pp. 1–16 (2021)

Herd immunity levels and multi-strain influenza epidemics in Russia: a modelling study

V. N. Leonenko*

Abstract — In the present paper, we consider a compartmental epidemic model which simulates the co-circulation of three influenza strains, A(H1N1)pdm09, A(H3N2), and B, in a population with the history of exposure to these virus strains. A strain-specific incidence data for the model input was generated using long-term weekly ARI incidence and virologic testing data. The algorithm for model calibration was developed as a combination of simulated annealing and BFGS optimization methods. Two simulations were carried out, assuming the absence and the presence of protected individuals in the population, with 2017–2018 and 2018–2019 epidemic seasons in Moscow as a case study. It was shown that strain-specific immune levels defined by virologic studies might be used in the model to obtain plausible incidence curves. However, different output parameter values, such as fractions of individuals exposed to particular virus strain in the previous epidemic season, can correspond to similar incidence trajectories, which complicates the assessment of herd immunity levels based on the model calibration. The results of the study will be used in the research of the interplay between the immunity formation dynamics and the circulation of influenza strains in Russian cities.

Keywords: Mathematical modelling, epidemiology, seasonal influenza, herd immunity, strain co-circulation.

MSC 2010: 37N25, 65C20, 92C60

Outbreaks of influenza, one of the oldest and the most widely spread human infectious diseases, result in 3 to 5 million cases of severe illness annually worldwide, and the mortality rate is from 250 to 640 thousand individuals per year [11]. In addition to induced mortality, influenza causes an increase of heart attacks and strokes [5], as well as other disease complications. To enhance the capabilities of influenza surveillance and, as a consequence, to find means of restraining influenza epidemics and reducing the mortality attributed to influenza complications, the healthcare organs widely use statistical and mechanistic models. One of the most important factors related to influenza dynamics, beside weather conditions [18, 21, 29, 31, 33, 34] and contact patterns in the population [2, 16, 23, 35], is the herd immunity to various influenza strains [3, 9, 15, 24].

It is generally known that the immunity level dynamics and the disease incidence dynamics in the population are intertwined but there is still a lot of open questions related to quantification of their connection. In earlier influenza modelling

*ITMO University, Saint-Petersburg, Russia. E-mail: vnleonenko@yandex.ru

This research was financially supported by the Russian Science Foundation, Agreement No. 20-71-00142.

efforts, such as those undertaken in the second part of the XX century [1, 4, 7, 30], it was found possible to explain the influenza dynamics on the basis of the assumption which related the variance in the disease incidence primarily to the seasonal virus mutation. Thus, different geographical areas were assumed to have approximately the same level of immunity to a circulating virus strain, and there was no need in considering the long-term history of exposure to different influenza strains in particular regions. However, since the early 1980s this assumption was put under doubt since modelling frameworks based on it started to show growing incoherence with the observed seasonal epidemic outbreak patterns. According to one of the opinions, the reason for that is in the growing immunity levels to influenza due to increasing speed of its circulation around the globe [13]. The experiments with the SEIR models calibration on the contemporary Russian ARI incidence data [19, 20] demonstrated that assuming the variance in immunity levels among Russian cities negligible, one can obtain the prediction of prospected epidemic peak height with satisfactory accuracy, but there is no chance to assess in advance the epidemic peak day. The experiments with calibrating metapopulational models to Russian ARI incidence data and data on inter–city transport flows within the country demonstrated, that the accuracy of peaks predictions made by homogeneous coupled SEIR models on the contemporary incidence data is much lower than it used to be in Soviet times and does not allow to use this approach for the disease forecasting [21]. Due to that matter, and also to the increased availability of serological data, the research area devoted to the analysis of the long-term influenza dynamics and the herd immunity levels is gaining popularity among other topics related to the epidemiology of influenza. As an example, a set of models was presented which was designed to model the long-term incidence of influenza in England caused by various viral strains, taking into account the dynamics of population immunity levels and vaccination coverage [9, 10].

In the current study, we use a deterministic compartmental model to analyze the multi-strain epidemics caused by co-circulation of different influenza virus strains in a fixed epidemic season. Apart from similar influenza modelling works, particularly, devoted to the circulation of influenza pandemic strains [6, 8], where herd immunity is not considered, we assume an exposure history to different influenza strains which influences the disease dynamics in the regarded epidemic season. A strain-specific incidence data, employed for model calibration, was generated by coupling long-term weekly ARI incidence with virologic testing data — both datasets being provided by Russian Research Institute of Influenza [28]. Using the data for two epidemic seasons in Moscow as a case study, we perform model calibration and assess its parameter values related to virulence of influenza strains, intensity of immunity waning and the levels of immunity induced by the preceding epidemic season. As a consequence, we obtain a possibility to compare the levels of immunity in the population estimated in two different ways: indirectly through the assessment of the epidemic model parameters and directly from serological monitoring data. The presented study is a part of the ongoing research, the ultimate aim of which consists in finding the connection between the two named immunity assessment

methods and thus gaining more insight into the interplay between the immunity formation dynamics and the circulation of influenza strains in Russian cities.

1. Data

The datasets used in the study were the following:

- Weekly ARI incidence (Moscow, 2009–2019), which includes total number of clinically registered cases of diseases caused by influenza strains and other ARI.
- Strain-specific virologic data (Moscow, 2009–2019) with the assessment of the number of detected influenza strains (see Figs. 1 and 2).

Since the second mentioned dataset became available to us, we managed to simplify the assessment of ARI cases related to influenza, compared to previous studies, where we had to rely on establishing the fixed [17, 20] or varied [22] ARI baselines according to the known epidemic thresholds. In the current research, we used the following algorithm to find strain-specific incidence:

- We calculated the weekly proportion of strain-specific influenza cases detected via virologic studies as a ratio between the number of detected particular influenza strains in the samples, and the total number of samples collected (assuming all of them corresponded to ARI cases).
- If the total number of collected samples was stated to be less than the sum of the samples corresponding to specific influenza strains, we considered the corresponding data erroneous and removed it from consideration.
- The cases marked 'strain A(type unknown)' were distributed among the cases of A(H1N1)pdm09 and A(H3N2) according to the proportion established by comparing the number of correctly specified samples.
- The obtained proportions of influenza strain incidence among ARI were multiplied by the weekly registered ARI incidence from the first dataset.

The resulting strain-specific incidence is shown in Fig. 3.

2. Model

In this study, a multi-strain compartmental model of influenza dynamics is used based on a deterministic system of difference equations, with the time step equal to one day. The model was derived from the original Baroyan–Rvachev model [4] and its modifications used by the author in the earlier studies [20, 22]. In this model version, we consider the co-circulation of three influenza strains, A(H1N1)pdm09, A(H3N2), and B, thus, we assume $n_s = 3$, where n_s is the total number of regarded

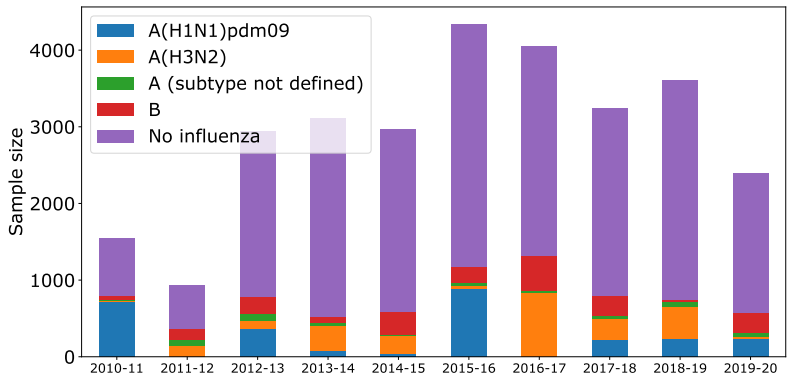

Figure 1. Proportion of different influenza strains summed by epidemic seasons, Moscow.

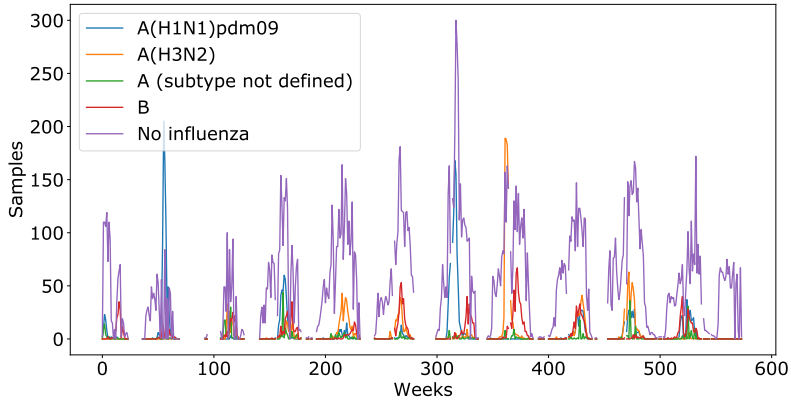

Figure 2. Weekly number of different influenza strains and ARI collected through virologic sampling, Moscow.

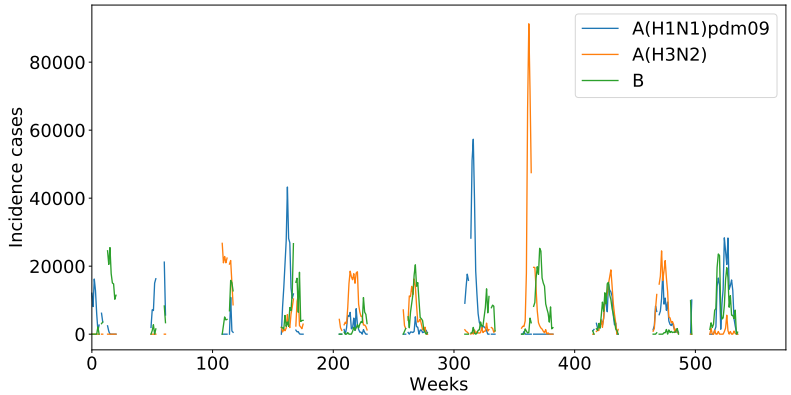

Figure 3. Resulting strain-specific incidence data for Moscow.

strains. Different strains of influenza B type are not distinguished and the dominant B type strain is regarded during each epidemic season. Let $x_t^{(h)}$ be the fraction of susceptible individuals in the population with exposure history $h \in \{1,\dots,n_s+1\}$, $y_t^{(m)}$ be the number of individuals newly infected at the moment t by the virus strain m, and $\overline{y}_t^{(m)}$ be the cumulative number of infectious persons by the time t transmitting the virus strain m, $m \in \{1,\dots,n_s\}$. A possibility of co-infection by multiple strains in the course of one season is not regarded, hence, the individuals recovered from the influenza caused by any of the circulating strains are considered immune. However, this assumed cross-immunity between virus strains is not transferred to the next epidemic season.

The susceptible individuals are divided into subgroups based on their exposure history h, $h \in \{1,\dots,n_s+1\}$. A group of susceptible individuals with exposure history state $h \in \{1,\dots,n_s\}$ is composed of those individuals who were subjected to infection by the strain m in the previous epidemic season, whereas a group with exposure history state $h = n_s+1$ is regarded as naive to the infection caused by any strain. The variable $\mu \in [0;1)$ reflects the fraction of population which do not participate in infection transmission (see Section 4.2). In the default case, $\mu = 0$. Following [9], we make a simplifying assumption that the preceding epidemic season is the one that solely affects the immunity levels of individuals, thus neglecting the effects of older exposures to influenza illness. Due to immunity waning, the individuals with the history of exposure to a fixed influenza strain in the preceding season might lose immunity to the same strain in the following epidemic season. We assume that the fraction a of those individuals, $a \in (0;1)$, becomes susceptible, whereas $1-a$ individuals retain their immunity during the modelled epidemic season. As a result, a function $f(h,m)$ is introduced into the model which defines the proportion of the individuals with exposure history state h, who are susceptible to virus strain m:

$$f(h,m) = \begin{cases} a, & m = h \\ 1, & m \neq h. \end{cases} \tag{2.1}$$

The modelling equation system is formulated in the following way:

$$x_{t+1}^{(h)} = \max\left\{0, \left(1 - \sum_{m=1}^{n_s} \frac{\beta^{(m)}}{\rho} \overline{y}_t^{(m)} f(h,m)\right) x_t^{(h)}\right\}, \quad h \in \{1,\dots,n_s+1\} \tag{2.2}$$

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

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

$$x_0^{(h)} = \alpha^{(h)}((1-\mu)\rho - \sum_{m=1}^{n_s} y_0^{(m)}) \geqslant 0, \quad h \in \{1,\dots,n_s+1\}$$

$$y_0^{(m)} = \varphi_0^{(m)} \geqslant 0, \quad m \in \{1,\dots,n_s\}. \tag{2.3}$$

The piecewise constant function g_τ gives a fraction of infectious individuals in the group of individuals infected τ days before the current moment t. The function reflects the change of individual infectiousness over time from the moment of acquiring influenza. It is assumed that there exists some moment $\bar{t}$: $\forall t \geqslant \bar{t}\ g_\tau = 0$, which reflects the moment of recovery. The values of $g(\tau)$ were set according to [4]: $g(0) = g(1) = 0$, $g(3) = 0.9$, $g(3) = 0.9$, $g(4) = 0.55$, $g(5) = 0.3$, $g(6) = 0.15$, $g(7) = 0.05$, $g(8) = g(9) = \ldots = 0$. The initial conditions are assumed to be non-negative.

We consider that $\beta^{(m)}$, the intensity of effective contacts (i.e., contacts followed by infection transmission) related to virus strain m, conforms to the following expression:

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

where $\lambda^{(m)}$ is virulence of the strain m, δ is the average number of contacts in the population [22].

3. Algorithm

3.1. Optimization criterion

Let $Z^{(\mathrm{dat})}$ be a set of incidence data points corresponding to one particular outbreak. Assuming that there are no gaps in incidence data, let t_1 be the length of the registered period of incidence, which coincides with the size of $Z^{(\mathrm{dat})}$. The optimum criterion is reached by minimizing the difference between the incidence values taken from data and the incidence calculated by the model:

$$F(Z^{(\mathrm{mod})}, Z^{(\mathrm{dat})}) = \sum_{i=0}^{t_1} w_i \cdot \left(z_i^{(\mathrm{mod})} - z_i^{(\mathrm{dat})}\right)^2 \longrightarrow \min. \tag{3.1}$$

Here $z_i^{(\mathrm{dat})}$ and $z_i^{(\mathrm{mod})}$ are relative incidences on the ith *week* taken from the input dataset and derived from the model, respectively. The weights w_i are included into the formula to reflect the varied impact of assessment quality for different incidence points: we assume that the accurate peak assessment is the most important task, and the 'importance' of the other incidence points is proportional to their distance from the peak. The corresponding formula is

$$w_d = \sigma^{-d}$$

where d is the distance of the current data point from the peak in time units (weeks), $\sigma > 1$ regulates the rate of decreasing of the 'importance' of incidence points for the fitting procedure.

To compare the daily incidences of the model output with the original weekly incidence data, the former are summed in groups of 7.

Table 1. Model parameters.

Variable	Description	Values
ρ	Population size, people	$12,263,854$
$\alpha^{(h)}$	A fraction of population exposed to the strain m in the preceding epidemic season, $h \in \{1,\dots,m\}$*	Varied
$\lambda^{(m)}$	Virulence of the virus strain m	Varied
a	The fraction of people who lost immunity after being exposed to the virus strain in the preceding epidemic season	Varied
δ	Average daily number of contacts in the population for a fixed individual	6.528 [22]
$I_0^{(m)}$	Initial number of individuals infected with the strain m, people	1
σ	Parameter for calculation of the weights w_d of data points	Varied

* A fraction $\alpha^{(m+1)}$ of the naive population was calculated as $\alpha^{(m+1)} = 1 - \sum_{h=1}^{m} \alpha^{(h)}$

3.2. Calibration algorithm

To perform the optimization and assess the model parameters, we used the scripts implemented in Python 3.6 with `numpy` and `matplotlib` libraries. The arising issue of executing numerical optimization procedure with the optimum criterion (3.1) is the convergence of the algorithm to local minima which results in finding non-optimal parameter values [20]. To deal with this problem, simulated annealing optimization algorithm [14] was used. Simulated annealing deals effectively with local minima problem, delivering the values which are close to optimum, but it converges very slowly to the optimum itself. Thus it is recommended to consequently apply another optimization method, using the values found by annealing as initial ones. For example, in [36], where the authors also performed influenza dynamics model calibration, Nelder–Mead method was used. According to the results of numerical experiments with the standard methods for the optimization function `scipy.optimize.minimize` [32], Nelder-Mead method showed poor performance during influenza dynamics model calibration compared to limited-memory BFGS optimization method [26], thus we decided to use the latter in the calibration algorithm. However, we assume that the calibration algorithm performance might depend on the type of input data and model structure. In this case, additional experiments could be beneficial to define the best optimization method for the case of the multi-strain model.

4. Experiments

To calibrate the model, we used generated strain-specific incidence corresponding to two epidemic seasons (Moscow, 2017–2018 and Moscow, 2018–2019). These particular outbreaks were chosen due to the availability of the immunity level assessments through virologic studies for this city and time period.

The model parameters used in the calibration procedures are listed in Table 1.

The procedure was performed according to the optimization algorithm described in the previous section. In the course of generating the parameter values, we controlled that the condition $\sum_{h=1}^{m} \alpha^{(h)} \leqslant 1$ is not violated by normalization of the values which exceeded 1 in sum.

4.1. Original model

Apart from the previous studies [20, 22] where we assumed a fraction of immune individuals in the population which was invulnerable to infection and thus excluded from the epidemic process, the structure of the current model supposes that each individual is in fact prone to the infection caused by any of the circulating strains, with the exposure history only slightly reducing the intensity of the infection process (for $a \neq 0$). As a consequence, the reservoir of susceptible individuals has immensely grown resulting in extremely high outbreak peaks. Neither automatic nor manual calibration was able to find parameter values corresponding to the fitting with determination coefficient $R^2_\sigma > 0.5$. In Fig. 4 one can see a typical set of trajectories which were manually fitted to match the incidence peak heights. The corresponding infection force does not allow reasonably fast outbreaks, resulting in very slow disease propagation through the population (by several orders longer than the real outbreak duration). The attempt to match the disease outbreak durations, on the contrary, caused unrealistic incidence numbers (see Fig. 5).

4.2. Introducing protected individuals

To make it possible for the model to reproduce real incidence, we had to introduce the proportion μ of the individuals with the protection from infection by any influenza strains. This proportion was chosen to be $\mu = 0.9$, following the order of the values obtained in [22]. The resulting modelling curves became more plausible and better corresponded to data. The automatic calibration with $\sigma = 1.3$ resulted in finding some erroneous local minima of the optimization function corresponding to zero incidence model curves for one or several strains. Since this output was not reflecting the actual incidence, we avoided those minima by narrowing the intervals for some of the model parameters (see Table 6, second column in [22]). Another set of experiments was conducted with enhanced and unified parameter value intervals, which reflects the situation of batch calibration of many consecutive epidemic seasons when manual tweaking of parameter ranges is not possible. We used the value of $\sigma = 1.5$ to increase the importance of fitting the maximum incidence at the expense of the incidence points situated farther from the peak. The fitting results for both experiments are shown in Figs. 6 and 7.

As one can see, for 2018–2019 epidemic season in Moscow, both fitting outcomes with different σ did not match the data. The apparent reason for that is the complicated form of the incidence curve induced by B strain, which includes several low peaks. It seems to correspond to an outbreak which did not gain momentum after several disease reintroductions.

The epidemic season of 2017–2018 was better described by the fitted modelling

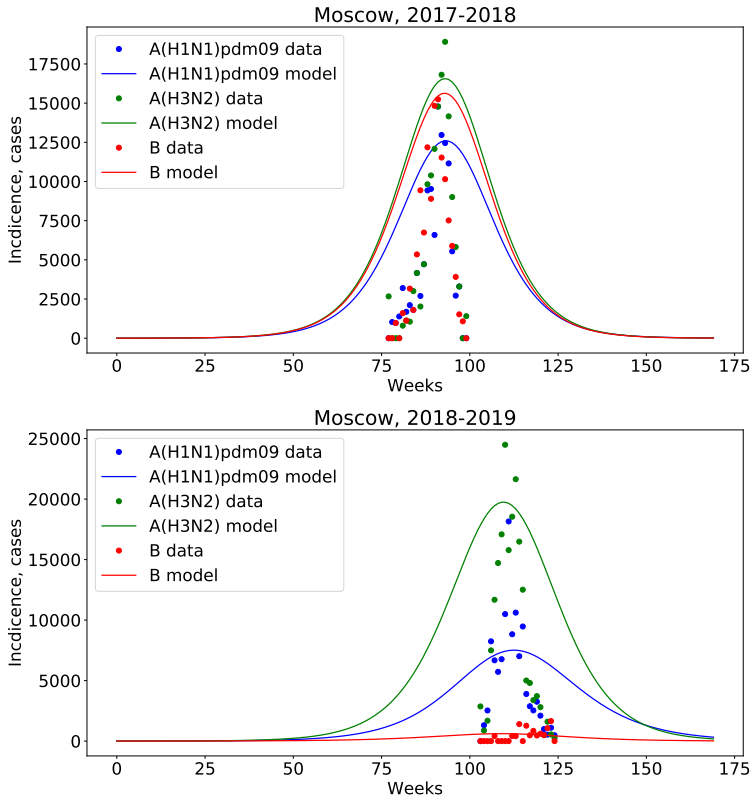

Figure 4. Model trajectories against incidence data in case with no protected individuals assumed in the population.

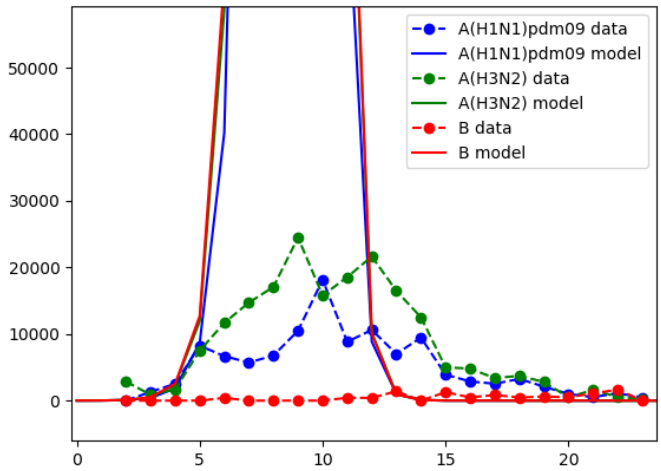

Figure 5. An example of explosive incidence growth demonstrated by short epidemics compared to Moscow data.

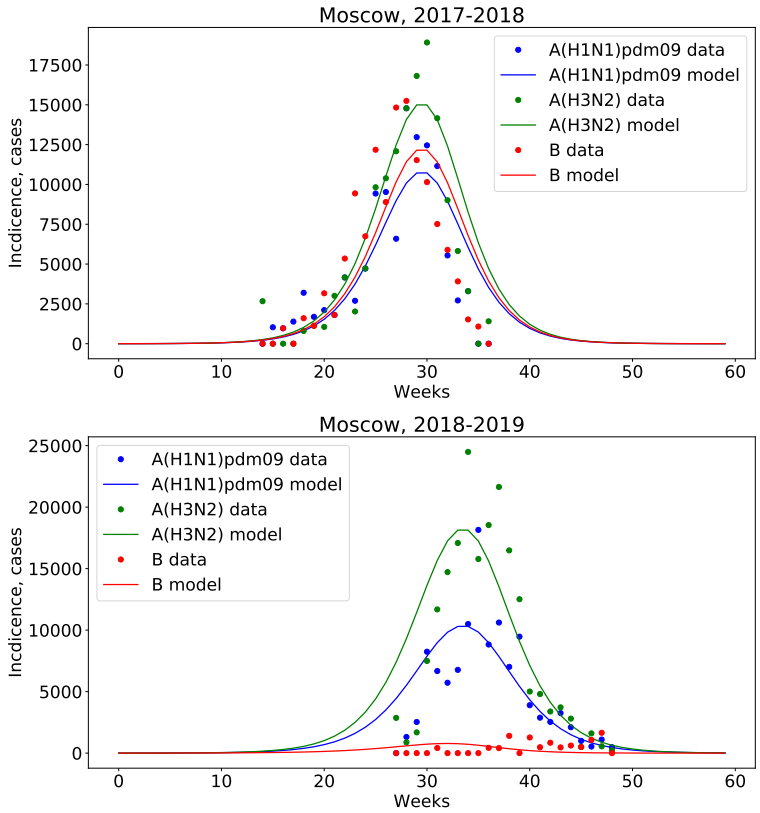

Figure 6. Model fitting to incidence data in the population with protected individuals, $\sigma = 1.3$.

Table 2. Variation of model parameters.

Variable	$\sigma = 1.3$	$\sigma = 1.5$	$\sigma = 1.3$	$\sigma = 1.5$
	Moscow, 2017–2018		Moscow, 2018–2019	
$\alpha^{\mathrm{A(H1N1)pdm09}}$	$(0.21, 0.27)$*	$(0.1, 0.4)$	$(0.1, 0.4)$	$(0.1, 0.4)$
$\alpha^{\mathrm{A(H3N2)}}$	$(0.32, 0.37)$	$(0.1, 0.4)$	$(0.1, 0.4)$	$(0.1, 0.4)$
α^{B}	$(0.37, 0.42)$	$(0.1, 0.4)$	$(0.1, 0.4)$	$(0.1, 0.4)$
$\lambda^{\mathrm{A(H1N1)pdm09}}$	$(0.07, 0.08)$	$(0.06, 0.13)$	$(0.06, 0.12)$	$(0.06, 0.13)$
$\lambda^{\mathrm{A(H3N2)}}$	$(0.07, 0.09)$	$(0.06, 0.13)$	$(0.08, 0.13)$	$(0.06, 0.13)$
λ^{B}	$(0.07, 0.09)$	$(0.06, 0.13)$	$(0.06, 0.13)$	$(0.06, 0.13)$
a	$(0.05, 0.3)$	$(0.05, 0.5)$	$(0.1, 0.5)$	$(0.05, 0.5)$

* the parameter range was narrowed to avoid erroneous local optimum with zero A(H1N1)pdm09 prevalence

curve. In the experiment with $\sigma = 1.3$ the first phase of multi-strain disease outbreak is portrayed rather close to reality, but the peak heights are underestimated for all the strains. In case of $\sigma = 1.5$ the increased 'value' of matching the peak according to the optimization function resulted in accurate depiction of one peak out of three, which is A(H3N2) (Fig. 7, green curve). However, that caused the mismatch in

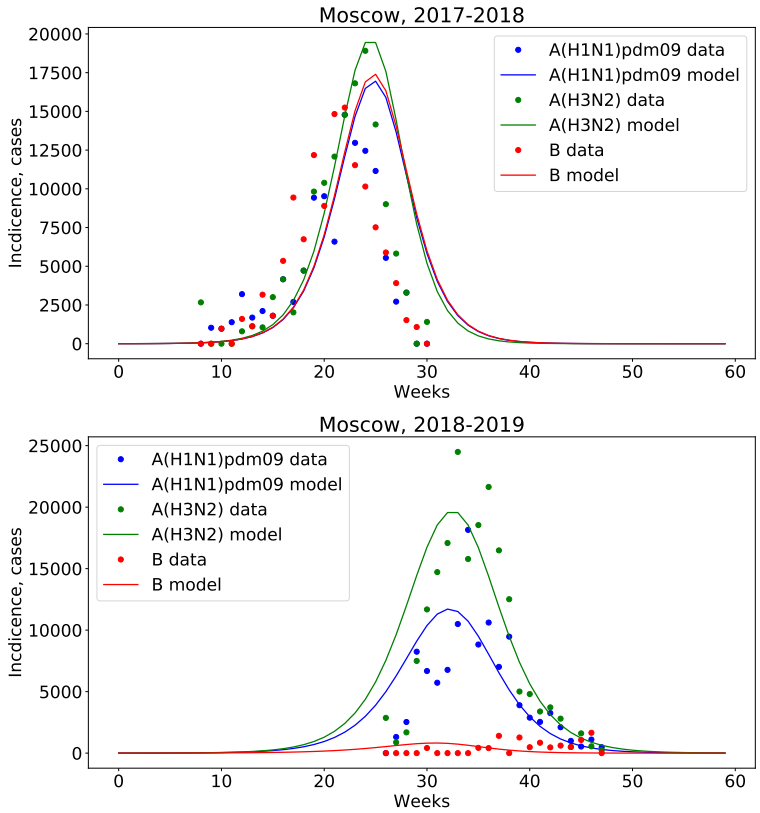

Figure 7. Model fitting to incidence data in the population with protected individuals, $\sigma = 1.5$.

assessing the other two peaks, along with the underestimation of disease incidence before the peaks and its overestimation after the peaks.

Table 3 demonstrates that changing of σ also caused changes in assessed values of parameters related to exposure history and thus consequentially to the immunity levels (for the 2017–2018 epidemic outbreak, the optimal value of $\alpha^{\mathrm{A(H1N1)pdm09}}$ was reduced in two times and α^{B} – in three times). Partly this might be caused by the known correlation between the parameters reflecting the immunity levels and the disease virulence [4, 12]. As a consequence of that fact, different combinations of λ and α might result in the similar incidence trajectories which greatly complicates the assessment of the immunity levels from the output of the model fitted to disease incidence trajectories.

In Table 4, the proportions of seropositives among the biological samples are presented. It is worth noting, that the search of optimal values for $\alpha^{(h)}$ close to the values drawn from the serological studies for Moscow, 2017–2018 (see the experiment with $\sigma = 1.3$, Table 2 under an asterisk) resulted in disease trajectories close to the real incidence data. Unfortunately, the inverse is not true: it seems that due to above–mentioned uncertainty of assessing the parameter values the seroprevalence

Table 3. Optimal parameters for the fitted models.

Variable	$\sigma = 1.3$	$\sigma = 1.5$	$\sigma = 1.3$	$\sigma = 1.5$
	Moscow, 2017–2018		Moscow, 2018–2019	
$\alpha^{\mathrm{A(H1N1)pdm09}}$	0.22	0.1	0.31	0.33
$\alpha^{\mathrm{A(H3N2)}}$	0.36	0.33	0.32	0.16
α^{B}	0.37	0.1	0.11	0.10
$\lambda^{\mathrm{A(H1N1)pdm09}}$	0.0757	0.0734	0.07895	0.0807
$\lambda^{\mathrm{A(H3N2)}}$	0.0867	0.0953	0.0804	0.0713
λ^{B}	0.0873	0.0734	0.0641	0.0641
a	0.28	0.11	0.31	0.32

Table 4. Fraction of samples seropositive to influenza strains in pre-epidemic period according to surveillance data.

Strain	Moscow, 2017–2018	Moscow, 2018–2019
A(H1N1)pdm09	0.29	0.76
A(H3N2)	0.40	0.71
B	0.45*	0.74/0.67**

* samples seropositive to B/Victoria strain
** samples seropositive to B/Colorado and B/Yamagata strains, respectively

results cannot be accurately assessed based on the model fitting output, at least, for the current experiment design.

It is interesting to note that some of the numbers in Table 4 have sums more than 1, assuming that some individuals in the population had immunity to several influenza strains at once, which contradicts the assumptions from [9] used in the current study. A possible explanation is the discrepancy between the detected antibodies in the sample and the actual immunity against the virus. In any case, this issue requires careful consideration.

5. Discussion

In the presented research, a compartmental model for influenza strains co-circulation was developed and calibrated to the generated strain-specific incidence data. The experiments performed with two different sets or parameter ranges on two epidemic seasons (Moscow, 2017–2018 and Moscow, 2018–2019) allow us to make the following conclusions.

First of all, the accurate reproduction of real incidence data seems to be not possible under the initial modelling assumptions (no fully protected individuals, $\mu = 0$). The model trajectories which are similar to data in terms of the peak height demonstrate exceedingly long and thus unrealistic epidemic durations (see Fig. 4) and the trajectories with plausible epidemic durations exhibit unnaturally high peaks (see Fig. 5). Regarding some fraction of individuals in the population as fully protected helped bring the model curve parameters to similarity with the incidence data. Nevertheless, the ideal model fit was not reached, especially for the case of Moscow,

2018–2019, where the epidemic incidence for B strain was not adequately described by the model (the coefficient of determination R^2_σ was less than 0 for both values of σ). In this particular case, the reason of discrepancy lies in the apparent late onset of the B strain outbreak compared to the outbreaks caused by A strains. This difference in onset timing also worsened the model fit for A(H1N1)pdm09 and A(H3N2) incidence curves. To properly describe such disease outbreak data, it might be fruitful to incorporate possible time differences between the outbreak onsets into the model by introducing strain-specific $t_0^{(m)} \geqslant 0$ instead of currently used $t_0^{(m)} \equiv 0$. This mechanism will reflect the onset of a full-fledged outbreak of particular virus strain taking place after several failed strain reintroductions into the population, the possibility of which is demonstrated by the stochastic models [25, 27].

In the other dataset used for calibration (Moscow, 2017–2018), the variation of σ caused big discrepancy in the resulting obtained optimal values of $\alpha^{(h)}$ and $\lambda^{(m)}$. Combined with the fact that modelling incidence curves for both values of σ (see Figs. 6 and 7) do not accurately reflect the data ($R^2_\sigma \in (0.5, 0.8)$), other local minima could exist and be revealed in the repetitive model calibration. This might add even more uncertainty to the parameter output. To overcome this issue, several options will be considered. First of all, it is necessary to enhance the model fit to data, at least achieving $R^2_\sigma > 0.8$. This aim might be reached by tweaking the weights of the fitting function to avoid some of the erroneous local optima (for instance, those related to zero model curves) or switching to the other methods of curve fitting, for instance, approximate Bayesian computation [9]. Finally, we could limit the state space by assuming equal strain virulence $\lambda^{(m)} \equiv \lambda > 0$ and narrowing uncertainty intervals for the parameters that might be assessed using additional virologic data or found in the papers published by other research teams. In any case, interval assessments instead of point assessments should be found for the herd immunity levels. Those assessments might be obtained by repeating the optimization procedures, altering the calibration parameters and regarding more data on influenza outbreaks for model calibration (different cities as well as different epidemic seasons). Once those interval assessments are gathered, we will be able to start matching them with the immunity levels measured in the course of virologic studies.

From the standpoint of epidemiological application of the model, it is crucial to understand why the adequate representation of epidemic incidence curves was made possible by the model only after the introduction of fully protected individuals. The reasons of that might be the following ones: (a) the part of the population is in fact not reachable by infection due to the heterogeneity in contact patterns, and hence excluded from the infection process, which cannot be captured by the mass action law principle used in the compartmental model; (b) we do not account for the fact that the provided data solely represent the detected part of the incidence cases, rather than all the cases occurred during an epidemic season. Also, the sizes of samples in virologic data, which are used to calculate strain-specific incidence, allow big biases in weekly registration of influenza cases and thus might result in high uncertainty of resulting strain-specific incidence assessment. We plan to tackle this problem by adding a separate model for healthcare services attendance and assessing the role

of population heterogeneity in forming indirect protection via multiagent models [23, 25]. This assessment might be later used to more accurately define the protected fraction of individuals in susceptible cohorts of a multi-strain compartmental model.

It is worth noting that two issues related to managing epidemiological data must be addressed, before the serological sampling results might be used in modelling efforts in the most efficient way. Firstly, the samples were collected from the adults (18 years of age and older), so to effectively match the results of serological studies to the immunity levels assessed by model calibration, an explicit age structure should be implemented in the compartmental model. Secondly, we need to develop a transparent interpretation of the sampling results which assume that an individual can be immune to several influenza strains at once (see Section 4.2, Table 4). If it is possible from the epidemiological point of view, contrary to the assumption in [9], we will need to create additional exposure history states corresponding to immunity to several strains. At last, we would like to explicitly add into the model the artificial immunity caused by vaccination.

References

1. A. Aguirre and E. Gonzalez, The feasibility of forecasting influenza epidemics in Cuba. *Memorias do Instituto Oswaldo Cruz* **87** (1992), No. 3, 429–432.
2. M. Ajelli and M. Litvinova, Estimating contact patterns relevant to the spread of infectious diseases in Russia. *J. Theor. Biology* **419** (2017), 1–7.
3. M. Baguelin, S. Flasche, A. Camacho, N. Demiris, E. Miller, and W. J. Edmunds, Assessing optimal target populations for influenza vaccination programmes: an evidence synthesis and modelling study. *PLoS Medicine* **10** (2013), No. 10, e1001527.
4. O. V. Baroyan, U. V. Basilevsky, V. V. Ermakov, K. D. Frank, L. A. Rvachev, and V. A. Shashkov, Computer modelling of influenza epidemics for large-scale systems of cities and territories. In: *Proc. WHO Symposium on Quantitative Epidemiology, Moscow*, 1970.
5. CDC, People with heart disease and those who have had a stroke are at high risk of developing complications from influenza (the Flu). URL: http://www.cdc.gov/flu/heartdisease/
6. V. Colizza, A. Barrat, M. Barthelemy, A.-J. Valleron, and A. Vespignani, Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Medicine* **4** (2007), No. 1, e13.
7. A. Flahault, S. Letrait, P. Blin, S. Hazout, J. Menares, and A.-J. Valleron, Modelling the 1985 influenza epidemic in France. *Statistics in Medicine* **7** (1988), No. 11, 1147–1155.
8. I. M. Hall, R. Gani, H. E. Hughes, and S. Leach, Real-time epidemic forecasting for pandemic influenza. *Epidemiology and Infection* **135** (2007), No. 3, 372–385.
9. E. M. Hill, S. Petrou, S. De Lusignan, I. Yonova, and M. J. Keeling, Seasonal influenza: Modelling approaches to capture immunity propagation. *PLoS Comput. Biology* **15** (2019), No. 10, e1007096.
10. E. M. Hill, S. Petrou, H. Forster, S. De Lusignan, I. Yonova, and M. J. Keeling, Optimising age coverage of seasonal influenza vaccination in England: A mathematical and health economic evaluation. *PLoS Comput. Biology* **16** (2020), No. 10, 1–32.
11. A. D. Iuliano, K. M. Roguski, H. H. Chang, D. J. Muscatello, R. Palekar, S. Tempia, C. Cohen, J. M. Gran, D. Schanzer, B. J. Cowling, et al., Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *The Lancet* **391** (2018), No. 10127, 1285–1300.

12. Yu. G. Ivannikov and A. T. Ismagulov, The epidemiology of influenza. Almaty, Kazakhstan, 1983 (in Russian).

13. Yu. G. Ivannikov and P. I. Ogarkov, An experience of mathematical computing forecasting of the influenza epidemics for big territory. *J. Infectology* **4** (2012), No. 3, 101–106 (in Russian).

14. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing. *Science* **220** (1983), No. 4598, 671–680.

15. O. S. Konshina, A. A. Sominina, E. A. Smorodintseva, K. A. Stolyarov, and I. Yu. Nikonorov, Population immunity to influenza virus A(H1N1)pdm09, A(H3N2) and B in the adult population of the Russian Federation long-term research results. *Russ. J. Infection Immunity* **7** (2017), No. 1, 27–33 (in Russian).

16. S. Kumar, K. Piper, D. D. Galloway, J. L. Hadler, and J. J. Grefenstette, Is population structure sufficient to generate area-level inequalities in influenza rates? An examination using agent-based models. *BMC Public Health* **15** (2015), No. 1, 947.

17. V. N. Leonenko and S. V. Ivanov, Fitting the SEIR model of seasonal influenza outbreak to the incidence data for Russian cities. *Russ. J. Numer. Anal. Math. Modelling* **31** (2016), No. 5, 267–279.

18. V. N. Leonenko, S. V. Ivanov, and Yu. K. Novoselova, A computational approach to investigate patterns of acute respiratory illness dynamics in the regions with distinct seasonal climate transitions. *Procedia Computer Science* **80** (2016), 2402–2412.

19. V. N. Leonenko, Yu. K. Novoselova, and K. M. Ong, Influenza outbreaks forecasting in Russian cities: is Baroyan–Rvachev approach still applicable? *Procedia Computer Science* **101** (2016), 282–291.

20. V. N. Leonenko and S. V. Ivanov, Influenza peaks prediction in Russian cities: comparing the accuracy of two SEIR models. *Math. Biosci. Engrg.* **15** (2018), No. 1, 209–232.

21. V. N. Leonenko and Yu. K. Novoselova, Influence of external factors on inter-city influenza spread in Russia: a modeling approach. In: *Trends in Biomathematics: Modeling, Optimization and Computational Problems*, 2018, pp. 375–389.

22. V. Leonenko and G. Bobashev, Analyzing influenza outbreaks in Russia using an age-structured dynamic transmission model. *Epidemics* **29** (2019), 100358.

23. V. Leonenko, S. Arzamastsev, and G. Bobashev, Contact patterns and influenza outbreaks in Russian cities: A proof-of-concept study via agent-based modeling. *J. Comput. Sci.* **44** (2020), 101156.

24. V. N. Leonenko and D. M. Danilenko, Modeling the dynamics of population immunity to influenza in Russian cities. *ITM Web of Conferences* **31** (2020), 03001.

25. V. N. Leonenko, Modeling co-circulation of influenza strains in heterogeneous urban populations: the role of herd immunity and uncertainty factors. In: *International Conference on Computational Science, June 2021*. Springer, Cham, pp. 663–669.

26. D. C. Liu and J. Nocedal, On the limited memory BFGS method for large scale optimization. *Math. Programming* **45** (1989), No. 1-3, 503–528.

27. N. V. Pertsev and N. V. Leonenko, Analysis of a stochastic model for the spread of tuberculosis with regard to reproduction and seasonal immigration of individuals. *Russ. J. Numer. Anal. Math. Modelling* **29** (2014), No. 5, 285–295.

28. Research Institute of Influenza, URL: http://influenza.spb.ru/en/

29. A. A. Romanyukha, T. E. Sannikova, and I. D. Drynov, The origin of acute respiratory epidemics. *Herald of the Russian Academy of Sciences* **81** (2011), No. 1, 31–34.

30. L. A. Rvachev and I. M. Longini, A mathematical model for the global spread of influenza.

Math. Biosci. **75** (1985), No. 1, 3–22.

31. N. E. Seleznev and V. N. Leonenko, Absolute humidity anomalies and the influenza onsets in Russia: a computational study. *Procedia Computer Science* **119** (2017), 224–233.
32. N. E. Seleznev and V. N. Leonenko, Boosting performance of influenza outbreak prediction framework. In: *Digital Transformation and Global Society, June 2017*. Springer, Cham, pp. 374–384.
33. J. Shaman, V. Pitzer, C. Viboud, B. T. Grenfell, and M. Lipsitch, Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biology* **8** (2010), No. 2, e1000316.
34. J. Tamerius, M. I. Nelson, S. Z. Zhou, C. Viboud, M. A. Miller, and W. J. Alonso, Global influenza seasonality: reconciling patterns across temperate and tropical regions. *Environmental Health Perspectives* **119** (2011), No. 4, 439.
35. A. I. Vlad, T. E. Sannikova, and A. A. Romanyukha, Transmission of acute respiratory infections in a city: agent-based approach. *Mathematical Biology and Bioinformatics* **15** (2020), No. 2, 338–356.
36. M. Waithaka, Likelihood-based estimation of dynamic transmission model parameters for seasonal influenza by fitting to age and season specific ILI data, 2014.