenic site C actually takes on null values only with

Table 5.1. R^2
Таблица 5.1. R^2

Year Год	1 year год	2 year года	3 year года	4 year года	5 year лет
2014	0,135	0,099	0,099	0,101	0,101
2015	–0,223	–0,204	–0,354	–0,355	–0,348
2016	0,043	–0,225	–0,278	–0,540	–0,546
2017	0,005	0,198	0,158	0,083	0,071
2018	0,092	0,001	0,100	0,071	0,052
2019	0,124	0,244	0,405	0,318	0,277
2020	0,596	0,592	0,577	0,561	0,543
2021	0,044	0,095	0,125	0,135	0,128
2022	0,600	0,472	0,497	0,540	0,561

Table 5.2. ±1 dilution
Таблица 5.2. ±1 разведение

Year Год	1 year год	2 year года	3 year года	4 year года	5 year лет
2014	67,84	67,94	67,94	67,24	67,24
2015	46,46	46,69	45,67	45,59	45,75
2016	40,43	38,41	36,50	34,48	34,36
2017	56,40	59,40	63,40	55,83	59,54
2018	65,27	59,62	62,97	60,88	61,09
2019	45,44	50,94	59,41	57,20	57,57
2020	57,70	57,03	58,33	57,32	56,79
2021	46,03	48,12	50,63	53,35	51,88
2022	55,27	50,18	54,16	55,66	60,49

Table 6. Regression results for 2022 data
Таблица 6. Результаты регрессии по данным 2022 г.

Customizable period Настраиваемые периоды	2022 г.	February 2022 report data Данные февральского отчета 2022 г.
Number of observations Количество наблюдений	8183	1994
Adjusted R^2 Скорректированный R^2	0,728	0,625
Standard deviation Стандартное отклонение	0,874	0,966
Forecast periods Прогнозные периоды		September 2022 report data Данные сентябрьского отчета 2022 г.
Number of observations Количество наблюдений		6675
Adjusted R^2 Скорректированный R^2		0,734
Standard deviation Стандартное отклонение		0,933
±1 dilution / разведение		69,33%
±2 dilution / разведения		98,73%

rare exceptions. Almost at all time intervals, a stable contribution, commensurate with the contribution of A and B, also demonstrates the antigenic site D, but only until 2018, followed by a sharp zeroing. Moderate compared to antigenic sites A and B, but at the same time stable significance is demonstrated by the antigenic site E with a noticeable increase by 2018–2021. The contribution of the antigenic site F varies diametrically depending on time intervals, from stable significant to zero. It should be noted that, in general, the results of the regression do not reveal a significant dependence on the number of years included in the adjusted period. At the same time, forecast results are significantly worse in those seasons when there is little data in the forecast period. In addition, our correlation analysis showed that there is no correlation between the values of the coefficients at antigenic sites and the number of substitutions in positions. Taking into account the fact that the number of observations only for 2022 is more than 1/3 of the total volume of observations for 2009–2022, it was decided to conduct separate calculations for observations in 2022. **Figure 6** shows the results of calculating the predictive ability of the developed cross-immunity model based on the published data of the RTGA 2022 (spring and autumn seasons).

Discussion

The analysis of the scientific literature demonstrated that in most studies the assessment of cross-immunity was carried out on a limited data set, including only reference strains, which reduces the objectivity of studies and makes it difficult to verify the forecast. Carrying out calculations based on the entire available array of HAI data published by WHO made it possible to bypass these limitations. With the help of the built-in additional module **Influenza DP**, which is a universal database processor (GISAID, NCBI, etc.), we have developed an algorithm that allows us to generate databases in real time according to the tasks for further analysis and calculations. To study the patterns in the cross–immunity model, subsets with the largest number of observations – Cell-Cell and Egg_s-Cell_t – were considered. We also introduced the concept of the degree of homology between two arbitrary strains, which takes the value of a function depending on the Hamming distance.

When considering functions for choosing a variant describing the degree of homology, a significant dependence of the values of the coefficient of determination on the type of function was noted. While all functions were subjected to the same necessary conditions corresponding to biological processes, the resulting adjusted R^2 values fluctuated quite significantly and took values from 0.327 to 0.513 for the Cell-Cell subset and 0.178 to 0.313 for the Egg_s-Cell_t subset, respectively. Such a significant difference between the upper and lower limits of the indicator indicates that the use of a simple Hamming distance [16, 18] as a measure of homology can impose significant restrictions on the models used. The ranking of the results for all functions was also carried out. When comparing the results with

the Cell-Cell and Egg_s-Cell_t subsets, almost complete correspondence of the function ranks by the value of the adjusted R^2 in relation to the subsets was observed. The results obtained may indicate a certain correspondence to the biological mechanisms of our approach to the choice of function.

Comparison of the adjusted R^2 values on the Cell-Cell and Egg_s-Cell_t subsets (0.523 and 0.313, respectively) demonstrates the validity of our decision to split the original dataset according to passage history. The difference in the adjusted R^2 also suggests that viruses cultured in chick embryos introduce a large error in the results of HAI assay.

In our calculations, we relied on the results of our own studies of combining amino acid positions into antigenic sites, described an additional F site, thereby expanding the number of antigenic sites to 6. The results of regression analysis on all Cell-Cell subsets showed that antigenic sites A and B made the greatest contribution to the immune response. They are moderately overtaken in stable values by the antigenic site E with a noticeable increase in 2018–2021. Antigenic site C practically does not contribute to the immune response in all considered intervals. Antigenic site D, which largely confidently demonstrates its significance at the beginning of the forecast periods, begins to lose ground in 2018–2021. The antigenic site F with its modestly unstable contribution to cross immunity turned out to be quite sensitive to time intervals.

In addition, we analyzed the stability of the predictive ability of our model for different time intervals (from 2009 to 2022). After dividing the Cell-Cell subset into 1 to 5 year time intervals, it was noted that the value of the adjusted R^2 in adjustable periods increases starting from 2019 reaching around 0.5 and above, except for 2021, which has very few observations (478) due to the COVID-19 pandemic. A similar picture is observed in the forecast periods. The pronounced high value of the adjusted R^2 for later data may be due to an increase in ongoing research (HAI assay) on cell culture and an improvement in their quality.

When comparing calculated values with published titers, reproducibility results per (±1) dilution are fairly stable across forecast years, with values greater than 50% (averaging around 60%), except for 2015, 2016, and 2021. At the same time, we see the main reason for the poorer results in a small amount of data for the forecast periods.

As shown in the results, the unprecedented number of observations for 2022 provided us with the opportunity to tune the model to the data at the beginning of 2022 (adjusted $R^2 = 0.728$), and check the results of the forecast for the autumn season. The prediction results (adjusted $R^2 = 0.734$) demonstrated titer reproducibility per (±1) dilution in 69.33% of cases and turned out to be significantly better than all previously considered periods, which can already be considered sufficient for practical application. The explanation for this result can be both the quality of the data and the forecast period limited by the season. In the work of T. Bedford et al. [15], who performed a regression analysis on the entire H_3N_2 HAI data set, obtained $R^2 = 0.372$, which is lower than our adjusted $R^2 = 0.523$ on the entire Cell-Cell subset. The results of colleagues could be affected by the lack of separation of data by passage history and time intervals. Also, the results may have been influenced by the use of a simple Hamming distance. It is important to recognize that, despite all the efforts to standardize the conduct of RTGA [32] and regardless of the chosen approach for studying the dependence of HAI titers on amino acid substitutions, the initially inherent high error of the HAI method (17%) [46] remains a factor that significantly affects the results.

Conclusions

Based on an epidemiological model and with a well-tuned cross-immunity model, we have developed a computer program that provides the ability to predict the most common strains, taking into account the influence of the immune landscape and select a vaccine strain for the upcoming season.

The ability to carry out calculations based on the entire available array of WHO-published seasonal HAI data, the observed improvement in results over recent years (starting in 2019), as well as a large volume of observations and the results of the forecast for 2022 (adjusted $R^2 = 0.734$; reproducibility of titers per one (± 1) dilution in 69.33% of cases) allow to assert that the proposed method can serve as a good tool for future forecasts with further study to confirm its stability.

In turn, the advantage of using the **Influenza IDE** computer program with a constantly updated database of various types and subtypes of the influenza virus in the future will allow reproducing the results obtained on other variants and subtypes of the influenza virus and thereby testing the developed model and expanding the spectrum of recommended vaccine strains. The experience and skills gained in determining antigenic sites in the future will make it possible to combine amino acid positions into antigenic sites according to several scenarios and making calculations already taking them into account. Further development of the cross-immunity model can be achieved by unilateral normalization of the titers of the entire array of test strains against reference strains, which will reduce the error associated with the individual characteristics of the experimental animals used. Also, the improvement of the model can be achieved by taking into account processes such as glycosylation.

REFERENCES

1. Russell C.A., Jones T.C., Barr I.G., Cox N.J., Garten R.J., Gregory V., et al. Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses. *Vaccine.* 2008; 26(Suppl. 4): 31–4. https://doi.org/10.1016/j.vaccine.2008.07.078
2. Belongia E.A., Simpson M.D., King J.P., Sundaram M.E., Kelley N.S., Osterholm M.T., et al. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. *Lancet Infect. Dis.* 2016; 16(8): 942–51. https://doi.org/10.1016/S1473-3099(16)00129-8
3. Jackson M.L., Chung J.R., Jackson L.A., Phillips C.H., Benoit J., Monto A.S., et al. Influenza vaccine effectiveness in the United

States during the 2015-2016 season. *N. Engl. J. Med.* 2017; 377(6): 534–43. https://doi.org/10.1056/NEJMoa1700153
4. Rolfes M.A., Flannery B., Chung J.R., O'Halloran A., Garg S., Belongia E.A., et al. Effects of influenza vaccination in the United States during the 2017-2018 influenza season. *Clin. Infect. Dis*. 2019; 69(11): 1845–53. https://doi.org/10.1093/cid/ciz075
5. Doyle J.D., Chung J.R., Kim S.S., Gaglani M., Raiyani C., Zimmerman R.K., et al. Interim estimates of 2018-2019 seasonal influenza vaccine effectiveness – United States. *MMWR. Morb. Mortal. Wkly Rep.* 2019; 68(6): 135–9. https://doi.org/10.15585/mmwr.mm6806a2
6. Zost S.J., Parkhouse K., Gumina M.E., Kim K., Diaz P.S., Wilson P.C., et al. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. Proc. Natl Acad. Sci. USA. 2017; 114(47): 12578–83. https://doi.org/10.1073/pnas.1712377114
7. Gouma S., Weirick M., Hensley S.E. Antigenic assessment of the H3N2 component of the 2019-2020 northern hemisphere influenza vaccine. Nat. Commun. 2020; 11(1): 2445. https://doi.org/10.1038/s41467-020-16183-y
8. Cobey S., Gouma S., Parkhouse K., Chambers B.S., Ertl H.C., Schmader K.E., et al. Poor immunogenicity, not vaccine strain egg adaptation, may explain the low H3N2 influenza vaccine effectiveness in 2012-2013. *Clin. Infect. Dis.* 2018; 67(3): 327–33. https://doi.org/10.1093/cid/ciy097
9. Klingen T.R., Reimering S., Guzmán C.A., McHardy A.C. In silico vaccine strain prediction for human influenza viruses. *Trends Microbiol*. 2018; 26(2): 119–31. https://doi.org/10.1016/j.tim.2017.09.001
10. Morris D.H., Gostic K.M., Pompei S., Bedford T., Łuksza M., Neher R.A., et al. Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol*. 2018; 26(2): 102–18. https://doi.org/10.1016/j.tim.2017.09.004
11. Timofeeva T.A., Asatryan M.N., Al'tshteyn A.D., Naroditskiy B.S., Gintsburg A.L., Kaverin N.V. Predicting the evolutionary variability of the influenza A virus. *Acta Naturae*. 2017; 9(3): 48–54. https://doi.org/10.32607/20758251-2017-9-3-48-54 https://elibrary.ru/zqitjr
12. Boev B.V. Modeling of the epidemic of influenza A(H1N1) in Russia season 2009-2010. *Epidemiologiya i vaktsinoprofilaktika.* 2010; (1): 52–8. https://elibrary.ru/laedxn (in Russian)
13. Huddleston J., Barnes J.R., Rowe T., Kondor R., Wentworth D.E., Whittaker L., et al. Integrating genotypes and phenotypes improves long-term forecasts of seasonal influenza A/H3N2 evolution. *eLife.* 2020; 9: e60067. https://doi.org/10.7554/eLife.60067
14. CDC. Center for Disease Control and Prevention. Influenza (Flu). Available at: https://www.cdc.gov/flu/
15. Bedford T., Suchard M.A., Lemey P., Dudas G., Gregory V., Hay A.J., et al. Integrating influenza antigenic dynamics with molecular evolution. *eLife.* 2014; 3: e01914. https://doi.org/10.7554/eLife.01914
16. Anderson C.S., McCall P.R., Stern H.A., Yang H., Topham D.J. Antigenic cartography of H1N1 influenza viruses using sequence-based antigenic distance calculation. *BMC Bioinformatics.* 2018; 19(1): 51. https://doi.org/10.1186/s12859-018-2042-4
17. Lee M.S., Chen J.S. Predicting antigenic variants of influenza A/H3N2 viruses. *Emerg. Infect. Dis*. 2004; 10(8): 1385–90. https://doi.org/10.3201/eid1008.040107
18. Lees W.D., Moss D.S., Shepherd A.J. A computational analysis of the antigenic properties of haemagglutinin in influenza a H3N2. *Bioinformatics*. 2010; 26(11): 1403–8. https://doi.org/10.1093/bioinformatics/btq160
19. Burnet F.M., Lush D. The action of certain surface active agents on viruses. *Aust. J. Exp. Biol. Med. Sci.* 1940; 18: 141–50.
20. Archetti I., Horsfall F.L. Persistent antigenic variation of influenza A viruses after incomplete neutralization in ovo with heterologous immune serum. *J. Exp. Med.* 1950; 92(5): 441–62. https://doi.org/10.1084/jem.92.5.441
21. Lapedes A., Farber R. The geometry of shape space: application to influenza. *J. Theor. Biol.* 2001; 212(1): 57–69. https://doi.org/10.1006/jtbi.2001.2347
22. Smith D.J., Lapedes A.S., de Jong J.C., Bestebroer T.M., Rimmelzwaan G.F., Osterhaus A.D., et al. Mapping the antigenic and genetic evolution of influenza virus. *Science.* 2004; 305(5682): 371–6. https://doi.org/10.1126/science.1097211
23. Wiley D.C., Skehel J.J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem*. 1987; 56: 365–94. https://doi.org/10.1146/annurev.bi.56.070187.002053
24. Wiley D.C., Wilson I.A., Skehel J.J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature.* 1981; 289(5796): 373–8. https://doi.org/10.1038/289373a0
25. Wilson I.A., Cox N.J. Structural basis of immune recognition of influenza virus hemagglutinin. *Annu. Rev. Immunol*. 1990; 8: 737–71. https://doi.org/10.1146/annurev.iy.08.040190.003513
26. Liao Y.C., Lee M.S., Ko C.Y., Hsiung C.A. Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus. *Bioinformatics.* 2008; 24(4): 505–12. https://doi.org/10.1093/bioinformatics/btm638
27. Asatryan M.N., Agasaryan V.G, Shcherbinin D.N., Timofeev B.I., Ershov I.F., Shmyr I.S., et al. Influenza IDE. Registration certificate № 2020617965; 2020. (in Russian)
28. Lawson C.L., Hanson R.J. *Solving Least Squares Problems*. New Jersey: Englewood Cliffs; 1974.
29. Khalafyan A.A. *Mathematical Statistics with Elements of Probability [Matematicheskaya statistika s elementami teorii veroyatnosti]*. Moscow: Binom; 2010. (in Russian)
30. Stephenson I., Gaines Das R., Wood J.M., Katz J.M. Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: an international collaborative study. *Vaccine.* 2007; 25(20): 4056–63. https://doi.org/10.1016/j.vaccine.2007.02.039
31. Wood J.M., Major D., Heath A., Newman R.W., Höschler K., Stephenson I., et al. Reproducibility of serology assays for pandemic influenza H1N1: collaborative study to evaluate a candidate WHO International Standard. *Vaccine*. 2012; 30(2): 210–7. https://doi.org/10.1016/j.vaccine.2011.11.019
32. Zacour M., Ward B.J., Brewer A., Tang P., Boivin G., Li Y. Standardization of hemagglutination inhibition assay for influenza serology allows for high reproducibility between laboratories. *Clin. Vaccine Immunol.* 2016; 23(3): 236–42. https://doi.org/10.1128/CVI.00613-15
33. The Francis Crick Institute. Worldwide Influenza Centre lab. Annual and interim reports. Available at: https://www.crick.ac.uk/research/platforms-and-facilities/worldwide-influenza-centre/annual-and-interim-reports
34. DuPai C.D., McWhite C.D., Smith C.B., Garten R., Maurer-Stroh S., Wilke C.O. Influenza passaging annotations: what they tell us and why we should listen. *Virus Evol.* 2019; 5(1): vez016. https://doi.org/10.1093/ve/vez016
35. Wu N.C., Zost S.J., Thompson A.J., Oyen D., Nycholat C.M., McBride R., et al. A structural explanation for the low effectiveness of the seasonal influenza H3N2 vaccine. *PLoS Pathog.* 2017; 13(10): e1006682. https://doi.org/10.1371/journal.ppat.1006682
36. Park Y.W., Kim Y.H., Jung H.U., Jeong O.S., Hong E.J., Kim H. Comparison of antigenic mutation during egg and cell passage cultivation of H3N2 influenza virus. *Clin. Exp. Vaccine Res*. 2020; 9(1): 56–63. https://doi.org/10.7774/cevr.2020.9.1.56
37. Popova L., Smith K., West A.H., Wilson P.C., James J.A., Thompson L.F. Immunodominance of antigenic site B over site A of hemagglutinin of recent H3N2 influenza viruses. *PLoS One.* 2012; 7(7): e41895. https://doi.org/10.1371/journal.pone.0041895

38. Klein N.P., Fireman B., Goddard K., Zerbo O., Asher J., Zhou J. Vaccine effectiveness of cell-culture relative to egg-based inactivated influenza vaccine during the 2017-2018 influenza season. *PLoS One*. 2020; 15(2): e0229279. https://doi.org/10.1371/journal.pone.0229279
39. GISAID. Eurosurveillance; 2017. Available at: https://gisaid.org/resources/commentary-on-gisaid/
40. Smith T.F., Waterman M.S. Identification of common molecular subsequences. *J. Mol. Biol*. 1981; 147(1): 195–7. https://doi.org/10.1016/0022-2836(81)90087-5
41. GetArea. Available at: http://curie.utmb.edu/getarea.html
42. RCSB PDB: Homepage. Available at: https://www.rcsb.org/
43. Shcherbinin D.N., Alekseeva S.V., Shmarov M.M., Smirnov Yu.A., Naroditskiy B.S., Gintsburg A.L. The analysis of B-cell epitopes of influenza virus hemagglutinin. *Acta Naturae.* 2016; 8(1): 13–20. https://doi.org/10.32607/20758251-2016-8-1-13-20 https://elibrary.ru/vsnklb
44. A standardised numbering for all subtypes of Influenza A hemagglutin (HA) sequences based on the mature HA sequence. Available at: https://antigenic-cartography.org/surveillance/evergreen/HAnumbering/
45. Smith D.J., Forrest S., Hightower R.R., Perelson A.S. Deriving shape space parameters from immunological data. *J. Theor. Biol*. 1997; 189(2): 141–50. https://doi.org/10.1006/jtbi.1997.0495
46. Kilbourne E.D., ed. *The Influenza Viruses and Influenza.* London: Academic Press Inc.; 1975.

ЛИТЕРАТУРА

1. Russell C.A., Jones T.C., Barr I.G., Cox N.J., Garten R.J., Gregory V., et al. Influenza vaccine strain selection and recent studies on the global migration of seasonal influenza viruses. *Vaccine.* 2008; 26(Suppl. 4): 31–4. https://doi.org/10.1016/j.vaccine.2008.07.078
2. Belongia E.A., Simpson M.D., King J.P., Sundaram M.E., Kelley N.S., Osterholm M.T., et al. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. *Lancet Infect. Dis.* 2016; 16(8): 942–51. https://doi.org/10.1016/S1473-3099(16)00129-8
3. Jackson M.L., Chung J.R., Jackson L.A., Phillips C.H., Benoit J., Monto A.S., et al. Influenza vaccine effectiveness in the United States during the 2015-2016 season. *N. Engl. J. Med.* 2017; 377(6): 534–43. https://doi.org/10.1056/NEJMoa1700153
4. Rolfes M.A., Flannery B., Chung J.R., O'Halloran A., Garg S., Belongia E.A., et al. Effects of influenza vaccination in the United States during the 2017-2018 influenza season. *Clin. Infect. Dis*. 2019; 69(11): 1845–53. https://doi.org/10.1093/cid/ciz075
5. Doyle J.D., Chung J.R., Kim S.S., Gaglani M., Raiyani C., Zimmerman R.K., et al. Interim estimates of 2018-2019 seasonal influenza vaccine effectiveness – United States. *MMWR. Morb. Mortal. Wkly Rep.* 2019; 68(6): 135–9. https://doi.org/10.15585/mmwr.mm6806a2
6. Zost S.J., Parkhouse K., Gumina M.E., Kim K., Diaz P.S., Wilson P.C., et al. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. Proc. Natl Acad. Sci. USA. 2017; 114(47): 12578–83. https://doi.org/10.1073/pnas.1712377114
7. Gouma S., Weirick M., Hensley S.E. Antigenic assessment of the H3N2 component of the 2019-2020 northern hemisphere influenza vaccine. Nat. Commun. 2020; 11(1): 2445. https://doi.org/10.1038/s41467-020-16183-y
8. Cobey S., Gouma S., Parkhouse K., Chambers B.S., Ertl H.C., Schmader K.E., et al. Poor immunogenicity, not vaccine strain egg adaptation, may explain the low H3N2 influenza vaccine effectiveness in 2012-2013. *Clin. Infect. Dis.* 2018; 67(3): 327–33. https://doi.org/10.1093/cid/ciy097
9. Klingen T.R., Reimering S., Guzmán C.A., McHardy A.C. In silico vaccine strain prediction for human influenza viruses. *Trends Microbiol.* 2018; 26(2): 119–31. https://doi.org/10.1016/j.tim.2017.09.001
10. Morris D.H., Gostic K.M., Pompei S., Bedford T., Łuksza M., Neher R.A., et al. Predictive modeling of influenza shows the promise of applied evolutionary biology. *Trends Microbiol*. 2018; 26(2): 102–18. https://doi.org/10.1016/j.tim.2017.09.004
11. Тимофеева Т.А., Асатрян М.Н., Альтштейн А.Д., Народицкий Б.С., Гинцбург А.Л., Каверин Н.В. Прогнозирование эволюционной изменчивости вируса гриппа А. *Acta Naturae*. 2017; 9(3): 104–11. https://doi.org/10.32607/20758251-2017-9-3-48-54 https://elibrary.ru/zqitor
12. Боев Б.В. Модель развития эпидемии гриппа А(H1N1) в России в сезон 2009–2010 годов. *Эпидемиология и вакцинопрофилактика.* 2010; (1): 52–8. https://elibrary.ru/laedxn
13. Huddleston J., Barnes J.R., Rowe T., Kondor R., Wentworth D.E., Whittaker L., et al. Integrating genotypes and phenotypes improves long-term forecasts of seasonal influenza A/H3N2 evolution. *eLife*. 2020; 9: e60067. https://doi.org/10.7554/eLife.60067
14. CDC. Center for Disease Control and Prevention. Influenza (Flu). Available at: https://www.cdc.gov/flu/
15. Bedford T., Suchard M.A., Lemey P., Dudas G., Gregory V., Hay A.J., et al. Integrating influenza antigenic dynamics with molecular evolution. *eLife.* 2014; 3: e01914. https://doi.org/10.7554/eLife.01914
16. Anderson C.S., McCall P.R., Stern H.A., Yang H., Topham D.J. Antigenic cartography of H1N1 influenza viruses using sequence-based antigenic distance calculation. *BMC Bioinformatics*. 2018; 19(1): 51. https://doi.org/10.1186/s12859-018-2042-4
17. Lee M.S., Chen J.S. Predicting antigenic variants of influenza A/H3N2 viruses. *Emerg. Infect. Dis*. 2004; 10(8): 1385–90. https://doi.org/10.3201/eid1008.040107
18. Lees W.D., Moss D.S., Shepherd A.J. A computational analysis of the antigenic properties of haemagglutinin in influenza a H3N2. *Bioinformatics*. 2010; 26(11): 1403–8. https://doi.org/10.1093/bioinformatics/btq160
19. Burnet F.M., Lush D. The action of certain surface active agents on viruses. *Aust. J. Exp. Biol. Med. Sci.* 1940; 18: 141–50.
20. Archetti I., Horsfall F.L. Persistent antigenic variation of influenza A viruses after incomplete neutralization in ovo with heterologous immune serum. *J. Exp. Med.* 1950; 92(5): 441–62. https://doi.org/10.1084/jem.92.5.441
21. Lapedes A., Farber R. The geometry of shape space: application to influenza. *J. Theor. Biol.* 2001; 212(1): 57–69. https://doi.org/10.1006/jtbi.2001.2347
22. Smith D.J., Lapedes A.S., de Jong J.C., Bestebroer T.M., Rimmelzwaan G.F., Osterhaus A.D., et al. Mapping the antigenic and genetic evolution of influenza virus. *Science.* 2004; 305(5682): 371–6. https://doi.org/10.1126/science.1097211
23. Wiley D.C., Skehel J.J. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem*. 1987; 56: 365–94. https://doi.org/10.1146/annurev.bi.56.070187.002053
24. Wiley D.C., Wilson I.A., Skehel J.J. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature.* 1981; 289(5796): 373–8. https://doi.org/10.1038/289373a0
25. Wilson I.A., Cox N.J. Structural basis of immune recognition of influenza virus hemagglutinin. *Annu. Rev. Immunol.* 1990; 8: 737–71. https://doi.org/10.1146/annurev.iy.08.040190.003513
26. Liao Y.C., Lee M.S., Ko C.Y., Hsiung C.A. Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus. *Bioinformatics*. 2008; 24(4): 505–12. https://doi.org/10.1093/bioinformatics/btm638
27. Асатрян М.Н., Агасарян В.Г, Щербинин Д.Н., Тимофеев Б.И., Ершов И.Ф., Шмыр И.С. и др. Influenza IDE. Свидетельство о регистрации № 2020617965; 2020.
28. Lawson C.L., Hanson R.J. *Solving Least Squares Problems.* New Jersey: Englewood Cliffs; 1974.
29. Халафян А.А. *Математическая статистика с элементами теории вероятности.* М.: Бином; 2010.

30. Stephenson I., Gaines Das R., Wood J.M., Katz J.M. Comparison of neutralising antibody assays for detection of antibody to influenza A/H3N2 viruses: an international collaborative study. *Vaccine.* 2007; 25(20): 4056–63. https://doi.org/10.1016/j.vaccine.2007.02.039
31. Wood J.M., Major D., Heath A., Newman R.W., Höschler K., Stephenson I., et al. Reproducibility of serology assays for pandemic influenza H1N1: collaborative study to evaluate a candidate WHO International Standard. *Vaccine*. 2012; 30(2): 210–7. https://doi.org/10.1016/j.vaccine.2011.11.019
32. Zacour M., Ward B.J., Brewer A., Tang P., Boivin G., Li Y. Standardization of hemagglutination inhibition assay for influenza serology allows for high reproducibility between laboratories. *Clin. Vaccine Immunol.* 2016; 23(3): 236–42. https://doi.org/10.1128/CVI.00613-15
33. The Francis Crick Institute. Worldwide Influenza Centre lab. Annual and interim reports. Available at: https://www.crick.ac.uk/research/platforms-and-facilities/worldwide-influenza-centre/annual-and-interim-reports
34. DuPai C.D., McWhite C.D., Smith C.B., Garten R., Maurer-Stroh S., Wilke C.O. Influenza passaging annotations: what they tell us and why we should listen. *Virus Evol.* 2019; 5(1): vez016. https://doi.org/10.1093/ve/vez016
35. Wu N.C., Zost S.J., Thompson A.J., Oyen D., Nycholat C.M., McBride R., et al. A structural explanation for the low effectiveness of the seasonal influenza H3N2 vaccine. *PLoS Pathog.* 2017; 13(10): e1006682. https://doi.org/10.1371/journal.ppat.1006682
36. Park Y.W., Kim Y.H., Jung H.U., Jeong O.S., Hong E.J., Kim H. Comparison of antigenic mutation during egg and cell passage cultivation of H3N2 influenza virus. *Clin. Exp. Vaccine Res*. 2020; 9(1): 56–63. https://doi.org/10.7774/cevr.2020.9.1.56
37. Popova L., Smith K., West A.H., Wilson P.C., James J.A., Thompson L.F. Immunodominance of antigenic site B over site A of hemagglutinin of recent H3N2 influenza viruses. *PLoS One.* 2012; 7(7): e41895. https://doi.org/10.1371/journal.pone.0041895
38. Klein N.P., Fireman B., Goddard K., Zerbo O., Asher J., Zhou J. Vaccine effectiveness of cell-culture relative to egg-based inactivated influenza vaccine during the 2017-2018 influenza season. *PLoS One*. 2020; 15(2): e0229279. https://doi.org/10.1371/journal.pone.0229279
39. GISAID. Eurosurveillance; 2017. Available at: https://gisaid.org/resources/commentary-on-gisaid/
40. Smith T.F., Waterman M.S. Identification of common molecular subsequences. *J. Mol. Biol*. 1981; 147(1): 195–7. https://doi.org/10.1016/0022-2836(81)90087-5
41. GetArea. Available at: http://curie.utmb.edu/getarea.html
42. RCSB PDB: Homepage. Available at: https://www.rcsb.org/
43. Щербинин Д.Н., Алексеева С.В., Шмаров М.М., Смирнов Ю.А., Народицкий Б.С., Гинцбург А.Л. Анализ В-клеточных эпитопов гемагглютинина вирусов гриппа. *Acta Naturae.* 2016; 8(1): 14–22. https://doi.org/10.32607/20758251-2016-8-1-13-20 https://elibrary.ru/vsjyhf
44. A standardised numbering for all subtypes of Influenza A hemaggluttin (HA) sequences based on the mature HA sequence. Available at: https://antigenic-cartography.org/surveillance/evergreen/HAnumbering/
45. Smith D.J., Forrest S., Hightower R.R., Perelson A.S. Deriving shape space parameters from immunological data. *J. Theor. Biol*. 1997; 189(2): 141–50. https://doi.org/10.1006/jtbi.1997.0495
46. Кильбурн Э.Д., ред. *Вирусы гриппа и грипп.* Пер. с англ. М.: Медицина; 1978.