On the use of hemagglutination-inhibition for influenza surveillance: Surveillance data are predictive of influenza vaccine effectiveness

Wilfred Ndifon*, Jonathan Dushoff1, Simon A. Levin

Department of Ecology & Evolutionary Biology, Princeton University, 106 Guyot Hall, Princeton, NJ 08544, United States

**A B S T R A C T**

The hemagglutination-inhibition (HI) assay is the main tool used by epidemiologists to quantify antigenic differences between circulating influenza virus strains, with the goal of selecting suitable vaccine strains. However, such quantitative measures of antigenic difference were recently shown to have poor predictive accuracy with respect to influenza vaccine effectiveness (VE) in healthy adults. Here, we re-examine those results using a more rigorous criterion for predictive accuracy – considering only cases when the vaccine (V) and dominant (D) circulating strains are antigenically different – and greater numbers of HI titers. We find that the Archetti–Horsfall measure of antigenic difference, which is based on both the normalized HI titer (NHI) of V relative to antisera raised against D and the NHI of D relative to V, predicts VE very well ($R^2 = 0.90, p = 2.5 \times 10^{-10}$). In contrast, the predictive accuracies of the NHI of D relative to V alone ($R^2 = 0.01$, and two other measures of antigenic difference based on the amino acid sequence of influenza virus hemagglutinin ($R^2 = 0.03$ for both measures) are relatively poor. Furthermore, while VE in the elderly is generally high in cases when D and V are antigenically identical (VE = 35%, S.E. = 5%), in other cases VE appears to increase with the antigenic difference between D and V ($R^2 = 0.90, p = 2.5 \times 10^{-10}$). This paradoxical observation could reflect the confounding effects of prior immunity on estimates of VE in the elderly. Together, our results underscore the need for consistently accurate selection of suitable vaccine strains. We suggest directions for further studies aimed at improving vaccine-strain selection and present a large collection of HI titers that will be useful to such studies.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

For at least a century, influenza A and B viruses have caused regular epidemics and occasional pandemics that have a significant negative impact on public health. On an almost annual basis, antigenically novel “drift” strains of these viruses emerge (primarily by means of mutations to the viral hemagglutinin protein), infecting about 5–15% of the world’s population and causing a large number of excess deaths [1,2]. Occasionally, antigenically novel “shift” strains of influenza A viruses capable of efficient human-to-human transmission emerge (via mutation of avian-adapted strains and/or exchange of viral segments among human- and avian-adapted strains), giving rise to pandemics characterized by significantly elevated rates of infection and of death [3]. The most deadly pandemic in recorded history is that of 1918, which killed tens of millions of individuals worldwide. The recent emergence of minimally transmission-competent and highly pathogenic shift strains of the influenza H5N1 virus subtype has raised awareness concerning the danger of a future pandemic [4].

To help forecast the emergence of, and assist in developing vaccines against, both epidemic and pandemic influenza virus strains, the World Health Organization oversees an extensive influenza surveillance network comprising researchers from over 80 countries [5,6]. The main tool used by these researchers to quantify (antigenic) changes to the hemagglutinin of circulating influenza viruses could have deleterious consequences for influenza vaccine effectiveness (VE) is the hemagglutination-inhibition (HI) assay [7]. However, measures of antigenic changes based on HI assay results (i.e., HI titers) were recently criticized for having limited predictive accuracy with respect to VE, and alternative measures were proposed [8–10]. This criticism calls for improved ways of interpreting HI titers, especially in light of the limited effectiveness of influenza vaccines administered in recent seasons (e.g., Refs. [11,12]). To shed light on the possible reasons for the reportedly poor predictive accuracy of HI titers, we describe briefly the conceptual basis of the HI assay.

The HI assay takes advantage of two facts about the biology of influenza viruses [13]: (i) the ability of an influenza virus to bind simultaneously to hemagglutinin receptors found on more than one red blood cell, causing the bound cells to become agglutinated, and
(ii) the fact that infection of an individual by a given influenza virus strain induces the production of antibodies with differing degrees of specificity for different strains. The degree to which antisera extracted from individuals infected by one strain (the “infecting” or “homologous” strain) prevent another strain (the “heterologous” strain) from agglutinating red blood cells – the heterologous HI titer – is used to measure the antigenic difference between the two virus strains. However, HI titers depend on several properties of virus strains that are not directly related to antigenic difference, including the capacity of strains to induce the production of antibodies [14]; the avidity of strains for red blood cells; and experimental conditions (i.e., temperature, pH, etc.). In addition, the concentrations of induced antibodies with specificity for heterologous strains are sometimes small, suggesting that fluctuations in those concentrations could be an important source of error in measured HI titers. We further address these issues in the next paper of this series.

In fact, HI titers measured in independent assays can be quite variable. For example, in three assays that compared the two influenza H3N2 virus strains A/Moscow/10/1999 and A/Sydney/5/97, the normalized HI titers (that is, the ratio of the heterologous HI titer to the homologous HI titer) of A/Moscow/10/1999 relative to antisera derived from ferrets infected by A/Sydney/5/97 were 1/2, 1, and 2, while the corresponding normalized HI titers for A/Sydney/5/97 were 1/64, 1/4, and 1/2 [15]. Due to the high intrinsic variability of HI titers, estimates of antigenic difference should be based on many replicated measurements of these titers, when possible. Here, we show that an estimate of antigenic difference based on multiple HI titers accurately predicts VE against influenza-like illness (ILI) in healthy adults, much more so than do other measures of antigenic difference that we tested. Furthermore, motivated by the need to improve access to HI titers we constructed a large database of HI titers, which we here make public. We hope that this database will serve as a catalyst for the establishment of other publicly accessible repositories of HI titers, which could dramatically improve the pace of research on the evolution of influenza viruses.

2. Materials and methods

2.1. HI data

The HI data used in this study were collected as part of a larger effort to make data on the antigenic evolution of influenza viruses more widely available. The data were assembled by searching relevant research articles and other documents published by the US Centers for Disease Control and Prevention and other collaborating centers of the World Health Organization’s global influenza surveillance network. Presently, the data include 112 HI tables containing a total of 12,500 HI titers obtained using influenza A/H1N1, A/H3N2, A/H5N1, and B virus strains. The data in each HI table are in the form of an m by n matrix H, with m (≥ n) the number of strains found in the HI table, and n the number of antisera raised against a subset of those strains. Each entry $H_{ij}$ of H represents the HI titer of strain D relative to antiserum raised against strain V. The data in all HI tables were obtained by following the same general experimental guidelines [7]. The data accompany this paper as Supplemental Material.

2.2. Using HI data to quantify antigenic difference

The standard measure of the antigenic difference between virus strains D and V is the reciprocal of the normalized HI titer of D relative to antisera raised against V [7]: $r_{ND} = H_{DV} / H_{DD}$. Strain D is said to be antigenically drifted relative to V if rND is greater than or equal to 4 [7]. An alternative measure of the antigenic difference between D and V is the reciprocal of the Archetti–Horsfall measure [9,16,17]: $r_{AHM} = \sqrt{H_{DD}H_{VV} / (H_{DV}H_{VD})}$. We expect rAHM to be less dependent on non-antigenic factors than rNHT (see the next paper of this series); thus, rAHM may provide a more accurate measure of antigenic difference. Often, we have the four HI titers, $H_{DD}$, $H_{DV}$, $H_{VD}$, and $H_{VV}$, from several independent HI assays of the antigenic difference between D and V. In such cases, we use the means of rAHM and rNHT computed using data from each assay. Observe that rNHT is an asymmetric measure of antigenic difference, while rAHM is symmetric.

2.3. Using antigenic difference to predict vaccine effectiveness

Vaccine effectiveness was defined in the usual way: $VE = (1 - RR) \times 100$, where RR is the relative risk of contracting influenza-like illness (e.g., Refs. [11,18]). RR is given by $\text{ad}/(bc)$, where b (respectively d) denotes the total number of vaccinated (respectively unvaccinated) individuals taking part in a given VE study, and a (respectively c) denotes the number of vaccinated (respectively unvaccinated) individuals diagnosed with ILI. We quantify the predictive accuracy of the antigenic difference between virus strains D and V with respect to VE by means of weighted linear regression. The weights used in the regression model control for differences in the variability of VE estimates. The weight associated with a particular VE estimate is defined as the reciprocal of the square of the width, $\text{SE}^2$, of the 95% confidence interval (CI) for that VE estimate. We compute CIs under the assumption that the number of vaccinated (respectively unvaccinated) individuals diagnosed with ILI follows a binomial distribution, with parameters $(a + c)$ and $a/(a + c)$ (respectively $(a + c)$ and $c/(a + c)$). The lower and upper 95% confidence limits for VE are computed using the method of Clopper and Pearson (e.g., see Ref. [19]). When VE is calculated as the average of m different VE estimates, the width of its CI is given by $1/m \sqrt{\sum_{i=1}^{m} \text{SE}_i^2}$, where $\text{SE}_i$ denotes the width of the CI for the i-th VE estimate.

It is important to note that in addition to the antigenic difference between D and V, VE depends on a range of other factors, including the age and immune status of vaccinated individuals and the presence in the population under study of other causative agents of ILI [20]. These factors are often not controlled for (e.g., Ref. [11]). Therefore, the correlation between antigenic difference and VE is not expected to be 100%. Conversely, a high correlation between antigenic difference and VE, although highly suggestive of a causal relationship, should be interpreted with caution. In addition, note that using estimates of VE against virologically confirmed influenza (rather than ILI) could lead to improvement in the calculated predictive accuracy of antigenic difference with respect to VE. However, we only have access to a limited number of such VE estimates (see Supplemental Material).

3. Results

We quantified the predictive accuracies (with respect to VE against ILI) of two HI-based measures of the antigenic difference between virus strains D and V, namely (i) the reciprocal of the normalized HI titer of D relative to antisera raised against V (denoted by rNHT), and (ii) the reciprocal of the geometric mean of both the normalized HI titer of D relative to V and the normalized HI titer of V relative to D (denoted by rAHM). For comparison, we also quantified the predictive accuracies of two measures of antigenic difference based on the amino acid sequence of influenza virus hemagglutinin (HA); one of these measures (denoted by $p_{sequence}$) is the Hamming distance between the HA amino acid sequences of D and V normalized by the length of HA, while the other measure (denoted by $p_{Peptope}$) is the maximum of the Hamming distances.
between corresponding HA epitopes normalized by the total number of amino acids found in each epitope [9,10]. The former measure of antigenic difference is essentially a lower bound on the normalized phylogenetic-tree distance between the HA amino acid sequences of D and V, while the latter measure is thought [9,10] to quantify changes to the “dominant” epitope of HA.

We focused on predicting VE in healthy adult vaccines, since VE in these individuals is more likely to be indicative of the antigenic difference between D and V than VE in the elderly, for whom there is a preponderance of confounding factors (e.g., Ref. [21]). We compiled a total of 22 VE estimates and the corresponding antigenic difference between D and V, for all previous influenza seasons in which (i) virus strains belonging to the H3N2 influenza virus subtype were dominant, and (ii) the relevant data are publicly available (see Table 1).2 Eleven of the 22 estimates correspond to cases when D and V are antigenically identical (i.e., both rAHM and rNHT equal 1, by definition). In these 11 cases, VE is consistently high (43%, S.E. = 5%; see Table 1). Note that the predictive accuracy of a given measure of antigenic difference can only be tested rigorously in cases when D and V are actually antigenically different. Considering only such cases, we find, using a weighted linear regression of VE against antigenic difference (see Section 2), that rAHM has good predictive accuracy with respect to VE ($R^2 = 0.62$, $p = 4.1 \times 10^{-4}$), while rNHT ($R^2 = 0.01$), $p_{\text{epitope}}$ ($R^2 = 0.03$), and $p_{\text{sequence}}$ ($R^2 = 0.03$) all have relatively poor predictive accuracies (see Table 1).

The results obtained using rAHM suggest the existence of a maximum antigenic difference between D and V (about 5) above which vaccines are likely to be ineffective: vaccines were effective (defined as a positive mean value of VE) in 16 out of 17 cases when the computed antigenic difference between D and V was less than 5, but were ineffective in 4 out of 5 cases in which the computed antigenic difference was greater than or equal to 5. This prediction is consistent with the antigenic difference threshold of 4, which is currently in use by epidemiologists [7]. Note that in the elderly, VE is generally high (35%, S.E. = 5%) when D and V are antigenically identical (see Table S1), although it is not as high as in healthy adults. In contrast to the situation in healthy adults, there is a statistically significant positive correlation between VE in the elderly and the antigenic difference between D and V, as measured using both rNHT ($R^2 = 0.53$, $p = 0.02$) and rAHM ($R^2 = 0.90$, $p = 2.5 \times 10^{-5}$) (see Table 2). This unexpected positive correlation could be due to, among other things, the confounding effects of chronic immunization of the elderly on estimates of VE (e.g., Ref. [21]).

In particular, previous empirical and theoretical results suggest that the immunogenicity (and hence the effectiveness) of influenza vaccines could decrease as the antigenic difference between V and the vaccine strain used in the previous influenza season decreases [21,29]. Consistent with those results, we find that there is a statistically significant ($p = 0.02$) interaction between the rAHM of the antigenic difference between D and V and the rAHM of the antigenic difference between V and the H3N2 vaccine strain used in the previous influenza season (see Table 2); VE is predicted to increase with the product of both antigenic differences. In addition, the antigenic difference between D and V is predicted to have a negative effect on VE, but this effect is not statistically significant (see Table 2). Importantly, the regression model that accounts for the abovementioned interaction effect explains a greater amount of the variation in VE than the model that only accounts for the antigenic difference between D and V ($R^2 = 0.98$ versus 0.90), and it also possesses a smaller Akaike information measure (88.27 versus 97.67), suggesting that it is indeed a better model (see Table 2).

4. Discussion

In this paper, we showed that vaccine effectiveness against influenza-like illness in healthy adults is high in seasons when vaccine (V) and dominant (D) circulating influenza virus strains are antigenically identical. In other seasons, VE is predicted accurately by the reciprocal Archetti–Horsfall measure (rAHM) of the antigenic difference between D and V [9,16,17], which is given by the reciprocal of the geometric mean of both the normalized hemagglutination-inhibition (HI) titer of D relative to antisera raised against V and the normalized HI titer of V relative to D. In contrast, measures of antigenic difference based on (i) the normalized HI titer of D relative to V alone, (ii) amino acid differences between the entire hemagglutinin sequences of D and V, and (iii) amino acid differences between “dominant” antibody-binding sites of the hemagglutinin sequences of D and V [9,10] were found to have relatively poor predictive accuracies with respect to VE. The superior predictive accuracy of rAHM could be explained by the fact that it has less dependence on non-antigenic factors than the other considered measures of antigenic difference (see the next paper of this series). Also, the considered sequence-based measures of antigenic difference provide only indirect information about the antigenic properties of influenza virus strains that determine the outcome of their interactions with vaccine-induced antibodies.3 The above results address previous criticisms of HI titers [8–10], and they suggest that rAHM should be used more often when selecting suitable vaccine strains.

Furthermore, we found that in seasons when D and V are antigenically identical, vaccines are generally effective in preventing influenza-like illness in the elderly, but in other seasons there is a significant positive correlation between VE and the rAHM of the antigenic difference between the two strains. This unexpected observation could be explained by a host of factors known to confound estimates of VE in the elderly, including impaired immune function and chronic immunization [20,21]. Indeed, we found that the correlation between VE and the antigenic difference between D and V becomes negative when the antigenic difference between D and the previous season’s vaccine strain was accounted for, although it was not statistically significant. Importantly, we found a statistically significant interaction between the antigenic difference between D and V and the antigenic difference between V and the previous season’s vaccine strain; VE was predicted to increase with the product of both antigenic differences. These results are consistent with the antigenic distance hypothesis [21,29], which postulates that prior immunization with a vaccine strain that has low antigenic difference from V could lead to low immunogenicity and, hence, effectiveness of vaccines containing V. Further studies of the mechanistic basis of these results could be of value. In addition, it would be helpful to investigate the relationship between antigenic difference and (more accurate) estimates of VE against virologically confirmed influenza, once a sufficient amount of those VE estimates becomes available. Overall, our results highlight the importance of consistently selecting the most suitable influenza vaccine strains.

Influenza vaccine-strain selection is currently very challenging, as evinced by the antigenic mismatches between vaccine and dominant strains observed recently [54,57–59]. These mismatches

---

2 We found another VE estimate from the 1968–1969 influenza season [63,64], but the rNHT of the corresponding dominant strain (Hong Kong/8/68) relative to the vaccine strain (Aichi/2/68) was 0.4 [65], which is lower than the expected value for two strains that are antigenically identical. This data point was therefore considered to be an outlier, and it was not included in our calculations.

3 The much lower predictive accuracies of the considered sequence-based measures of antigenic difference compared to their previously reported accuracies could be explained, in part, by the fact that we used a more rigorous criterion for predictive accuracy than was used in previous studies [9,10].
Table 1
Estimates of antigenic difference and of VE in healthy adults.

<table>
<thead>
<tr>
<th>Influenza season</th>
<th>Vaccine strain</th>
<th>Dominant strain</th>
<th>Antigenic difference</th>
<th>VE data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>% S/NHT/AHM p_{epitope} p_{sequence}</td>
<td>a/b/c/d</td>
</tr>
<tr>
<td>1968</td>
<td>Hong Kong/8/68</td>
<td>Hong Kong/8/68</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1969</td>
<td>Aichi/2/68</td>
<td>Aichi/2/68</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1980–1981</td>
<td>Bangkok/1/79</td>
<td>Bangkok/1/79</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1984–1985</td>
<td>Philippines/2/82</td>
<td>Mississippi/1/85</td>
<td>4.76/5.00 [33–35]</td>
<td>.190</td>
</tr>
<tr>
<td>1985–1986</td>
<td>Philippines/2/82</td>
<td>Mississippi/1/85</td>
<td>4.76/5.00 [33–35]</td>
<td>.190</td>
</tr>
<tr>
<td>1987–1988</td>
<td>Leningrad/100/86</td>
<td>Shanghai/11/87</td>
<td>2.00/2.00 [34]</td>
<td>.143</td>
</tr>
<tr>
<td>1993–1994</td>
<td>Beijing/32/92</td>
<td>Beijing/32/92</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1994–1995</td>
<td>Shangdong/9/93</td>
<td>Shangdong/9/93</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1995–1996</td>
<td>Johannesburg/33/94</td>
<td>Nanchong/33/94</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>1999–2000</td>
<td>Sydney/5/97</td>
<td>Sydney/5/97</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2001–2002</td>
<td>Panama/2007/99</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Shown are VE estimates from influenza seasons in which H3N2 virus strains were dominant. When VE estimates were obtained from multiple VE studies, the average of those estimates was used. Antigenic difference was quantified using two measures based on hemagglutination–inhibition titers (rAHM and rNHT; see Section 2); a measure based on the entire amino acid sequence of hemagglutinin (p_sequence); and a measure based on hemagglutinin epitopes (p_epitope) [9,10]. See Table S3 for details on the amino acid positions found in these epitopes. Except when noted otherwise, estimates of p_epitope and p_sequence were taken from [9]. In cases when p_epitope and p_sequence were computed by us the Genbank IDs of the sequences used are given. For 2003–2004 the more accurate [55] VE estimate, which does not take into account individuals vaccinated <2 weeks prior to developing symptomatic influenza, was used. The vaccine and dominant strains for a given season were taken from publications reporting VE estimates for that season, if the strains and/or their antigenic similarity were indicated. Otherwise, the vaccine and dominant strains reported by the World Health Organization were used.

* a – number of vaccinated individuals diagnosed with ILI; b – total number of vaccinated individuals; c – number of unvaccinated individuals diagnosed with ILI; d – total number of unvaccinated individuals.

** The width of the confidence interval associated with each VE estimate was calculated as described in Section 2.

*** Antigenic data for the correct vaccine strain (A/Hong Kong/31-36/68) were not available, so the vaccine strain (A/Hong Kong/1/68) recommended by the World Health Organization was used instead.
can have deleterious consequences for VE, especially in the elderly. Also, in the unfortunate event that avian influenza virus strains (e.g., H5N1 viruses) acquire the ability to spread efficiently among humans, an antigenic mismatch between those strains and the recommended “pre-pandemic” vaccine strain [60] could have very serious consequences. When selecting vaccine strains, researchers rely on epidemiologic, genetic and, most importantly, antigenic data (primarily from HI assays) collected by sampling circulating strains [5]. The development and application of accurate sequence-based measures of antigenic difference to data from ongoing influenza virus sequencing projects (e.g., Ref. [61]) could allow sequence data to be used to complement antigenic data. This could help to increase the diversity of sampled strains that are phenotyped antigenically, as well as the probability of identifying suitable vaccine strains. In addition, the collection of better-quality antigenic data (e.g., corresponding to the means of replicate HI titers) combined with the development of quantitative methods for increasing the signal-to-noise ratio of these data (e.g., Ref. [62]) could improve the selection of suitable vaccine strains and allow greater understanding of influenza virus evolution. In order to help spur the development of such quantitative methods and encourage studies on other aspects of influenza virus evolution, we presented a large collection of HI titers culled from various sources.

Acknowledgments

W.N. acknowledges support from a Burroughs-Wellcome Graduate Fellowship, while J.D. acknowledges support from the Natural Sciences and Engineering Research Council of Canada. We are also pleased to acknowledge support from the Defense Advanced Research Projects Agency (DARPA), grant HR0011-05-1-0057. We are grateful to two anonymous reviewers for their very helpful comments on this manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi: 10.1016/j.vaccine.2009.02.047].

References


Centers for Disease Control and Prevention. Information for the vaccines and related biological products advisory committee; 2005.


