# Methods

## Overview of modeling strategy

We reconstructed patterns of childhood imprinting in the human population using models that considered the first childhood IAV exposure only (one-hit model), the first and second exposures (two-hit model), or the first second and third exposures (three-hit model). These reconstructions allowed us to estimate the fraction of each birth year with a specific sequence of matched or mismatched childhood exposures. Matched exposures are defined as childhood exposures to a seasonal IAV in the same HA group as a future avian challenge, and are assumed to contribute to imprinting protection. Mismatched exposures (to the other HA group) do not provide protection. Throughout the manuscript, subscript m signifies a matched exposure while subscript o signifies a mismatch. In children too young to have had three exposures at the time of data observation, an o is also used to signify the absence of any exposure.

The one-hit model yielded only two possible imprinting outcomes against a group 1 or group 2 avian challenge: {m, o}. The two-hit model yielded four possible outcomes, {mm, mo, om, oo}, and the three-hit model yielded eight possible outcomes, {mmm, mmo, mom, omm, moo, omo, oom, ooo}. We assumed the realized strength of imprinting protection against H5N1 or H7N9 would depend on both the order and number of matched childhood exposures. We fit multinomial models to a combined dataset of H5N1 and H7N9 cases to estimate the relative risk of severe, detectable avian influenzas infection for each imprinting outcome in each of the three tested models.

### Reconstruction of childhood imprinting patterns

To reconstruct population-level imprinting patterns, we expanded on methods developed in Gostic et al., 2016. Briefly, we first estimated , or the probability that an individual would have had their first, second and third IAV exposures at ages *a1, a2* and *a3.* We assumed only one exposure occurred per year, which imposed the constraint *a1 < a2 < a3.* Two immunologically significant exposures in the same year are unlikely, as cellular immunity remains up-regulated for weeks to months after an influenza virus infection (CITE), which is thought to dramatically reduce the probability of a second influenza infection in the same season (CITE). We also assumed infants under 12 months (age 0) could have immunologically significant exposures. This imposed the constraint, *a1 ≥* 0, where *a1 =* 0 implies the first exposure occurred in the year of birth, *b.* Finally, we assumed the maximum possible age of third exposure was 25, and for cohorts over age 25 in the year of data observation, we normalized so that (similar to methods in Gostic et al., 2016).

We modeled childhood influenza exposures as independent events with exponentially distributed waiting times. In other words, we assumed that previous exposures do not impact the probability of additional immunologically significant exposures in the future. Crucially, we define immunologically significant exposures as any viral challenge that stimulates an anamnestic response, but that may or may not lead to active, symptomatic infection. Although immunity from previous exposures would clearly impact the probability of future influenza *infections* (which we define as symptomatic disease accompanied by within-host viral replication), we assume probabilities of *exposure* (which may or may not culminate in active, symptomatic infection) are stochastic and memoryless. In other words, we assume that immunity from an influenza infection last year won’t prevent classmates from sneezing on your child and stimulating your child’s immune system. Although we acknowledge that immunity from last year may prevent your child from getting sick, our models consider only probabilities of exposure, not disease outcomes.

Granted, active infections would most likely stimulate a more robust immune response and stronger boosting than asymptomatic exposures, but the mapping between past exposures and current immunity remains poorly understood for influenza. Thus, quantifying dependencies between past and future infections into our modeling framework would require strong *a priori* assumptions, which themselves could strongly shape imprinting reconstructions. Instead, we have opted for a more parsimonious approach, which assumes all exposures (i.e. boosting events) are functionally equal.

If influenza attack rates (annual probabilities of exposure, denoted *)* did not vary from year-to-year, the following equation could be used to estimate the probability that an individual’s first and second exposures occur at ages (*a1, a2*).

(1)

Here, the first bracketed factor describes the probability of waiting a1 years between birth and the year of first exposure, and the second bracketed factor describes the probability of waiting an additional *a2-(a1+1)* years between the first and second exposure.

In reality, influenza attack rates vary from year-to-year, so is not constant. Using methods described previously (Gostic et al., 2016), we estimated annual attack rates for all years from 1918 until the present using data on the historical intensity of seasonal influenza circulation. We then modified equation 1 to incorporate variable annual attack rates:

(2)

It is conceptually straightforward to generalize this approach to account for three exposures (or more):

] (3)

Next, for each possible year of birth (b), we cross-referenced virologic data on which strains circulated in the years corresponding to all possible (*a1, a2, a3*)combinations, to obtain probabilities that the first, second and third exposure was to H1, H2, or H3, i.e. . Then, we aggregated across all possible (a1, a2, a3) combinations to obtain the overall probability of each subtype-specific imprinting outcome, for a given birth year:

(4)

Finally, we re-aggregated from subtype-specific imprinting patterns, to group-specific imprinting patterns. HA group 1 contains seasonal H1, and H2, as well as avian H5, while group 2 contains seasonal H3 and avian H7. Aggregation was performed by summing relevant subtype-specific probabilities. For example, to obtain the birth year-specific probability of two consecutive matches to group 1, followed by a mismatch to group 1, the following summation was performed, with the subtype-specific imprinting probabilities on the right-hand side given by the right-hand side of equation I.

(5)

Using the strategy described in equation 5, we calculated probabilities of all possible group 1 and group 2 imprinting outcomes in the three-hit model, {mmm, mmo, mom, omm, moo, omo, oom, ooo}g1, and {mmm, mmo, mom, omm, moo, omo, oom, ooo}g2,and for each individual year of birth from 1918 to 2017. Each probability was entered into vector w*xxx*, whose entries, wb,*xxx*, described the expected fraction of individuals of imprinting status *xxx* within a given birth year. Note that *,* as 100% of the population falls into one of the 8 possible imprinting classifications listed above upon challenge with a given HA group.

To obtain imprinting probabilities in simpler models, we summed across all possible outcomes at later, unmodeled exposures. For example, in the two-hit model, . The outcome at the third exposure is irrelevant in the two-hit model, so we can simply sum across all three-hit outcomes in which the first two exposures align with the desired two-hit outcome.