INTRODUCTION

AGE-SPECIFIC DIFFERENCES IN H1N1 VS. H3N2. LARGE EPIDEMICS AND PANDEMICS OFTEN SHOW ODD AGE DISTRIBUTIONS OF INFECTION. E.G…. The seasonal influenza viruses endemic to humans also show differences in age distribution… LAST YEAR (H3N2), incidental (HENSLEY), long-term (H1N1 vs. H3N2).

The ability to predict age distributions of influenza infection would provide myriad advantages, for epidemic preparedness and personalized medicine. But the extent to which these age distributions are predictable remains unclear, especially for seasonal influenza viruses.

Pre-existing immunity is a strong driver of observed influenza age distributions, and the strength of pre-existing immunity against a particular influenza challenge is predictably linked to birth year. Individuals develop particularly strong, lifelong immune memory against the influenza viruses they encountered in childhood, a phenomenon alternatively referred to as original antigenic sin, antigenic seniority or immunological imprinting. Recently, we showed that birth year-specific patterns of immunological imprinting are a surprisingly good predictor of observed case age distributions for two novel, avian influenza subtypes H5N1 and H7N9.

Similar birth year-specific imprinting effects are also hypothesized to impact age distributions of seasonal influenza infection. For example, H1N1 viruses cause relatively low incidence and mild disease in older adults, and numerous studies have hypothesized that older adults may enjoy preferential protection against contemporary H1N1 viruses as a result of childhood exposures to the same influenza subtype [CITE CITE CITE]. However, this evidence is largely supported by qualitative inference, and using data collected in a single influenza season or from a limited geographical range. Here, we use a formal statistical framework to analyze data that span multiple influenza seasons, and multiple countries to test systematically whether childhood imprinting patterns predictably shape seasonal influenza age distributions.

ONE KEY CHALLENGE IS THAT age-specific effects shape observed distributions of seasonal influenza cases, alongside birth year-specific effects. Age and birth year are intrinsically correlated (unless one could obtain a data set spanning multiple decades, or ideally an entire human lifetime), and thus the respective impacts of age-specific and birth year-specific effects are easily confounded.

To address this challenge, we fit a single curve to describe age-intrinsic risk from any seasonal influenza A virus (H1N1 or H3N2), and then use model comparison (AIC) to assess the extent to which birth year-specific imprinting can improve the model’s ability to explain residual, subtype-specific differences in case age distribution.

Age-specific effects include age-specific social mixing patterns (*38*), age-specific vaccine coverage, age-specific risk of severe disease, and immunosenescence (CITE). Behavioral factors may also impact age-specific probabilities of case observation, as large influenza data sets are typically collected using passive surveillance (i.e. cases are only detected if infected patients seek outpatient treatment). Infected individuals may seek medical treatment at different rates due to age-specific differences in employment status or mobility. Furthermore, cases attended at nursing homes or by pediatricians may be reported at different rates, or excluded entirely from certain data sets. All the above age-specific factors should have subtype-independent impacts on influenza risk. In other words, age-intrinsic risk should take roughly the same shape in an influenza epidemic caused by seasonal subtype H1N1 or H3N2.

On the other hand, birth year-specific risk is distinctly subtype dependent. Birth year is an established proxy for influenza immune history (CITE), which roughly describes how an individual’s childhood influenza exposures, and boosting from subsequent exposures determines the strength of pre-existing immunity against particular IAVs.

We tested three imprinting hypotheses (Table 1). First, imprinting protection against seasonal influenza may only provide protection at the HA subtype level, in which case individuals would gain preferential, lifelong protection against viruses of the same HA subtype as the first influenza A virus encountered in childhood. HA subtype-level imprinting protection is consistent with the fact that antibody responses against seasonal influenza typically provide narrow, within-subtype protection. Second, imprinting protection may act at the HA group level. HA group-level protection is consistent with patterns of protection against avian influenzas viruses (CITE). Although immune responses providing broad cross-protection at the HA group level are rarely expressed against familiar, seasonal influenza variants, these broadly protective responses may provide an effective, second line of defense against drifted seasonal variants, in which case memory of more immunodominant, variable epitopes may be absent or ineffective. Finally, imprinting protection may act at the NA subtype level.

**Methods**

Both age-specific effects and birth year-specific differences in immune history shape an individual’s risk of infection with seasonal influenza A viruses. Age-specific risk factors include age-specific social mixing patterns (MOSSONG),

But because age and birth year are intrinsically correlated, their respective impacts are confounded by most standard statistical approaches.

To tease apart age-specific and birth year-specific effects

In the category of age-specific risk factors, social mixing patterns influence transmission networks[CITE]. Immune function and overall health also change with age [CITE], as do rates of vaccination, antiviral use, and the presence of underlying health complications (Fig. XX). Additionally, age-specific behavioral factors modulate probabilities of case observation, as certain age groups may be more likely to seek a doctor’s care upon development of flu-like symptoms.

Unfortunately, few of these age-specific factors have well-quantified distributions across different age groups, and the specific impacts from and interactions between individuals, mechanistic risk factors are even less well understood. However, few if any of these age-specific risk factors should be markedly different in influenza epidemics caused by H1N1 or H3N2—the majority reflect age-specific differences in behavior, and the few physiological risk factors noted should modulate risk from *any* influenza virus challenge.

Thus, we were able to break the correlation between age and birth year-specific effects by fitting a single age-specific risk curve to observed case age distributions from influenza epidemics caused by any seasonal subtype. We reason that the majority of age-specific factors should have similar impacts regardless of the circulating seasonal influenza subtype (H1N1 or H3N2).

This approach allows us to use model comparison to test whether any of a suite of imprinting hypotheses are able to improve the age-specific model’s fit to observed data by explaining residual, subtype-specific differences in birth-year specific risk. These subtype-specific differences arise predictably, as childhood imprinting predicts different levels of protection against H1N1 and H3N2 (Fig. ##, table ##). Following similar logic, Gangon et al. (CITE) recently suggested that birth year-specific effects could be identified by comparing the proportion of cases in a given season, and within a given age group, that were caused by H1N1 or H3N2. Our likelihood-based framework follows similar logic, but instead of comparing case ratios from individual influenza seasons, and in individual age groups, allows us to simultaneously examine birth year-specific differences in H1N1 and H3N2 incidence across all ages, and to simultaneously integrate data from across numerous influenza seasons. Our approach is designed to facilitate comparison of three distinct childhood imprinting hypotheses, where we assume that imprinting to a particular seasonal subtype may provide protection later in life against viruses with (1) HA in the same group, (2) HA of the same subtype or (3) NA of the same subtype as the first influenza A virus encountered in childhood (Table ##).

**Likelihood**

Define **pc,s** as a vector whose entries, pa,c,srepresent the probability that a randomly drawn case was observed in an individual of age a during season s in country c. Then, use a linear model to define the expected case age distribution. Information on individual case medical histories was available in the INSIGHT dataset, and so we were able to model impacts from vaccination prior to the influenza season of interest (***V****)*, of underlying symptoms (***U***) and of antiviral treatment (***T***) explicitly when fitting to these data. However, when fitting to the Arizona dataset, these patient details were not reported, and so the linear model used to represent **pc,s** only included age-specific risk *(****A****)*, and birth year-specific risk (***I***)from imprinting. We assumed underlying symptoms (***U***) and antiviral treatment *(****T***) would have similar impacts on H1N1 and H3N2’s age distributions and therefore fit a single curve to describe these impacts on both H1N1 and H3N2, however, we fit distinct impacts of vaccination (***V***), and of imprinting protection (***I***) against each subtype.

When fitting to data from the INSIGHT study, the most complicated model for **pc,s** was defined as follows. All other tested models were nested within the full model, and can be obtained by removing factors ***T, U, V*** or ***I***.

***p*c,*s*** *=****A****\*****T*c,*s****\*****U*c,*s*** *\** **𝟙**H1(***V*c,*s,H1*** *\*****I*c,*s,H1***)*\** **𝟙**H3(***I*c,*s,H3****\*****V*c,*s,H3***)

Note that **𝟙**Hx is an indicator function that takes value 1 if ***p*c,*s*** is being fit to data on the age distribution of HxNy cases, thus allowing some factors to be subtype-specific (only fit to H1N1 or H3N2’s observed age distributions), whereas others are subtype-independent (fit to data on both subtypes). The model for age-specific risk was defined as:

***A*** *=* [**𝟙**18-24\*r18-24\*b +**𝟙**25-31\*r25-31\*b +**𝟙**32-38\*b+**𝟙**39-45\*r39-45 +**𝟙**46-52\*r46-52 +

**𝟙**53-59\*r53-90+**𝟙**60-66\*r60-66 +**𝟙**67-73\*r67-73 +**𝟙**74-80\*r74-80 +**𝟙**81-90\*r81-90]

Here, 𝟙i-j is a binary vector, whose entries take value 1 for ages within the range i-j, and 0 otherwise. Constant *b* is a fixed, baseline value for age-specific risk, which we arbitrarily set to 1, after verifying that other arbitrarily chosen values produced identical maximum likelihood estimates of all other free parameters. We arbitrarily designated the 32-38 age group as the baseline risk group, as this choice was convenient for optimization (SEE CODE). In all age groups other than the baseline group, the relevant indicator function is multiplied by *b,* and by an age-specific free parameter *ri-j,* which modulates relative risk in the age group of interest. Thus, we fit nine free parameters to define a stepwise function that characterizes age-specific risk, independent on the circulating subtype.

Next, ***T*c,*s*** was defined as:

***T*c,*s = f****c,s,T \*rT+(1-****f****c,s,T)*

Here, ***f****c,s,T* is a season-specific vector whose entries *fa,c,s,T* represent the fraction of individuals of age *a* who were treated with antivirals in season *s.* For countries, seasons and ages where one or more cases was observed, *fa,c,s,T* represented the observed fraction of casesin which antiviral treatment was reported. For countries, seasons and ages where no cases were reported, *fa,c,s,T*  entries took a value representing a predicted probability of antiviral treatment from within the study population of individuals in country *c* and of age *a*. Predicted, age-specific probabilities were obtained by fitting a local polynomial regression of treatment status on age for all subjects recorded from a given country. We chose a loess smoother of degree 2 and span 0.9, which visually produced a smooth curve that captured observed, country-specific trends without being overly flexible (Supplement). Inputting these predicted probabilities of antiviral treatment when no corresponding cases were observed was necessary because, for example, inputting a 0 probability of antiviral treatment at all entries corresponding to 0 observed cases would have created a spurious signal that the absence of antiviral treatment was associated with low infection risk.

Identical methods were used to define risk factors from the presence of underlying symptoms, and from vaccination, ***U*c,*s***,***V*c,*s,H1***,and ***V*c,*s,H3*.**

***U*c,*s = f****c,s,U \*rU+(1-****f****c,s,U)*

***V*c,*s,Hx = f****c,s,V,Hx \*rV,Hx+(1-****f****c,s,V,Hx)*

Finally, we reconstructed country and birth year-specific probabilities of childhood imprinting as described previously (CITE). We used these reconstructions to build vectors ***f****c,s,I,Hx*, whose entries describe the probability that an individual of age a, in country c, and season s is protected by their childhood imprinting against seasonal challenge HxNy. We built several variants of these vectors to reflect the three tested hypotheses that imprinting may provide protection against seasonal challenges with HA of the same subtype, with HA in the same group, or with NA of the same subtype as the first IAV encountered in childhood.