INSIGHT – Methods – GLMMS

**Data**

We analyzed two datasets, one from an outpatient influenza clinical study (INSIGHT 002), and another from a clinical study of influenza in hospitalized patients, (INSIGHT 003). INSIGHT 002 enrolls adult patients 18 and over who present with fever, cough and/or sore throat, and who have a suspected influenza infection. On enrollment, all study participants reported influenza-like illness. Participating physicians recorded demographic information (age, country of enrollment), vaccination status and other relevant medical history for each patient. Infection status was determined to subtype (A/H1N1, A/H3N2, influenza B or no confirmed infection).

INSIGHT 003 enrolled hospitalized patients with a PCR-confirmed influenza diagnosis. The study tracks a number of outcomes, including duration of hospitalization, days in ICU and death.

We excluded five cases who were over the age of 90, and two cases who were under age 18 from the outpatient data set. We chose 90 as an upper age cutoff because we cannot confidently reconstruct historical influenza circulation patterns before 1918. Excluding older cases insures that no one in our data set could have been born prior to the 1918 pandemic. We also excluded ## cases in whom cross-infection by multiple subtypes was detected, or in whom the infecting subtype was unknown.

**Overview of methods**

We used generalized additive mixed models with a logit link to test whether imprinting status is a significant predictor of confirmed infection, or of a severe infection outcome. The baseline model included age, vaccination status, antiviral treatment, and the presence of underlying symptoms as predictors of infection risk (Table 1). The country and season of case observation were both treated as random effects.

We tested this baseline model against models that added effects from childhood imprinting to the same HA group, the same HA subtype or the same NA subtype as the relevant seasonal influenza A challenge. We used AIC to compare the performance of models that included each of three possible imprinting effects against the baseline model.

**Birth year-specific imprinting predictions**

Birth year-specific probabilities of imprinting to a given HA subtype, NA subtype or HA group were estimated exactly as in (Gostic et al., 2016), except the following modification. Previously, we estimated each case’s birth year as *y-a*, where *y* represents the year of case observation and *a* represents age. We have now updated reconstruction methods to better address uncertainty inherent to this approach*.* For example, consider an infant of age 0 observed in May 2000. Using our old method, this case would have been assigned a birth year of *y-a*=2000. However, in reality, this individual could have been born any time between late May 1999 and early May, 2000. This implies 7.5 possible birth months in 1999 [*y-a-1*], and only 4.5 possible birth months in 2000 [*y-a*].

We have generalized the above argument to estimate the probability that each case was born in year *y-a-1*, or in year *y-a* (Fig. S1). Then, we reconstructed imprinting probabilities for each of two possible birth years, and took a weighted average, where weights were determined by the fraction of possible birth months in year *y-a-1* or year *y-a* for each case in the data(Fig S1).

With few exceptions, the enrolled study population would not have been vaccinated against influenza as infants or children, and thus, it was not necessary incorporate impacts from vaccination in infancy or early childhood into imprinting reconstructions. In the last decade, a number of countries have adopted policies that promote influenza vaccination in children [CITE CITE CITE], and the United States now recommends vaccination of all healthy infants at 6 months of age [CITE]. However, childhood vaccination coverage levels remain low globally [PALACHE], and even study participants born in countries that now have high childhood influenza vaccine coverage would have been old enough when these policies were adopted in the late 00’s that they almost certainly would have had an active influenza infection before they were vaccinated.

**Outcomes of interest**

A key challenge in this analysis is that age and subtype-specific protection status are intrinsically correlated, due to the history of influenza circulation in humans. For example, older individuals have higher probabilities of protection against H1N1, whereas younger cohorts have higher probabilities of protection against H3N2. Due to collinearity between age and probabilities of imprinting protection against either seasonal subtype, it is not possible to include both in the same linear model as predictors of the probability of confirmed H1N1 infection, or of confirmed H3N2 infection.

However, for a given birth year, the probability of imprinting protection against H1N1 is the probabilistic complement of protection against H3N2. Thus, it was possible to break the collinearity between age and imprinting status by formulating the model to simultaneously predict probabilities of confirmed infection with H1N1 or H3N2. For example, 60-year-olds should have similar age-specific risk against any seasonal influenza virus, but given the history of influenza circulation, these cohorts would have much different birth year-specific protection against H1N1 (strongest possible protection), than against H3N2 (no protection predicted). This approach requires the assumption that purely age-specific risk factors should have similar age-specific impacts on risk from all influenza subtypes. For example, age-specific rates of vaccination, age-specific behavioral and social mixing patterns, immunosenescence, and age-specific risk of a severe infection outcome should all should all impact seasonal influenza risk, but should have roughly the same impacts regardless of whether a seasonal epidemic is caused by A/H1N1, A/H3N2 or influenza B. After constraining age-specific risk to take the same shape against H1N1 and H3N2, it becomes possible to test for additional, subtype-specific imprinting effects, which may arise from differences in childhood immune history. To do so, we analyzed the interaction between challenge subtype {H1N1, H3N2}, and imprinting protection probability against the relevant challenge subtype (Equation 5).

When confirmed infection was the outcome of interest, we only analyzed cases from countries and seasons in which one subtype was responsible for 90% or more of seasonal circulation. Then, it was possible to input each case’s birth year-specific probability of imprinting protection against the dominant circulating subtype. (On the other hand, in seasons with mixed H1N1 and H3N2 circulation, it was impossible to determine whether the birth year-specific probability of imprinting protection against H1N1 or against H3N2 was relevant, and so we excluded these cases). When the outcome of interest was disease severity (e.g. duration of infection, or ICU admission), given an influenza infection confirmed to subtype, imprinting protection against the subtype responsible for the observed infection was clearly the relevant choice, and so it was not necessary to exclude cases from seasons with mixed IAV circulation.

**Models tested**

We used a generalized additive mixed model with a logit link to describe probabilities of infection in the outpatient data set. Here, Yics was a binary variable that took value 1 if patient i had a confirmed H1N1 or H3N2 infection, and value 0 otherwise. Subscripts c and s indicate the country and season in which the patient was observed. The baseline model included linear fixed effects from patient age, the presence of underlying conditions, and antiviral treatment. The baseline model (Equations 1-4) also included the interaction between recent vaccination and the relevant seasonal challenge subtype (H1N1 or H3N2). We included this interaction because vaccine efficacy is known to differ between H1N1 and H3N2. Country and season were also included in the baseline model as random effects. Table 1 summarizes factors tested.

(1)

(2)

(3)

(4)

We classified cases observed from October-March as falling into the corresponding year’s Northern Hemisphere (NH) influenza season, and cases observed from April-September into the corresponding Southern Hemisphere influenza season. One limitation of this analysis is that, due to a change in study protocol, some subjects were asked to report influenza vaccination in the last 6 months while others were asked to report influenza vaccination in the past year, which may introduce some noise into the data on history of recent influenza vaccination.

In addition to the baseline model, we tested effects from childhood imprinting to the same HA group, to the same HA subtype and to the same NA subtype as the seasonal challenge of interest. Data visualization showed a nonlinear relationship between imprinting protection and probability of seasonal influenza infection (Fig. S##), so for each imprinting hypothesis, we tested two models. One included a linear fixed effect of imprinting (Equation 5), and the other used a penalized cubic regression spline to incorporate nonlinear effects (Equation 6).

(5)

(6)

**Model fitting**

Model fitting was performed in R (Version ##) using the gam function from the mgcv package. All relevant code is provided in the supplementary appendix.

**Results**

* In the best 2 models, imprinting effects were shrunk to have no significant impact.
* In the 3rd best model, (HA group, nonlinear), imprinting protection had a significant impact on H1N1 risk, but only a near-significant impact on H3N2 risk.

|  |  |  |  |
| --- | --- | --- | --- |
| **Model** | **Imprinting form** | **AIC** | **ΔAIC** |
| baseline + HA subtype imprinting | nonlinear | 4445.693 | 0 |
| baseline + NA subtype imprinting | nonlinear | 4447.12 | 1.43 |
| baseline + HA group imprinting | nonlinear | 4449.42 | 3.72 |
| baseline + HA group imprinting | linear | 4453.70 | 8.00 |
| baseline + HA subtype imprinting | linear | 4453.97 | 8.27 |
| baseline |  | 4454.19 | 8.50 |
| baseline + NA subtype imprinting | linear | 4455.94 | 10.25 |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Description | Possible values | Included in baseline model? |
| age | Continuous: patient age | 18-93 | Yes |
| vaccination | Binary: vaccination in past year noted? | {0, 1} | Yes |
| antiviral treatment | Binary: antiviral treatment noted? | {0, 1) | Yes |
| underlying conditions | Binary: any underlying condition present? | {0, 1} | Yes |
| season | Categorical: treated as a random effect | Earliest: NH ’09-‘10\*  Latest: NH ’16-‘17 | Yes |
| country | Categorical: treated as a random effect | {Denmark, Spain, Germany, Estonia, USA, Belgium, Portugal, Poland, Austria, UK, Australia, Thailand, Argentina, Chile, Greece, Peru, Japan} | Yes |
| imprinting protection, HA group level | Continuous: probability of imprinting to the same HA group as the current seasonal challenge | [0, 1] | No |
| imprinting protection, HA subtype level | Continuous: probability of imprinting to the same HA subtype as the current seasonal challenge | [0, 1] | No |
| imprinting protection, HA group level | Continuous: probability of imprinting to the same NA subtype as the current seasonal challenge | [0, 1] | No |

Table 1: Independent variables

\*NH abbreviates northern hemisphere influenza season, defined as spanning October-March in this study. SH abbreviates the southern hemisphere influenza season, defined as spanning April-September in this study

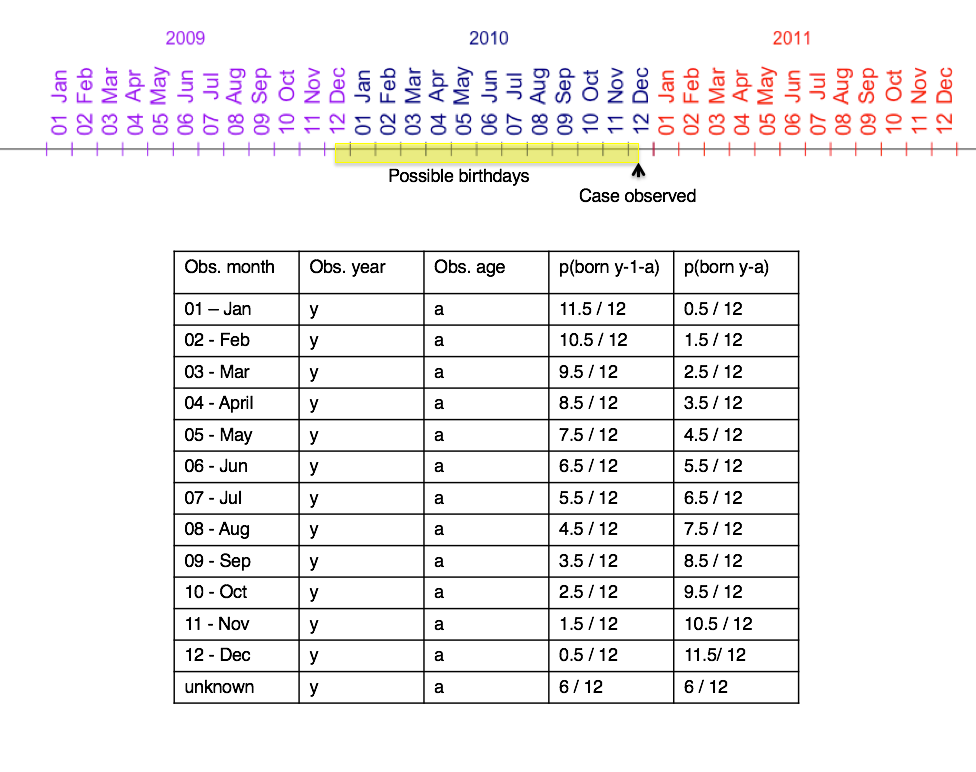


Figure S1. Calculation of birth year from month of case observation and at at time of case observation.