**Multinomial methods**

~~A key challenge in this analysis is to separate age-specific risk factors from birth-year specific risk factors. INSIGHT data was collected over eight years, form 2009 to 2017, so each birth cohort grows eight years older from the earliest season in our data to the most recent. This provides some distinction between age and birth year, but these two factors remain intrinsically correlated.~~

Both age-specific risk factors and birth-year specific risk factors (cohort effects) are likely to influence observed age distributions of H1N1 and H3N2 cases. Here, we assume that age-specific risk factors are fixed, regardless of the viral challenge. For example, age-specific social mixing patterns influence transmission networks[CITE]. Immune function[CITE] and overall health[CITE] are known to change predictably with age. Rates of vaccination, antiviral use, and the presence of underlying health complications also vary by age (Fig. XX). Finally, certain age groups may be more likely to visit a doctor’s office upon developing flu-like symptoms, and these behavioral patterns will influence age-specific probabilities of case observation. Taken together, this suite of age-specific factors should roughly define the expected age distribution of observed infections for *any* influenza subtype *in an immunologically naïve population*. Put another way, we assume that if pre-existing population immunity did not exist, that an epidemic caused by any influenza subtype, or even by a non-influenza virus with similar modes of transmission and symptoms, would show similar age distributions of infection as a result of the combined effects of the age-specific factors described above.

Of course, in reality, pre-existing immunity does play an important role in shaping actual and observed age distributions of cases. We propose that consistently observed differences in observed age distribution of H1N1 and H3N2 cases can largely be attributed to cohort-specific differences in pre-existing immunity.

Following this logic, our analysis strategy can be summarized as follows:

1. Characterize the hypothetical, expected age distribution in the absence of pre-existing immunity and cohort effects. We will refer to this distribution as the “**null**” for the duration of the manuscript.
2. Compare the null age distribution to observed age distributions of confirmed H1N1 infection, and confirmed H3N2 infection.
3. Develop a suite of *a priori* predictions describing how cohort effects may cause observed distributions of H1N1 and H3N2 to differ from the null.
4. Use model comparison (AIC) to determine which, if any of the cohort effect hypotheses provides the best fit to observed data.

**Null age distribution**

Ultimately, we developed an empirical method to characterize the null age distribution of cases. Essentially, we assumed that the observed age distribution of tested cases should approximate the null distribution. Observed age distributions of tested cases average across age distributions from overlapping epidemics caused by A/H1N1, A/H3N2, influenza B viruses and other respiratory pathogens. While the age distribution for any single pathogen is likely to depart from the null distribution due to pathogen-specific cohort effects, we assume that averaging across observed distributions for many pathogens is a reasonable way to approximate the pathogen-independent, null.

In an attempt to verify that this approach is robust, we developed two strategies to randomly subsample the full data and build the null distribution. We repeated null construction ## times, using each strategy and verified that the core results of our analysis were unchanged.

*Characterization of null age distribution:*

We assumed null age distributions would differ at the country level, because healthcare-seeking behavior and social mixing patterns vary geographically. Our analysis considered individuals age 18-90, for a total of 73 single-year age bins.

We wished to analyze data from countries with at least 400 total cases enrolled (slightly over five times the total number of age bins in the distribution), and with at least 100 each of confirmed H1N1 cases and confirmed H3N2 cases. Argentina (n = 3520), Belgium (n = 1311) and Thailand (n = 3274) were the only three countries that met these criteria.

*Subsampling strategy 1:*

We wished to include an equal number of confirmed H1N1 and confirmed H3N2 cases included in the null distribution, so that neither seasonal subtype had disproportionate influence. In our first sampling strategy, we defined m as the minimum number of subjects from the country of interest in any of the following four diagnosis categories: (1) confirmed H1N1, (2) confirmed H3N2, (3) confirmed influenza B or (4) PCR negative. We then randomly sampled m patient ages from each of the four categories, and set the null distribution for the country of interest equal to the smoothed density of the resulting sample. Smoothed density was computed using the density() function in R. The main analysis used a Gaussian kernel with default bandwidth. Sensitivity analyses verified that the core results of our analysis are robust to bandwidth choice. Full code is provided in the supplement.

*Subsampling strategy 2:*

The alternate subsampling strategy differed only in that we included all ages from categories 3 and 4 (confirmed influenza B, or PCR negative), instead of subsampling from these categories.

TABLES OF NUMBERS OF CASES ANALYZED?

ACKNOWLEDGE LIMITATIONS, BUT ROBUST ACROSS TIME, SPACE AND DATA COLLECTION.

Initially, we attempted to simulate a null age distribution using a dynamical, age-structured transmission model. However, we quickly realized that this mechanistic, simulation approach required an unjustifiable number of influential assumptions about the nature of cross-immunity between influenza subtypes, the temporal duration of immune memory, annual vaccine uptake and efficacy, and host of other factors. An additional pitfall of this approach is that even if we were confidently able to simulate the null age distribution of cases, the output would still need to be filtered through a poorly understood observation process before it could be compared to the observed data. Certain age groups may be more likely to seek medical care and appear in the data than others (e.g. our data excludes cases under age 18, and the number enrolled young adults, approximately ages 18-25, is conspicuously low). Thus, observed distributions of cases cannot be assumed to represent the overall age distribution of infections in the population. A key advantage of our empirical approach is that age-specific biases in case observation are automatically wrapped into the null distribution. Since very few 18-year-olds appear to have been tested for influenza in the first place, our empirical null distribution assumes that 18-year-olds will account for a proportionately low number of confirmed influenza cases.

**Imprinting hypotheses**

TALK IN INTRODUCTION ABOUT RATIONALE.

We assumed childhood imprinting might act at one of

**Other factors tested**

**All models tested**

**Likelihood**

**Make sure you talk about data and exclusion criteria somewhere!!**