Models for influenza surveillance and early warning

Wang Dayan

WHO Collaborating Center for Reference and Research on Influenza
National Institute for Viral Disease Control and Prevention, China CDC

Hanoi, Vietman
29-31, May, 2012
Models

\[ q = C v(t) \]

\[ i = \frac{\partial q}{\partial t} = \frac{q_{k+1} - q_k}{t_{k+1} - t_k} \]

\[ i = C \frac{\partial v}{\partial t} \approx C_k \frac{v_{k+1} - v_k}{t_{k+1} - t_k} \]

\[ \psi(x) \rightarrow e^{i\alpha(x)} \psi(x) \]

\[ \frac{\partial}{\partial x} \psi(x) \rightarrow e^{i\alpha(x)} \frac{\partial}{\partial x} \psi(x) + e^{i\alpha(x)} \]

CHINESE CENTER FOR DISEASE CONTROL AND PREVENTION
Influenza Virus

Genetic composition of a flu virus

Negative single-stranded RNA virus.

8 RNA segments encoding 11 proteins:
- NP, PA, PB1 and PB2 code for viral replication machinery
- NS for non-structural proteins 1 and 2 (NS-1, 2),
- MP for matrix proteins 1 and 2 (M-1 and 2).
- HA for hemagglutinin
- NA for neuraminidase

The surface glycoprotein, HA, controls viral entry into the cells and is the major antigenic target of the host antibody responses.
The complexity of influenza A viruses

HA: 1-16 H17
NA: 1-9
Influenza A viruses have a broad host-range
Prediction and early warning is the key step of infection prevention and control

The critical role of epidemic prediction for the selection of vaccine strains

Early warning for pandemic?

CHINESE CENTER FOR DISEASE CONTROL AND PREVENTION
Difficult of vaccine recommendation lies in the rapid antigenic drift

Flu virus mainly uses two strategies to change its antigenicity to escape host immune surveillance.

1. Gene mutation

2. Gene reassortment

Therefore, rapidly and accurately capture the antigenic variation is key to the vaccine recommendation.
Technical difficulties for vaccine recommendation

- Limitation of traditional method
  - HI experiment: not so sensitive, inconsistency
  - MN experiment: complicated, not so accurate

- Long period for vaccine production
## Challenge for vaccine recommendation: vaccine mismatch

<table>
<thead>
<tr>
<th>Season</th>
<th>Vaccine strain (Predominant virus)</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>96-97</td>
<td>Nanchang/933/95 (Nanchang/933/95)</td>
<td>++++</td>
</tr>
<tr>
<td>97-98</td>
<td>Nanchang/933/95 (Sydney/5/97)</td>
<td>+</td>
</tr>
<tr>
<td>98-99</td>
<td>Sydney/5/97 (Sydney/5/97)</td>
<td>++++</td>
</tr>
<tr>
<td>99-00</td>
<td>Sydney/5/97 (Sydney/5/97)</td>
<td>++++</td>
</tr>
<tr>
<td>01-02</td>
<td>Panama/2007/99 (Panama/2007/99)</td>
<td>++++</td>
</tr>
<tr>
<td>02-03</td>
<td>Panama/2007/99 (Panama/99+Fujian/02)</td>
<td>++++</td>
</tr>
<tr>
<td>03-04</td>
<td>Panama/2007/99 (Fujian/411/02)</td>
<td>++</td>
</tr>
<tr>
<td>04-05</td>
<td>Fujian/411/02 (WY/3/03) (Fujian/411/02)</td>
<td>++++</td>
</tr>
<tr>
<td>05-06</td>
<td>CA/7/04 (NY/55/04) (California/7/04)</td>
<td>++++</td>
</tr>
</tbody>
</table>

*From USCDC 1996-2006*
✓ New techniques
✓ Models
The method readily allows monitoring of antigenic differences among vaccine and circulating strains and thus estimation of the effects of vaccination.
Cartography

H3N2 colored by clade
1 = blue, 2 = lightblue, 3A = pink, 3B = red, 3C = maroon, 5 = yellow, 6 = green
rapid and high-quality of HAs or even whole genomes of influenza viruses in influenza surveillance

development of sequence-based computational approaches became an indispensable effort to understand the antigenic properties and characteristics of influenza virus evolution

Sequence data
Traditional methods for predicting antigenic variation with seq-1

Counting amino acid changes on HA, particularly those in the epitope regions or based on phylogenetic techniques.

Wilson et al. 1990, Annu Rev Immunol 8, 737
Lee et al. 2004, Emerg Infect Dis 10, 1385
Munoz et al. 2005, Vaccine 23, 1114
Gupta et al. 2006, Vaccine 24, 3881
Lee et al. 2007, Vaccine 25, 8133
Traditional methods for predicting antigenic variation with seq-2

<table>
<thead>
<tr>
<th>Position</th>
<th>Antigenic domain</th>
<th>Residue frequency among the 45 viruses in the training dataset</th>
<th>GM4</th>
<th>GM5</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA82</td>
<td>E</td>
<td>17 (E), 28 (K)</td>
<td>0.998</td>
<td>1.037</td>
</tr>
<tr>
<td>AA92</td>
<td>E</td>
<td>1 (E), 44 (K)</td>
<td>0.941</td>
<td>0.920</td>
</tr>
<tr>
<td>AA121</td>
<td>D</td>
<td>28 (I), 10 (N), 7 (T)</td>
<td>0.495</td>
<td>0.546</td>
</tr>
<tr>
<td>AA124</td>
<td>A</td>
<td>13 (D), 20 (G), 2 (N), 10 (S)</td>
<td>0.298</td>
<td></td>
</tr>
<tr>
<td>AA129</td>
<td>B</td>
<td>1 (E), 44 (G)</td>
<td>1.748</td>
<td></td>
</tr>
<tr>
<td>AA135</td>
<td>A</td>
<td>1 (E), 23 (G), 6 (K), 15 (T)</td>
<td>0.954</td>
<td>1.021</td>
</tr>
<tr>
<td>AA144</td>
<td>A</td>
<td>13 (D), 3 (I), 5 (N), 24 (V)</td>
<td>0.716</td>
<td>0.683</td>
</tr>
<tr>
<td>AA145</td>
<td>A</td>
<td>1 (I), 18 (K), 22 (N), 1 (R), 1 (S)</td>
<td>1.209</td>
<td>1.282</td>
</tr>
<tr>
<td>AA155</td>
<td>B</td>
<td>30 (H), 2 (T), 13 (Y)</td>
<td>1.202</td>
<td>1.582</td>
</tr>
<tr>
<td>AA156</td>
<td>B</td>
<td>8 (E), 1 (H), 27 (K), 9 (Q)</td>
<td>0.400</td>
<td>0.294</td>
</tr>
<tr>
<td>AA157</td>
<td>B</td>
<td>26 (L), 19 (S)</td>
<td>0.423</td>
<td>0.448</td>
</tr>
<tr>
<td>AA158</td>
<td>B</td>
<td>29 (E), 7 (G), 9 (K)</td>
<td>0.761</td>
<td>0.715</td>
</tr>
<tr>
<td>AA160</td>
<td>B</td>
<td>1 (A), 33 (K), 1 (R), 1 (S), 7 (T)</td>
<td>1.072</td>
<td>1.073</td>
</tr>
<tr>
<td>AA173</td>
<td>D</td>
<td>34 (K), 11 (N)</td>
<td>1.285</td>
<td>1.301</td>
</tr>
<tr>
<td>AA174</td>
<td>D</td>
<td>40 (F), 4 (S), 1 (V)</td>
<td>0.613</td>
<td>0.633</td>
</tr>
<tr>
<td>AA188</td>
<td>B</td>
<td>42 (D), 1 (E), 1 (N), 1 (V)</td>
<td>1.087</td>
<td>1.234</td>
</tr>
<tr>
<td>AA189</td>
<td>B</td>
<td>8 (K), 5 (Q), 8 (R), 24 (S)</td>
<td>0.721</td>
<td>0.684</td>
</tr>
<tr>
<td>AA240</td>
<td>D</td>
<td>44 (G), 1 (R)</td>
<td>0.690</td>
<td>0.708</td>
</tr>
<tr>
<td>AA273</td>
<td>C</td>
<td>44 (P), 1 (S)</td>
<td>0.779</td>
<td>0.738</td>
</tr>
<tr>
<td>AA276</td>
<td>C</td>
<td>9 (K), 14 (N), 22 (T)</td>
<td>1.830</td>
<td>2.287</td>
</tr>
</tbody>
</table>

Agreement rate in the training dataset (N=181) 93.37% 92.82%
Agreement rate in the validation dataset (N=96) 91.67% 91.67%

Considering the specific amino acid changes at certain positions on HA observed during influenza evolution (site-dependent models).

Liao et al. 2008, Bioinformatics 24(4), 505
Huang et al. 2009, BMC Bioinformatics 10 (Sup 1), S41
Lees et al. 2010, Bioinformatics
Mapping of H3N2 influenza antigenic evolution in China reveals a strategy for vaccine strain recommendation

Xiangjun Du\textsuperscript{1,2,*}, Libo Dong\textsuperscript{3,*}, Yu Lan\textsuperscript{3}, Yousong Peng\textsuperscript{1,2}, Aiping Wu\textsuperscript{1}, Ye Zhang\textsuperscript{3}, Weijuan Huang\textsuperscript{3}, Dayan Wang\textsuperscript{3}, Min Wang\textsuperscript{3}, Yuanji Guo\textsuperscript{3}, Yuelong Shu\textsuperscript{3} & Taijiao Jiang\textsuperscript{1}

Chinese Academy of Sciences
National Institute for Viral Disease Control and Prevention, Chinese CDC
Development of data-mining methods for exploring the evolutionary mechanisms of influenza virus
Variation of antigen result from the change of interaction between HA and antibody
Development of PREDAC for modeling the antigenic evolution of the H3N2 virus

machine-learning model using a Naive Bayes classifier to integrate the structural and physicochemical features of HA
Systematic antigenic evolution of influenza H3N2 virus in China based on PREDAC

1. China CDC carried out a large-scale sequencing of HA from 1071 influenza A (H3N2) viruses isolated between 1968 and 2009 from a variety of representative regions in mainland China.
2. Phylogenetic analysis is unable to delineate its antigenic evolution.
3. Applied PREDAC to construct an ACnet and predict antigenic clusters for these influenza A (H3N2) viruses.
4. In total 20 antigenic clusters, 17 dominant antigenic clusters, which included ~99% of the viruses, and each persisted for one to five seasons.
5. The predicted antigenic clusters vividly describe the influenza antigenic evolution patterns in China.
Automatic vaccine strain recommendations based on the H3N2 viruses monitored in China
Threshold percentage and prediction date affect vaccine strain predictions

<table>
<thead>
<tr>
<th>Date/threshold</th>
<th>5%</th>
<th>10%</th>
<th>15%</th>
<th>20%</th>
<th>25%</th>
<th>30%</th>
<th>40%</th>
<th>50%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dec 31</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Jan 15</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Jan 30</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Feb 15</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Feb 28</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Mar 15</td>
<td>5</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Mar 30</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Apr 15</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Apr 30</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
</tr>
</tbody>
</table>

The predictions were made on data from China at different dates and threshold percentages.
Systematic antigenic evolution of influenza H1N1 virus in China based on PREDAC
Systematic antigenic evolution of influenza H5N1 virus in China based on PREDAC
Model for predicting mortality burden caused by influenza virus based on HA sequence

Correlation of Influenza Virus Excess Mortality with Antigenic Variation: Application to Rapid Estimation of Influenza Mortality Burden

Aiping Wu¹, Yousong Peng¹,²,³, Xiangjun Du¹,², Yuelong Shu³, Taijiao Jiang¹*

Mortality burden of a newly emerged virus

CHINESE CENTER FOR DISEASE CONTROL AND PREVENTION
Evolution of a human influenza virus, is considered as serial replacements of the antigenic strains driven by the cross immunity induced by the previous prevalent antigenic strains of same (sub)type in the human population.

Given that the extent of cross protection between two viruses depends largely on the antigenic distance between them, it is likely that the antigenic distances between a new antigenic strain and the previous antigenic strains largely determine the mortality burden of the new strain on the human population.

Simplified host-virus interaction model
Correlation analysis of influenza virus excess mortality and antigenic variation

Table 2. The Spearman and Pearson Correlation Coefficients between the excess all-cause mortalities and the integrated antigenic distances relative to the previous 1-5 antigenic strains as background strains.

<table>
<thead>
<tr>
<th>Virus (sub)type</th>
<th>Correlation method</th>
<th>No. of background strains</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>A(H1N1)</td>
<td>Spearman</td>
<td>0.64(0.14)</td>
</tr>
<tr>
<td></td>
<td>Pearson</td>
<td>0.79(0.03)</td>
</tr>
<tr>
<td>A(H3N2)</td>
<td>Spearman</td>
<td>0.71(0.004)</td>
</tr>
<tr>
<td></td>
<td>Pearson</td>
<td>0.58(0.03)</td>
</tr>
<tr>
<td>B</td>
<td>Spearman</td>
<td>0.26(0.46)</td>
</tr>
<tr>
<td></td>
<td>Pearson</td>
<td>0.36(0.31)</td>
</tr>
</tbody>
</table>

The numbers in parenthesis indicate the P-values of corresponding coefficients. The largest coefficient for each (sub)type is highlighted in bold.

*Not applicable due to limited number of antigenic strains.

d_{ij} = \ln \left( \frac{\sqrt{h_i h_j}}{H_i H_j} \right)

Archetti-Horsfall method
The quantitative relationship between the excess all-cause mortality and antigenic distance

Figure 2. The quantitative relationship between the excess all-cause mortality and antigenic distance for A(H1N1). (A) The nonparametric (the dashed line) and ordinary linear (the black line) regression between the excess all-cause mortalities caused by A(H1N1) antigenic strains and their integrated antigenic distances to the previous first and second antigenic strains. The nonparametric regression is done using the local polynomials method (called loess method [11]). (B) The observed and estimated excess all-cause mortality for recent seasonal A(H1N1) virus A/Brisbane/59/2007. The error bar shows the standard deviation of the prediction.
doi:10.1371/journal.pcbi.1000882.g002
Development of a computational method for predicting antigenic distance: \textbf{(EADpred, Epitope-based antigenic distance prediction)}

Correlation between calculated antigenic distance (with EADpred) and excess mortality

Enable an estimation as early as possible of excess mortalities for emerging antigenic strains
China’s new WHO flu monitoring center seeks to reverse criticism

BEIJING — China has not always been a world leader when it comes to infectious disease surveillance. Severe acute respiratory syndrome caught the country by surprise in 2003, and, two years later, government officials went into denial after reports surfaced that H5N1 avian influenza had infected people and birds. But since those decades, China has ramped up its screening efforts, building several infectious-disease institutes and more than 400 labs devoted to flu surveillance and testing, plus adding sentinel equipment to some 550 hospitals. So when H1N1 ‘swine flu’ struck four years later, the world’s most populous country was much better prepared.

“China has set up the world’s largest influenza surveillance network,” Yuelong Shu, director of the National Influenza Center, part of the Chinese Center for Disease Control and Disease Control and Prevention. “Now China is quite open to the outside.”

On a domestic level, Liu also thinks that the National Influenza Center’s status as a WHO organ—complete with data sharing between the five global centers—will promote a spirit of greater cooperation among the Chinese research community. “The reference center will help Chinese institutions and pharmaceutical companies get influenza viral strains and other data more quickly and more comprehensively, facilitating their work to develop vaccines and therapies against particular viral strains,” he says.

Even so, gaps still remain in China’s monitoring. According to a recent report by leading officials at the Chinese CDC and its US counterpart. China still needs to invest more in pathogen-based surveillance systems
Acknowledgement

- Shu’s group (CNIC)
- Yu Lan
- Libo Dong
- Lei Yang
- Yuelong Shu
- … …

- Jiang’s Group (CAS)
- Xiangjun Du
- Aiping Wu
- Yousong peng
- Yang Cao
- Taijiao Jiang
- … …
请各位专家批评指正