# Methods

## Estimation of age from birth year from age in AZDHS data

The AZDHS data contained three variables, influenza season, birth year and confirmed subtype. For most cases, birth year was extracted directly from the reported date of birth in patient medical records. To fit age-specific risk curves to the AZDHS data, we estimated patient age at the time case observation using the formula [year of observation]-[birth year]. To insure that the minimum estimated age was 0, where the second year in the corresponding influenza season was considered the calendar year of observation (e.g. 2013 for the 2012-2013 season).

## Estimation of birth year from age in the INSIGHT data

The INSIGHT data contained patient age, and the exact date of case enrollment, but not birth year. We estimated birth year using a method that took advantage of the precise temporal resolution of INSIGHT dates of enrollment. The simplest approximation of birth year would have been (observation year)-(age), but this approximation is slightly biased, as cases observed earlier in the year (e.g. in January) are less likely to have passed a birthday in the current calendar year. Using logic laid out in Fig. S9, we used the following three formulas to estimate three possible birth years for cases observed in the NH influenza season: (current year)-(age)-1, (current year)-(age)-, and (current year)-(age)-+1. We then took a weighted average of the three relevant, birth year-specific imprinting protection probabilities, using weights 0.0625, 0.875 and 0.0625, respectively. Meanwhile, for cases observed in a SH season, only two birth years were possible. Here, we used the formulas: (current year)-(age)-1, and (current year)-(age) to calculate both possible birth years, with each receiving a probabilistic weight of 0.5.

## Splines

In Figures 2-3, smoothing splines were fit to aid visual interpretation of noisy data. We fit splines using the command *smooth.spline(x = AGE, y = FRACTIONS, spar = 0.8)* in R version 3.5.0. Variables *AGE* and *FRACTIONS* were vectors whose entries represented single years of age, and the fraction of cases observed in the corresponding age group. The smoothing parameter 0.8 was chosen to provide a visually smooth fit. Alternative smoothing parameter choices (0.6 & 1.0) are shown in Figs. S##-##. Although the exact shape of fitted splines changed was sensitive to our choice of smoothing parameter, qualitative differences between H1N1 and H3N2’s age distributions were robust.

## Model formulation

For each unique country and season in which cases were observed, define *p* as a vector whose entriesrepresent the expected probability that a randomly drawn H1N1 or a randomly drawn H3N2 case was observed in an individual of age *a*. Each model defined, *p* as a linear combination of age-specific risk, birth year-specific risk (i.e. imprinting effects), and other medical history variables, and *p* took slightly different shapes for expected H1N1 and H3N2 case age distributions. All tested models were nested within the equation:

Note ***1H1N1*** is an indicator function that takes value 1 if *p* describes the expected age distribution of H1N1 cases, and 0 otherwise, thus including subtype-specific risk factors only in relevant subtype-specific predictions. Similarly, ***1H3N2*** takes value 1 if *p* describes the expected age distribution of H3N2 cases. Subtype-independent risk factors always modulate *p,* regardless of the focal subtype.

### Denominator data (D)

When fitting to INSIGHT data, *D* was a vector whose entries were proportional to the age distribution of all tested cases within a given country and year. As noted above, corresponding denominator data were not available in the AZDHS dataset, and so factor D was not included in the. Model.

### Age-specific risk (A)

Age-specific risk was defined as a step function, in which relative risk was fixed to value 1 in an arbitrarily chosen age bin, and then z-1 free parameters were fit to describe relative risk in all other age bins. Below, ***1i*** are indicator functions specifying whether each vector entry is a member of age bin *i.* To obtain the predicted fraction of cases observed in each single year of age, we normalized risk distribution given by equation 2 so that predicted risk across all age groups summed to 1.

### Antiviral treatment (T)

Within each country and season, *fT* defined a vector whose entries describe the fraction of tested cases of a given age that had received antiviral treatment. Free parameter *rT* defines the relative risk of any confirmed influenza infection, given antiviral treatment. Then, risk factor *T* is defined as:

### Underlying conditions (U)

The underlying conditions risk factor takes the same form as factor *T*:

### Vaccination (V) and Imprinting (I)

Factors describing risk from vaccination and imprinting took forms similar to risk factors *T* and *U*, but with subtype-specific impacts. An indicator function defined whether a given prediction vector described risk of confirmed H1N1 or H3N2 infection. Let *fV* and *fI* be vectors describing the fraction of cases of each age that were vaccinated against influenza, or that were protected against strain *HxNy* by their childhood imprinting.

## Model fitting and model comparison

We simultaneously estimated all free parameter values using the optim() function in R. We calculated likelihood profiles and 95% profile confidence intervals for each free parameter.

## Code and data availability

Code to perform all analyses and construct all plots is available \#\#HERE\#\#. AZDHS data is available as a supplementary data file. Data from the INSIGHT study are available by application, pending approval from the study's scientific review committee (<http://insight.ccbr.umn.edu/index.php>). Because we are not free to share the INSIGHT data, the posted code contains an INSIGHT data file with scrambled column entries.