Exploring the interaction of influenza A subtypes H1N1 and H3N2 based on an evolution-driven transmission model

Gang Wang · Bing Zhang · Shiwen Liang · Feng Tang · Yang Zhao · Dechao Tian · Xiangjun Du

Received: 23 February 2022 / Accepted: 19 June 2022 / Published online: 30 June 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract Two influenza A subtypes (H1N1 and H3N2) co-circulate and pose a serious disease burden globally. It is unclear about the extent of cross-protection between these two influenza A subtypes as well as the impact of cross-protection on influenza dynamics. Based on an evolution-driven mathematical transmission model, the interaction between H1N1 and H3N2 was explored based on monthly incidence data in the USA from the years 2012 to 2019. The results showed that significant cross-protection between H1N1 and H3N2 exists. Besides, the cross-protection for H1N1 from H3N2 is much stronger than the other way around. The results also show that cross-protection between H1N1 and H3N2 has substantial influence on the epidemics. The model was also validated in Germany, Austria and Norway. A forecasting framework was proposed based on the model with feasible accuracy for both retrospective and prospective forecasts of H1N1 and H3N2 in the USA. Our findings quantify and highlight the importance of cross-protection on the co-circulating of seasonal influenza A H1N1 and H3N2.

Keywords Influenza subtypes · Antigenic changes · Cross-protection · Mathematic model · Forecast

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11071-022-07661-7.
1 Introduction

Influenza epidemics have created a serious medical burden in many regions of the world and cause significant excess mortality rates in different ages [1]. Of the two influenza types, influenza A (H1N1 and H3N2) is associated with more severe symptoms and clinical outcomes than influenza B [2]. Infection of one type/subtype can provide cross-protection against other types/subtypes mediated by high-affinity antibodies against conserved epitopes [3–5]. Virus-specific T cell can target influenza highly conserved epitopes and have potential for cross-protection [6–8]. In ferrets and mice, although infection with one subtype did not prevent infection with the other, reinfection resulted in shorter viral shedding times and reduced clinical symptoms [9, 10]. Population cohort studies have shown that cross-subtype protection exists in humans during continuous and/or concurrent epidemics caused by two influenza subtypes [11], while another study identified no heterosubtypic protection following infection [12]. Serological findings also indicate that the cross-protection may be weak and short-lived [13]. Overall, the extent of cross-protection between influenza subtypes is still controversial and needs to be quantitatively studied at the population level.

In temperate regions, different influenza subtypes share the co-seasonality under the common influence of environmental factors [14, 15]. Local outbreaks of seasonal influenza are often associated with an increased incidence of more than one influenza type or subtype [16]. Early circulation of one influenza subtype is associated with a reduced total incidence of the other subtypes, consistent with the presence of interference between subtypes [17]. The traditional time series correlation analysis methods are not suitable to this case when accounting for autocorrelation and confounding factors in time series, which may result in false positive or negative correlation [18]. Dynamic models are widely used to explore pathogen interactions at the population level, inference ability depends on the strength and time course of the underlying mechanism: stronger and more persistent interactions are easier and more precisely quantified [19–21]. Previous studies [22, 23] have used mathematical models to study the interaction of influenza viruses. However, the results may be biased due to the nonnegligible effect of the punctuated antigenic changes on influenza dynamics, especially for H3N2.

It has been demonstrated that significant differences exist for models with or without viral antigenic changes as covariates regarding either capture the temporal patterns or make forecasts [24, 25]. An evolution-informed model has already been constructed and used to make predictions for H3N2, but the interaction between H3N2 and H1N1 was simplified and neglected [24]. The theoretical modeling results illustrate a cross-protection of 50% between the two subtypes is sufficient to explain why a single influenza illness peak was observed in temperate countries [16]. Accounting for cross-immunity driven by exposures in previous outbreaks explicitly is expected to improve the accuracy of epidemiological modeling forecasts during influenza outbreaks [26]. More importantly, forecasts separated by subtype generally produced more accurate predictions and, when summed, produced more accurate predictions of total influenza incidence [27].

Based on the previous works, an evolution-driven mathematical transmission model was built to explicitly and quantitatively explore the interaction between influenza A H1N1 and H3N2 with the realistic evolutionary change information of the virus incorporated into the model in the present study. Based on the model, retrospective and prospective forecasts can be made, which will help decision makers to take timely control measures and allocate medical resources.

2 Materials and methods

2.1 Epidemiological data and virological data

Hemagglutinin (HA) protein sequences of influenza A virus H1N1 and H3N2 were downloaded from the Global Initiative on Sharing Avian Influenza Data (GISAID, https://www.gisaid.org/) and then were aligned with MUSCLE v3.7 using default settings [28]. Undetermined amino acids were replaced by gaps, and only the HA1 domain was retained for further analysis [30]. Previous studies have used mathematical models to study the interaction of influenza viruses. However, the results may be biased due to the nonnegligible effect of the punctuated antigenic changes on influenza dynamics, especially for H3N2.

It has been demonstrated that significant differences exist for models with or without viral antigenic changes as covariates regarding either capture the temporal patterns or make forecasts [24, 25]. An evolution-informed model has already been constructed and used to make predictions for H3N2, but the interaction between H3N2 and H1N1 was simplified and neglected [24]. The theoretical modeling results illustrate a cross-protection of 50% between the two subtypes is sufficient to explain why a single influenza illness peak was observed in temperate countries [16]. Accounting for cross-immunity driven by exposures in previous outbreaks explicitly is expected to improve the accuracy of epidemiological modeling forecasts during influenza outbreaks [26]. More importantly, forecasts separated by subtype generally produced more accurate predictions and, when summed, produced more accurate predictions of total influenza incidence [27].

Based on the previous works, an evolution-driven mathematical transmission model was built to explicitly and quantitatively explore the interaction between influenza A H1N1 and H3N2 with the realistic evolutionary change information of the virus incorporated into the model in the present study. Based on the model, retrospective and prospective forecasts can be made, which will help decision makers to take timely control measures and allocate medical resources.
Influenza Surveillance and Response System (GISRS). Yearly national level population estimates were downloaded from World Health Organization (WHO). The final weekly subtype-specific incidence was calculated as the product of influenza-like illness (ILI) incidence rate, proportion of ILI samples that tested positive for influenza, subtype-specific proportion and population size. Weekly incidence data were then aggregated to monthly data.

2.2 Time-series analysis

The wavelet decomposition of the time series from the years 2012 to 2019 for all subtypes was performed to disentangle periodicity in time-frequency space using the R package developed by Grinsted et al. (2004) [31].

The wavelet transform decomposes a signal using narrow functions (wavelets) when high-frequency features are present and wide functions on low-frequency structures [32]. Before wavelet decomposition, the monthly incidence of influenza A virus H1N1, H3N2, and influenza B was square-rooted. The wavelet transform decomposes a signal using narrow functions (wavelets) when high-frequency features are present and wide functions on low-frequency structures [32]. Before wavelet decomposition, the monthly incidence of influenza A virus H1N1, H3N2, and influenza B was square-rooted. The wavelet transform decomposes a signal using narrow functions (wavelets) when high-frequency features are present and wide functions on low-frequency structures [32]. Before wavelet decomposition, the monthly incidence of influenza A virus H1N1, H3N2, and influenza B was square-rooted. The wavelet transform decomposes a signal using narrow functions (wavelets) when high-frequency features are present and wide functions on low-frequency structures [32]. Before wavelet decomposition, the monthly incidence of influenza A virus H1N1, H3N2, and influenza B was square-rooted. The wavelet transform decomposes a signal using narrow functions (wavelets) when high-frequency features are present and wide functions on low-frequency structures [32]. Before wavelet decomposition, the monthly incidence of influenza A virus H1N1, H3N2, and influenza B was square-rooted. The wavelet transform decomposes a signal using narrow functions (wavelets) when high-frequency features are present and wide functions on low-frequency structures [32]. Before wavelet decomposition, the monthly incidence of influenza A virus H1N1, H3N2, and influenza B was square-rooted.

2.3 Two-subtype epidemiological model

An evolution-driven mathematical transmission model was proposed to explore the interaction between influenza A viruses H1N1 and H3N2 based on monthly incidence time series data of the USA from the years 2012 to 2019. The model considers both the transmission characteristic and antigenic evolution. The transmission model divided the population into several compartments according to their infectious and immune status (Fig. 1), where $S$ represents the susceptible population, $I_{H3}$ and $I_{H1}$ are the infected populations by H3N2 and H1N1. $R_{cH3}$ and $R_{cH1}$ are the population recovered from infection H3N2 and H1N1 and may transiently cross-protected from H1N1 and H3N2, respectively. $R_{H3}$ and $R_{H1}$ are the population completely protected from H3N2 and H1N1 without cross-protection from the other. Population $J_{H3}$ and $J_{H1}$ are infected by H3N2 and H1N1, but also immune to H1N1 and H3N2, respectively, due to repeated infection, and $R$ is recovered class with immune to both subtypes. The variables and related parameters for these two subtypes are described by corresponding subscripts (H3 and H1). Each differential equation represents the number of people flowing in and out of the compartment per unit time as indicated by the arrows in Fig. 1. The complete model is given by:

$$
\frac{dS}{dt} = \left( \mu N(t) + \frac{dN(t)}{dt} \right) - \beta(t) S \left( \frac{\xi \times I_{H3} + J_{H3}}{N(t)} \right)^\alpha - \beta(t) S \left( \frac{\xi \times I_{H1} + J_{H1}}{N(t)} \right)^\alpha - 2\tau - \mu S
$$

$$
\frac{dI_{H3}}{dt} = \beta(t) S \left( \frac{\xi \times I_{H3} + J_{H3}}{N(t)} \right)^\alpha + \tau - \frac{I_{H3}}{D} - \mu I_{H3}
$$

$$
\frac{dI_{H1}}{dt} = \beta(t) S \left( \frac{\xi \times I_{H1} + J_{H1}}{N(t)} \right)^\alpha + \tau - \frac{I_{H1}}{D} - \mu I_{H1}
$$

$$
\frac{dR_{cH3}}{dt} = \frac{I_{H3}}{D_{H3}} - c_{H3} \beta(t) R_{cH3} \left( \frac{\xi \times I_{H1} + J_{H1}}{N(t)} \right)^\alpha - \mu R_{cH3} - \frac{R_{cH3}}{P_{H3}}
$$

$$
\frac{dR_{cH1}}{dt} = \frac{I_{H1}}{D_{H1}} - c_{H1} \beta(t) R_{cH1} \left( \frac{\xi \times I_{H3} + J_{H3}}{N(t)} \right)^\alpha - \mu R_{cH1} - \frac{R_{cH1}}{P_{H1}}
$$

$$
\frac{dR_{H3}}{dt} = \frac{R_{cH3}}{P_{H3}} + \frac{R}{L_{H3}(t)} - \frac{R_{H3}}{L_{H3}(t)} - \beta(t) R_{H3} \left( \frac{\xi \times I_{H1} + J_{H1}}{N(t)} \right)^\alpha - \mu R_{H3}
$$

$$
\frac{dR_{H1}}{dt} = \frac{R_{cH1}}{P_{H1}} + \frac{R}{L_{H3}(t)} - \frac{R_{H1}}{L_{H1}(t)} - \beta(t) R_{H1} \left( \frac{\xi \times I_{H3} + J_{H3}}{N(t)} \right)^\alpha - \mu R_{H1}
$$

$$
\frac{dJ_{H3}}{dt} = c_{H1} \beta(t) R_{cH1} \left( \frac{\xi \times I_{H3} + J_{H3}}{N(t)} \right)^\alpha + \beta(t) R_{H1} \left( \frac{\xi \times I_{H3} + J_{H3}}{N(t)} \right)^\alpha - \frac{J_{H3}}{\omega \times D} - \mu J_{H3}
$$
Three aspects of cross-protection on the dynamics were considered in the model.

\[ \frac{dJ}{dt} = cH3\beta(t)R_{H3} \left( \frac{\xi \times I_{H1} + J}{N(t)} \right) - \frac{J}{\omega \times D} \]

\[ \frac{dR}{dt} = \frac{J_{H3}}{\omega \times D} + \frac{J_{H1}}{\omega \times D} - \frac{R}{L_{H3}(t)} - \mu R \]

(1) The influence of cross-protection on susceptibility: individuals infected with one subtype differ in susceptibility to the other, and the parameters \( cH3 \) and \( cH1 \) are the two parameters quanties the strength of partial cross-protection from subtypes H1N1 and H3N2, respectively. These two parameters represent the difference in the probability of reinfection with the other subtype between cross-protected and uncross-protected populations. \( P_{H3} \) and \( P_{H1} \) represent the duration of cross-protection.

(2) The influence of cross-protection on transmission: individuals with cross-protection differ in transmission compared to those without cross-protection, yielding the force of infection

\[ \lambda(t) = \beta(t) \left( \frac{\xi \times I_{H1} + J}{N(t)} \right)^\alpha \]

The parameter \( \alpha \) was used to represent the heterogeneous mixing effects on the spread of pathogen in the population. Those effects include immune response difference due to age effect and complex transmission pattern due to intrinsic property of psychosocial behavior, etc.\(^{13,33} \)

(3) The influence of cross-protection on infectious period: individuals with cross-protection differ in infectious period compared to those without cross-protection. The parameter \( D \) denotes the infectious period. The parameter \( \omega \) represents the infectious period difference between \( I \) and...
where $J_i$ is the weight to infect people [24]. The weight $\theta$ is taken to be a decaying function of the intersequence interval. The evolutionary index $E_i(t)$ was used to measure the degree of evolutionary change of the virus, the index is calculated as a weighted average distance $\bar{d}_{(s,t)}$ between the viruses in current month $t$ and those in the past two seasons that this period is close to the average time of cluster replacements [35]. The biology meaning behind $E(t)$ relates to the immune memory or protection existing in the human population against new viruses: The more similar this immune memory or protection existing in the human population against new viruses: The more similar this variant is to viruses in the past, the less likely it will be to infect people [24]. The number of births per unit time was quantified as $L_i$ where $L_i$ is the basic immune protection period and $w_i$ is the scaling factor. The evolutionary index $E_i(t)$ was calculated between the viruses in current season and those in the past two seasons due to limited sequence data, which is given by:

$$E_i(t) = \sum_{s=1}^{2} \bar{d}_{(s,t)} e^{\delta}$$

For the evolutionary index $E_i(s_0)$ for Germany, Austria and Norway, the evolutionary distance was calculated between the viruses in current season $s_0$ and those in the past two seasons due to limited sequence data, which is given by:

$$E_i(s_0) = \sum_{s=1}^{2} \bar{d}_{(s,s_0)} e^{\delta}$$

The other parameters $\tau$ and $\mu$ are the external importation rate and death rate, which were fixed to 36.5 per year and 0.015 per year, respectively, as the previous study [24]. The number of births per unit time was quantified as $\mu N(t) + dN(t)/dt$, to reproduce the natural increase of susceptible population.

### 2.4 Parameter estimation

$H_i(t)$ was defined to be the total number of new infections come to $I_i$ and $J_i$, the monthly case reports, $C_i(t)$, are normally distributed, that is,

$$C_i(t) \sim N(\rho H_i(t), \sigma_s)$$

where $\rho$ is the reporting ratio and $\sigma_s$ is the standard deviation which is different in summer (April–September) and winter (November–March in the next year). For simplicity, the variance was assumed to scale linearly with the mean, that is $\sigma_s = \sigma_s^0 \rho H_i(t)$, $\sigma_s^0$ is the scaling coefficient (s = 1 for summer and s = 2 for winter, respectively).

Both parameters and initial conditions were estimated on the basis of the likelihood function using MIF in the R package pomp [36], which is given by:

$$L(t) = \sum_{i} p_{normal}(C_i(t), \rho H_i(t), \sigma_i)$$

The priors of all the parameters were uniform distributions (Table S1). The iterated filtering (IF2) algorithm will be free to search parameter space outside the box. The maximum likelihood estimation of target parameters and its confidential interval (CI) were obtained based on the Monte Carlo adjusted profile algorithm [37].

### 2.5 Forecasts

The model was trained on the data set before the target season and the initial conditions were estimated based on exactly the same procedure described above. Next, forecasts were obtained through forward simulations for the target season starting from these initial conditions. Some external variables or drivers whose values must be specified independently from the dynamics. The projected population size $N(t)$ was produced by the United States Census Bureau. Especially for the weighted average distance $\bar{d}_{(s,t)}$, we use the mean distance of $\bar{d}_{(s,s)}$ from the months April to September in the last season and make it constant for the next season as a rough approximation in the USA.

### 2.6 Model validation and Sensitivity analysis

To establish whether the inference framework can offer valid and robust parameters estimation, parameters were re-estimated based on synthesized data that generated based on pre-defined parameters. Only the critical interaction parameters $c_{H3}$ and $c_{H1}$ were considered, and three scenarios were tested: $c_{H3} = 0.5, c_{H1} = 1$; $c_{H3} = 1, c_{H1} = 0.5; c_{H3} = 1, c_{H1} = 1$. Other parameters in the dynamics model are fixed at maximum
likelihood estimations. Based on the simulated time series, the target parameters were estimated and compared with the pre-defined values. To further test whether the model framework would detect weak cross-protection, one of the two parameters \( c_{H3} \) and \( c_{H1} \) was fixed at 1 and the value of the other parameter \( (c_{H1} \text{ or } c_{H3}) \) shifts from 0.25 to 0.95. Furthermore, we compare the shape of the log-likelihood profiles as the size of the data varies under three scenarios \((c_{H3} = 0.5, c_{H1} = 1; \; c_{H3} = 1, c_{H1} = 0.5; \; c_{H3} = 1, c_{H1} = 1)\) to investigate the precision of inference as the length of time series varies.

The difference of infectious period of population with or without cross-protection was specifically considered in the study using the scaling parameter \( \omega \). The product between parameters \( D \) and \( \omega \) makes a strong correlation and undistinguishable between them. To test if this correlation could affect the other key parameters estimation, the infectious period \( D \) was fixed at 0.04 or 0.02 year without \( \omega \) in the model. The other key parameters \((c_i, \xi)\) were randomly searched on the basis of the likelihood function.

3 Results

3.1 Data and Epidemiological model

The monthly incidence of influenza A virus H1N1, H3N2, and influenza B from the years 2012 to 2019 for the whole USA shows annual epidemic (Fig. 2A and Figure S1). The cross-correlation between H1N1 and H3N2 has the largest value at 1-year lag, which is the same as that of H1N1 and B (Figure S2). However, the cross-correlation between H3N2 and B shows the strongest correlation at 0-year lag. Since the cross-protection between influenza B and A is weak [38, 39], and the overall prevalence of influenza B is much lower than the others in the USA during the study period, only interaction of H1N1 and H3N2 was explored in the present study.

An evolution-driven mathematical transmission model was built by incorporating the interaction between H1N1 and H3N2. The transmission model divided the population into several compartments according to their infectious and immune status including the first infected \((I_{H3} \text{ and } I_{H1})\), the second infected by another virus \((J_{H3} \text{ and } J_{H1})\), the recovered that can be cross-protected \((R_{eH3} \text{ and } R_{eH1})\), and the recovered without cross-protection \((R_{H3} \text{ and } R_{H1})\). The model also specifically incorporates three aspects of cross-protection on epidemics (Fig. 1). The multi-annual variation of H1N1 and H3N2 dynamics can be simulated by the model simultaneously and accurately (Fig. 2B). The maximum likelihood estimations for all the parameters are shown in Table S1. The monthly evolutionary index \( E_i(t) \) was used as a driving factor to account for the influence of antigenic evolution on the epidemics, which is illustrated in Fig. 2C.

3.2 Evaluating the influence of cross-protection on epidemics

The inference framework was tested on synthesized data sets and the key parameters could be correctly re-estimated, no matter the strength of cross-protection is different between H1N1 and H3N2 (Figure S3). Our results show that there exists cross-protection for H1N1 from H3N2 \((c_{H3} = 0.36(0.05 \sim 0.53))\), but not the other way around \((c_{H1} = 0.98(0.55 \sim 1.23))\) (Fig. 3A and Table S1). The probability of people recovering from H3N2 infection being reinfected with H1N1 decreased to 0.36 compared to people not infected with H3N2 previously. However, populations were equally likely to be reinfected with H3N2 regardless of whether they were infected with H1N1 or not. The cross-protection from H1N1 offered by H3N2 has larger influence on the H1N1 incidence than H3N2, since the multiples of cumulative H1N1 incidence change faster and larger as the parameter \( c_{H3} \) increases (Fig. 3B). The sensitive analysis showed that the variation of parameters \( \xi \) and \( \omega \) could also affect the epidemic intensity and long-term trend of influenza epidemics obviously (Figure S4 and Figure S5). Moreover, the change of H1N1 incidence was more pronounced than H3N2 (Figure S6). The coexisting heterosubtypic and homosubtypic immunity would last for around years \( (P_{H3} : 2.78(1.68 \sim 3.28), \; P_{H1} : 1.69(1.21 \sim 2.71)) \) (Table S1), which is consistent with the finding that the cross-correlation between H1N1 and H3N2 has the largest value at 1-year lag (Figure S2). The homosubtypic immunity alone is much longer and would sustain for several years not considering the antigenic changes \((L^0_{H3} : 54.65(5.05 \sim 61.75), \; L^0_{H1} : 83.94(46.91 \sim 125.51))\) (Table S1). However, the cross-protection effect could be false negative when estimated if the magnitude
itself was weak (Figure S7). At the same time, with longer time series, the narrower 95% confident intervals can be obtained and the accuracy of inference will be improved (Figure S8).

3.3 Applying the model to other countries

In order to verify the generality of our model and the key parameters \((c_i, \xi, \omega)\), the model was applied to Germany, Austria and Norway with sufficient data, respectively, as external validation. Because of limited or biased epidemiological and antigenic data of these countries, it is not sufficient to estimate all parameter values simultaneously [40]. Therefore, key parameters obtained from the USA were constrained by letting

\[
0 < c_{H3} < 1, c_{H1} = 1, 0 < \xi < 1, 0 < \omega < 1, \text{ and the other parameters were randomly searched on the basis of the likelihood function.} \]

When calculating the evolutionary index \(E_i(s_0)\), the average evolutionary distance between seasons was calculated and used due to lack of sequence data in the summer (Figure S9).

The model is capable to capture the dynamics of H1N1 and H3N2 in Germany, Austria and Norway (Fig. 4A, 4B and 4C, respectively). Monthly predictions and observations are significantly correlated (Figure S10). The evolutionary index of H3N2 is often greater than that of H1N1 in either the USA (Fig. 2D), Germany, Austria or Norway (Fig. 4A, 4B and 4C, respectively), which causes more rapid loss of homosubtypic immunity in the model.
3.4 Retrospective and prospective forecast

To test the predicting ability, the model was trained on the data set before the target season and tested on the data set after the season 2015/2016. Because the model is nonautonomous, some covariates should be specified manually and independently before making predictions. For the weighted average distance $d_i(s,t)$, the mean distance of $d_i(s,t)$ from the months April to September in the last season was used and keep it constant for the next season as a rough approximation. The retrospective forecasts of 2016/2017, 2017/2018, 2018/2019 and a prospective forecast 2019/2020 show an increasing trend for H1N1 and a decreasing trend for H3N2. Most observations fall within the 95% uncertainty intervals, which matches well with historical data except for A(H3N2) in 2018/2019 (Fig. 5A). There are significant correlations between predictions and observations of H1N1 and H3N2 ($r = 0.85$ and $0.54$, respectively, with both $P < 0.01$) from the retrospective and prospective forecasts (Fig. 5B). Our results demonstrate the feasibility of epidemiological forecasts by incorporating the interaction between H1N1 and H3N2, the transmission characteristic and antigenic evolution.

4 Discussion

An evolution-driven mathematical transmission model that specifically incorporates antigenic evolution and cross-protection between subtypes was proposed in this study for exploration of interaction between H1N1 and H3N2 based on data in the USA from the years 2012 to 2019. Key parameters and quantitative description of strength of the interaction between them were revealed. A forecasting approach was also proposed based on the model and the forecast feasibility was tested. Our findings highlight the importance of cross-protection on the co-circulating of H1N1 and H3N2.

The model is capable to simulate the epidemic trend and relative epidemic intensity of H1N1 and H3N2 not only in the USA but also in other countries. There exists significant cross-protection for H1N1 from H3N2 ($c_{H1} : 0.36(0.05 \sim 0.53)$), which is consistent with the general perception that H3N2 outcompetes H1N1, as a result, causes more frequent and larger epidemics [41]. The cross-protection for H3N2 from H1N1 was not detected in this study ($c_{H1} : 0.98(0.55 \sim 1.23)$), but sensitivity analysis indicates that it could be much weaker than the other way around if it existed (Figure S7). Furthermore, individuals without cross-protection contribute less to
influenza transmission than individuals with cross-protection \( (\xi : 0.01(0.00 - 0.41)) \), which can probably be explained by much more health-seeking behavior and more intense self-isolation than individuals with cross-protection [42–44], who have shorter infectious period as well \( (\omega : 0.32(0.29 - 0.59)) \). Further sensitivity analysis showed that the direction and strength of the cross-protection between H1N1 and H3N2 is not influenced by the other assumptions (Table S2).

Fig. 4 Applying the model to other countries. Monthly incidence and evolutionary change \( E(t) \) for Germany A, Austria B, and Norway C, respectively. The black curve is the monthly observed incidence. The colored curve is the predicted monthly median incidence with shaded 95% uncertainty intervals (2.5 and 97.5% quantiles) from 5000 random simulations. The evolutionary index \( E_i(s_0) \) was calculated at the scale of season due to limited sequence data (check Materials and Methods for more details).

A forecasting approach was proposed based on the evolution-driven mathematical model to make predictions for H1N1 and H3N2 simultaneously ahead of season. Earlier forecasts of incidence dynamics can aid policy makers to take better preventive and control measures, plan and allocate of resource, and develop vaccines. The prediction of incidence dynamics depends on the prediction of the evolution index. However, due to the discontinuity of influenza surveillance data in many countries, the model cannot be applied to more countries and used to make
predictions. The influenza virus surveillance should be further strengthened to provide a data basis for exploration of influenza subtypes epidemic patterns and make earlier forecasts available.

Several limitations should be acknowledged. First, the compartmental model did not include age structure for lack of data in different age groups. People of different ages have different immune responses to natural infection. The immune imprint left by the initial infection lasts for life and affects the immune response to later infection with antigen-similar viruses [13, 45, 46]. Second, it is not always appropriate to calculate hamming distance based on mutations in HA to measure the degree of antigen changes. Taking into account variations in other parts of the influenza virus and changes in the influenza phenotype will help us to better measure the extent of antigenic changes [47–49]. Third, we considered only the immune

Fig. 5 Incidence forecasts for the USA. A Retrospective forecasts (for influenza season from 2016/2017 to 2018/2019, in green), and a prospective forecast for the 2019/2020 influenza season (in orange). The black curve is the monthly observed incidence; the colored curve is the predicted monthly median incidence with shaded 95% uncertainty intervals (2.5 and 97.5% quantiles) from 5000 random simulations with the maximum likelihood estimations. B Spearman rank’s correlations were computed between monthly median predictions and observations of H1N1 and H3N2 from the retrospective and retrospective forecasts as Fig. 4A. The error bars represent 95% uncertainty intervals of the predictions.
response caused by natural infection and did not distinguish the difference between natural infection and vaccine induced immune response [50, 51]. Fourth, the model consistently underestimates cumulative incidence especially for H3N2 epidemics although the observed incidence lies within the 95% uncertainty intervals. The underestimation of epidemics could result from combinations of many factors as mentioned above, such as population heterogeneity, measurement of antigenic changes and vaccination, which should be investigated further.

**Author Contribution** XD designed the study. GW, SL and FT collected the data. GW, BZ and SL built the model. GW performed the analysis. XD, GW, BZ, SL, YZ and DT interpreted the data. XD and GW prepared the manuscript. XD, GW, BZ, YZ and DT edited the paper. All authors reviewed and approved the submitted manuscript.

**Funding** This research was funded by Shenzhen Science and Technology Program under grant KQTD2018041114332605, National Natural Science Foundation of China under Grants 31970643 and 816128002, the Guangdong Frontier and Key Tech Innovation Program under grants 2021A111112007, 2019B020228001 and 2019B11103001, the Natural Science Foundation of Guangdong Province, China, under grant 2021A1515011592.

**Data and code availability** https://github.com/DuLab-SYSU/FluSubtypes_Evolution_Driven_Model

**Declarations**

**Conflict of interest** The authors declare no competing interests.

**References**

1. Iuliano, A.D., et al.: Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. Lancet (London, England). **391**, 1285–1300 (2018)
2. Chen, L., Han, X., Li, Y., Zhang, C., Xing, X.: Severity and outcomes of influenza-related pneumonia in type A and B strains in China, 2013–2019. Infect. Dis. Poverty **9**, 42 (2020)
3. Ekiert, D.C., et al.: Antibody recognition of a highly conserved influenza virus epitope. Science (New York, N.Y.). **324**, 246–251 (2009)
4. Dreyfus, C., et al.: Highly conserved protective epitopes on influenza B viruses. Science (New York, N.Y.). **337**, 1343–1348 (2012)
5. Ekiert, D.C., et al.: A highly conserved neutralizing epitope on group 2 influenza a viruses. Science (New York, N.Y.). **333**, 843–850 (2011)
6. Reber, A.J., et al.: Extensive T cell cross-reactivity between diverse seasonal influenza strains in the ferret model. Sci. Rep.-UK **8**, 6112 (2018)
7. Gautam, A., et al.: Peritoneal cells mediate immune responses and cross-protection against influenza a virus. Front. Immunol. **10**, 1160 (2019)
8. Dong, W., et al.: Cross-protective immune responses induced by sequential influenza virus infection and by sequential vaccination with inactivated influenza vaccines. Front. Immunol. **9**, 2312 (2018)
9. Gooch, K.E., et al.: Heterosubtypic cross-protection correlates with cross-reactive interferon-gamma-secreting lymphocytes in the ferret model of influenza. Sci. Rep.-UK **9**, 2617 (2019)
10. Kreijtz, J.H.C.M., et al.: Primary influenza a virus infection induces cross-protective immunity against a lethal infection with a heterosubtypic virus strain in mice. Vaccine. **25**, 612–620 (2007)
11. Sonoguchi, T., Naito, H., Hara, M., Takeuchi, Y., Fukumi, H.: Cross-subtype protection in humans during sequential, overlapping, and/or concurrent epidemics caused by H3N2 and H1N1 influenza viruses. J. Infect. Dis. **151**, 81–88 (1985)
12. Cowling, B. J. et al.: Incidence of influenza virus infections in children in Hong Kong in a 3-year randomized placebo-controlled vaccine study, 2009–2012. **517–524** (2014)
13. Ranjeva, S., et al.: Age-specific differences in the dynamics of protective immunity to influenza. Nat. Commun. **10**, 1660 (2019)
14. Deyle, E.R., Maher, M.C., Hernandez, R.D., Basu, S., Sugihara, G.: Global environmental drivers of influenza. Proc. Natl. Acad. Sci. **113**, 13081–13086 (2016)
15. Dalziel, B.D., et al.: Urbanization and humidity shape the intensity of influenza epidemics in U.S. cities. Science (New York, N.Y.) **362**, 75–79 (2018)
16. Lavenu, A., Valleron, A., Carrat, F.: Exploring cross-protection between influenza strains by an epidemiological model. Virus Res. **103**, 101–105 (2004)
17. Goldstein, E., Cobey, S., Takahashi, S., Miller, J.C., Lipsitch, M.: Predicting the epidemic sizes of influenza A/H1N1, A/H3N2, and B: a statistical method. PLoS Med. **8**, e1001051 (2011)
18. Nickbakhsh, S., et al.: Virus–virus interactions impact the population dynamics of influenza and the common cold. Proc. Natl. Acad. Sci. **116**, 27142–27150 (2019)
19. Shrestha, S., et al.: Identifying the interaction between influenza and pneumococcal pneumonia using incidence data. Sci. Transl. Med. **5**, 184r–191r (2013)
20. Shrestha, S., King, A.A., Rohani, P.: Statistical inference for multi-pathogen systems. PLoS Comput. Biol. **7**, e1002135 (2011)
21. Shrestha, S., et al.: The role of influenza in the epidemiology of pneumonia. Sci. Rep.-UK **5**, 15314 (2015)
22. Yang, W., Lau, E.H.Y., Cowling, B.J.: Dynamic interactions of influenza viruses in Hong Kong during 1998–2018. PLoS Comput. Biol. **16**, e1007989 (2020)
23. Gog, J.R., Grenfell, B.T.: Dynamics and selection of many-strain pathogens. P. Natl. Acad. Sci. USA. 99, 17209–17214 (2002)
24. Du, X., King, A.A., Woods, R.J., Pascual, M.: Evolution-informed forecasting of seasonal influenza A (H3N2). Sci. Transl. Med. 9, 11 (2017)
25. Axelsen, J.B., Yaari, R., Grenfell, B.T., Stone, L.: Multiannual forecasting of seasonal influenza dynamics reveals climatic and evolutionary drivers. P. Natl. Acad. Sci. USA. 111, 9538–9542 (2014)
26. Sachak-Patwa, R., Byrne, H.M., Thompson, R.N.: Accounting for cross-immunity can improve forecast accuracy during influenza epidemics. Epidemics-Neth. 34, 100432 (2020)
27. Kandula, S., Yang, W., Shaman, J.: Type- and subtype-specific influenza forecast. Am. J. Epidemiol. 185, 395–402 (2017)
28. Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. 1792–1797 (2004)
29. 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 289, 373–378 (1981)
30. Price, M.N., Dehal, P.S., Arkin, A.P.: FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS ONE 5, e9490 (2010)
31. Grinsted, A., Moore, J.C., Jevrejeva, S.: Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Proc. Geophys. 11, 561–566 (2004)
32. Cazelles, B., Chavez, M., Magny, G.C.D., Guégan, J., Hales, S.: Time-dependent spectral analysis of epidemiological time-series with wavelets. J. R. Soc. Interface. 4, 625–636 (2007)
33. Scarpino, S.V., Allard, A., Hébert-Dufresne, L.: The effect of a prudent adaptive behaviour on disease transmission. Nat. Phys. 12, 1042 (2016)
34. Bhadra, A., Ionides, E.L., Laneri, K., Pascual, M., Dhiman, R.C.: Malaria in Northwest India: data analysis via partially observed stochastic differential equation models driven by Ivy noise. J. Am. Stat. Assoc. 106, 440–451 (2011)
35. Du, X., et al.: Mapping of H3N2 influenza antigenic evolution in China reveals a strategy for vaccine strain recommendation. Nat. Commun. 3, 709 (2012)
36. Ionides, E.L., Breto, C., King, A.A.: Inference for nonlinear dynamical systems. P. Natl. Acad. Sci. USA. 103, 18438–18443 (2006)
37. Ionides, E.L., Breto, C., Park, J., Smith, R.A., King, A.A.: Monte Carlo profile confidence intervals for dynamic systems. J. Royal Soc. Interf. 14, 132 (2017)
38. Möst, J., Weiss, G.: Consecutive infections with influenza a and b virus in children during the 2014–2015 seasonal influenza epidemic. J. Infect. Dis. 214, 1139–1141 (2016)
39. Le Sage, V., et al.: Pre-existing heterosubtypic immunity provides a barrier to airborne transmission of influenza viruses. PLoS Pathog. 17, e1009273 (2021)
40. D’Agostino McGowan, L., Grantz, K. H. & Murray, E.: Quantifying uncertainty in infectious disease mechanistic models. Am. J. Epidemiol (2021).
41. Bedford, T., et al.: Global circulation patterns of seasonal influenza viruses vary with antigenic drift. Nature 523, 217–220 (2015)
42. Baltrusaitis, K. et al.: Health-care seeking behavior for respiratory illness among flu near you participants in the United States during the 2015–16 through 2018–19 influenza season. J. Infect. Dis, (2020).
43. Ma, W., Huo, X., Zhou, M.: The healthcare seeking rate of individuals with influenza like illness: a meta-analysis. Infect. Dis. 50, 728–735 (2018)
44. Peppa, M., John Edmunds, W., Funk, S.: Disease severity determines health-seeking behaviour amongst individuals with influenza-like illness in an internet-based cohort. BMC Infect. Dis. 17, 238 (2017)
45. Meade, P., Kuan, G., Strohmeier, S., Maier, H.E., Amanat, F., Balmaseda, A., Kramer, F.: Influenza virus infection induces a narrow antibody response in children but a broad recall response in adults. MBio 11(1), e03243-19 (2020)
46. Gouma, S., et al.: Middle-aged individuals may be in a perpetual state of H3N2 influenza virus susceptibility. Nat. Commun. 11, 4566 (2020)
47. Neher, R.A., Bedford, T., Daniels, R.S., Russell, C.A., Shraiman, B.I.: Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses. P. Natl. Acad. Sci. USA. 113, E1701–E1709 (2016)
48. Li, C., et al.: Selection of antigenically advanced variants of seasonal influenza viruses. Nat. Microbiol. 1, 16058 (2016)
49. Petrova, V.N., Russell, C.A.: The evolution of seasonal influenza viruses. Nat. Rev. Microbiol. 16, 47–60 (2018)
50. Arinaminpathy, N., et al.: Estimating direct and indirect protective effect of influenza vaccination in the United States. Am. J. Epidemiol. 186, 92–100 (2017)
51. Backer, J.A., van Boven, M., van der Hoek, W., Wallinga, J.: Vaccinating children against influenza increases variability in epidemic size. Epidemics-Neth. 26, 95–103 (2019)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.