Supplementary Material for: A novel approach for robust evaluation of broadly reactive influenza vaccine candidates

Table of contents

# 1. Reproducibility instructions

TODO

# 2. Extended methods

The methods included in the main text are a summary of the most important methods we used to conduct our analysis. Here, we include a more detailed explanation of many aspects of the methods used.

## 2.1 Antigenic distance calculation

We calculated the antigenic distance for all of the influenza strains that were used in the cohort data. We calculated the distances pairwise between each set of two strains, and we calculated all distances separately between H1N1 and H3N2 influenza strains. That is, we did not calculate the distance between any H1N1 and H3N2 strains. We used three different distance metrics to measure antigenic distance. All three have been previously described.

The first metric we used was the temporal antigenic distance. Many existing papers in the literature either order strains in year order and then count the distances between them as equal, regardless of the number of years elapsed (1,2); or assign distance based on the year of isolation (3,4). The temporal distance is based on the year of isolation for each strain, and measures how much chronological time has passed since the strains were first observed and sequenced. Letting and be two influenza strains from the same subtype, and be the year of isolation for an arbitrary strain, the temporal distance is given by

That is, the temporal distance is equal to the absolute difference between the year of isolation for the strains. For our study, the absolute value is not necessary because we do not have any assay strains that were isolated after the vaccine strain to which they were compared. However, in such a study, the absolute temporal distance should be compared with the directional temporal distance as future years might have a greater impact on evolution than past years.

The second metric we used was a sequence-based metric called the dominant p-epitope metric (5). First, the -epitope distance is calculated for each of the five antigenic sites on the hemaaglutinin head domain. The five antigenic sites are composed of residues with known positions using a standard numbering scheme for HA proteins which varies by subtype (6). Notably, the numbering for the residues refers to the sequence after the sequence for the signal peptide has been removed. For H1 HA proteins, the signal peptide is typically 18 residues in length. For a given epitope site, the -epitope distance is calculated as

where is the standard Hamming distance (7) which counts the number of differences between two strings. The dominant -epitope distance is then the maximum of the -epitope distance for each set:

Here the index set contains the letters which are typically used as names for each of the antigenic sites. A table of the site residues can be found in the original reference (5).

We obtained the sequence for every strain used in the cohort study data from GenBank. The accession numbers for each of the sequences we used are shown in [Table 1](#tbl-accession). In order to calculate the -epitope distance, we first preprocessed the sequences by removing the signal peptide sequence from the beginning of each sequence. After preprocessing, we computed the -epitope distance for each of the epitopes, and then we took the maximum across all epitopes for each strain pair to obtain the dominant -epitope distance.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: GenBank accession number for the protein sequences for each strain included in our study data.   | Strain | Protein Sequence Accession Number | | --- | --- | | SC/18 | AAD17229.1 | | PR/34 | AGU93019.1 | | Wei/43 | ABD79101.1 | | FM/47 | AAA67338.1 | | Den/57 | ABD15258.1 | | NJ/76 | AGB51356.1 | | USSR/77 | ABD95350.1 | | Bra/78 | ABO38065.1 | | CA/78 | ABP49338.1 | | Chi/83 | ABO38340.1 | | Sing/86 | ABO38395.1 | | TX/91 | ACF41933.1 | | Bei/95 | ACF41867.1 | | NC/99 | AAP34324.1 | | SI/06 | ABU99109.1 | | Bris/07 | ADE28750.1 | | CA/09 | ACP41953.1 | | MI/15 | AMV49034.1 | |

The third method we used to calculate antigenic distance was antigenic cartography (8,9). The goal of antigenic cartography is to estimate distances between strains from immunological data by a dimension reduction algorithm. We used the HAI titers against the panel of strains from our cohort data for antigenic cartography. First we had to create a titer matrix, which has one row for each individual in the dataset and one row for each strain observed in the dataset.

We used the methods from the Racmacs (10) R package for antigenic cartography, which accepted our titer matrix as the input. Racmacs uses a variant of classical multidimensional scaling (MDS) for the dimension reduction algorithm, which finds a two-dimensional representation of the data that minimizes a particular loss function. In simpler terms, Racmacs finds a way to represent our data in two dimensions that preserves the most amount of information possible. Classical MDS does not account for missing or censored data, which are both present in our titer matrices. However, Racmacs implements an optimization routine called dimensional annealing to numerically minimize the classical MDS loss function.

First, the censored and missing data points are randomly imputed, taking into account only the range of the immunological values provided. We pass the imputed data to a standard implementation of classical MDS and calculate the resulting stress (another term for the loss function in this scenario). Then, the data are passed to a numerical optimization routine (L-BGFS) which relaxes the coordinates in the reduced dimensional space in order to reduce the stress. We performed 100 optimization rounds for our cartographic maps in order to account for the randomness in the algorithm. We repeated the entire map creation process 100 times to allow for different initial conditions which could lead the optimizer to local minima, and chose the best overall map after all optimizations.

Once we obtained the optimized antigenic map, we found the positions of each strain in the resulting two-dimensional space and calculated the pairwise Euclidian distances, giving us the pairwise cartographic distances.

## 2.2 Likelihood adjustment for censoring

In the following sections about estimating our various metrics, we will employ the same correction to the regression model likelihood to adjust for censoring. Letting represent an HAI titer measurement, and the outcome for a regression model, if we say that

we can adjust for censoring in by writing the likelihood function as

where is the probability density function of and is the normal density function. Here, is the lower censoring bound and is the upper censoring bound.

The censoring bounds are easy to determine as long as we know how our assay of interest was conducted. HAI is a serial dilution assay, and so produces interval censored values, and has both a lower and an upper limit of detection (LoD). For an observed HAI value , we assume there is some latent “true” dilution which will be a positive real number that represents the minimal dilution where hemagglutination is not observed. The latent dilution can never be observed because of how we conduct the assay. First, we choose a starting dilution, which is constrained by the proportion of reagents we need to add to our assay. We call the starting dilution and it is 10 in our dataset. That is, the HAI assays in our study were performed with an initial dilution of 1:10 serum to other assay ingredients.

If we observe hemagglutination at this starting titer, we record the assay value as below the limit of detection – by convention, this is often recorded as half the limit of detection, as it is in our study data. Although it is important to note that as long as we keep track of which assay measurements are below the limit of detection, these values are really meaningless and such not be treated as numbers (but they usually are).

Regardless, we will dilute the assay solution 2-fold and look for agglutination again, continuing to dilute the amount of serum in solution. The maximal dilution we perform for the assay, is the upper limit of detection. While 20480 was the theoretical upper LoD in our study, we did not observe any values at this level.

Any result that is within the limits of detection is still interval censored, i.e., we only know that it lies within a particular interval. For example, if we observe agglutination at a dilution of 1:40, but not at 1:20, we know that the true minimal inhibition dilution is somewhere between 20 and 40. We know it is higher than 20, and less than 40, but not where it falls in that range.

For an HAI titer assay conducted in this manner, we can write the censoring bounds and for a given titer (leaving out subscripts from this formula for simplicity) on the log scale as

We use this same censoring correction in all of the corrected models.

## 2.3 Current immunogenicity metric details

Our methods for calculating the metrics are laid out without much technical detail in the separate methods document, so we refer readers looking for a tutorial to that document instead. In this section we will briefly list the relevant technical details for our estimation models.

We will use the following notation to define our formulas for the current immunogenicity metrics.

* indexes study subjects, where is the sample size.
* The variable indexes different strains, and is the homologous strain.
* We define seroprotection as a post-vaccination HAI titer greater than 40, and seroconversion as a post-vaccination fold-change of 4-fold or higher along with seroprotection.

### 2.3.1 Homologous GMT (magnitude)

The simple formula for the homologous GMT without considering censoring is

The regression model we fit to estimate the homologous GMT in the same framework as our novel metrics is given as

In this model, the parameter estimates the mean of the outcome values.

### 2.3.2 Seroconversion rate (breadth)

The simple formula for the seroconversion rate without considering censoring or the overall estimation framework is

where is the indicator function.

The regression model we fit to estimate the seroconversion rate is

where estimates the overall seroconversion rate.

### 2.3.3 Overall GMT (total strength)

The simple formula for the overall GMT is

and the regression model is

This regression model is exactly the same as the model for , but for the model we only include homologous strains in the fitting, and for the , which is estimated by as well, we include all strains in the model fitting to get an overall estimate.

## 2.4 Novel immunogenicity metric details

In order to calculate our novel immunogenicity metrics, we first need to fit a summary antibody landscape model to the titer data. Letting be the antigenic distance from the current season’s vaccine to assay strain (recall that we treat all seasons independently and fit a separate model to each season) we define a multilevel linear regression model to construct the summary antibody landscape. The multilevel model is defined as

where we have a random intercept and random antigenic distance slope for each individual , and the random effects for each individual are allowed to be correlated with a correlation matrix shared by all individuals. We will discuss priors in the model implementation section, but for all other models we have chosen weakly informative independent priors. However, allowing the random effects to be correlated as we do here is often beneficial for model fitting.

In order to calculate our metrics from this model we need to calculate what brms calls expectation predictions or epreds. This refers to the predicted values for some values of the predictors. Specifically, we use the population average epreds which are estimated conditionally on the random effects but do not include random effects deviations in the predictions, i.e.,

We calculated these epreds on an interpolated grid of normalized antigenic distance values, . The granularity of the predictions can be increased if desired, but we found this granularity to give a good balance between resolution and memory usage.

Our novel metrics are then derived from the epreds.

### 2.4.1 Intercept (magnitude)

The intercept is simply defined as

That is, we average over the posterior samples of the epreds for an antigenic distance value of 0.

### 2.4.2 Proportion above threshold (breadth)

The proportion above the threshold is defined as the distance value where the horizontal line (on the log scale) intersects the summary antibody landscape. That is,

which for the linear regression model has a simple closed form of

but might not have a simple closed form for more complicated models, so we find the interpolated value of in our grid which minimizes .

### 2.4.3 Area under the curve (total strength)

The formula for the area under the curve (between the normalized antigenic distance values of 0 and 1 is

For each posterior sample, we approximate this integral numerically using the trapezoid method implemented by pracma::trapz() with default iteration settings.

## 2.5 Model implementation general details

Whenever we summarize posterior samples to obtain a point estimate and credible interval, we calculate the mean and 95% highest density continuous interval (HDCI), implemented as tidybayes::mean\_hdci().

For all of our regression models, we specified general weakly informative priors. Because the models converged and moved away from the priors, the specific priors are not very important. That being said, we used priors for regression coefficients and priors for variance parameters, where is the location-scale Student’s distribution with degrees of freedom, location parameter and scale parameter ; and is the half-Student’s distribution constrained to be strictly positive with the same arguments. We use 3 degrees of freedom for all of our Student’s priors because the variance is infinite for smaller degrees of freedom. Using allows for the prior distribution to have fat tails, so if the data support a large parameter value the model likelihood will allow that, but this tends to be much less pathological than trying to sample from priors with or . For our multilevel summary landscape models, we specified the priors on the covariance matrix using a Cholesky factor decomposition with an prior (11) on the Cholesky factor and independent priors on the vector of variance parameters.

## 2.6 Intraclass correlation (ICC) analysis details

To estimate the intraclass correlation coefficients, we employed a second Bayesian model. We fit a one-way random effects model where the outcome was the posterior samples of a given metric (from a given season) and the predictors were a global intercept parameter and a random intercept for the subsample. We can write the regression model for calculating the ICC as

where indexes the subsamples from the same cohort. We also fit the models separately for each of the six different metrics, using the p-epitope and cartographic distances for the novel metrics, and for metrics calculated with and without the censoring correction.

Once we fit the regression model, we calculated the posterior samples of the ICC as

and summarized this as the mean and 95% HDCI over the posterior samples, which is what we report as our ICC results.

# 3. Supplementary results

We conducted a number of additional analyses which are tangential to those presented in the main text. However, they provide more context to our results so we present them here.

## 3.1 Additional strain information

We use abbreviated strain names throughout our manuscript. The abbreviated strain names along with the full strain name used on GenBank is shown in [Table 2](#tbl-names).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2: Full strain names and abbreviations for each strain used in our study.   | Short Name | Full Name | | --- | --- | | SC/18 | A/H1N1/South Carolina/1/1918 | | PR/34 | A/H1N1/Puerto Rico/8/1934 | | Wei/43 | A/H1N1/Weiss/1943 | | FM/47 | A/H1N1/Fort Monmouth/1/1947 | | Den/57 | A/H1N1/Denver/1957 | | NJ/76 | A/H1N1/New Jersey/8/1976 | | USSR/77 | A/H1N1/Ussr/90/1977 | | Bra/78 | A/H1N1/Brazil/11/1978 | | CA/78 | A/H1N1/California/10/1978 | | Chi/83 | A/H1N1/Chile/1/1983 | | Sing/86 | A/H1N1/Singapore/6/1986 | | TX/91 | A/H1N1/Texas/36/1991 | | Bei/95 | A/H1N1/Beijing/262/1995 | | NC/99 | A/H1N1/New Caledonia/20/1999 | | SI/06 | A/H1N1/Solomon Islands/3/2006 | | Bris/07 | A/H1N1/Brisbane/59/2007 | | CA/09 | A/H1N1/California/07/2009 | | MI/15 | A/H1N1/Michigan 45/2015 | |

For the duration of our study, the A(H1N1) vaccine component was only updated once. The Fluzone standard dose vaccine included the CA/09 strain from 2013/14 through 2016/17, and was updated to contain the MI/15 strain in 2017/18.

The panel of historical strains for A(H1N1) stayed fairly consistent for the duration of our study, and was the same at all study sites within a given season. The 2013/14 season used a panel of 16 strains, the 2014/15 and 2015/16 seasons used a panel of 15 seasons, and the 2016/17 and 2017/18 seasons used a panel of 16 strains. The additional strain used in 2013/14 was PR/34, which was only used during that season. In 2016/2017 after the MI/2015 virus was isolated and sequenced, that virus was added to the panel of viruses for all following study seasons.

## 3.2 Demographics information

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 3: Participant demographics for each of the seasonal cohorts in our study sample.   | **Characteristic** | **2013/14**  N = 128 | **2014/15**  N = 236 | **2015/16**  N = 209 | **2016/17**  N = 328 | **2017/18**  N = 255 | **Overall**  N = 1156 | | --- | --- | --- | --- | --- | --- | --- | | Sex assigned at birth, n (%) |  |  |  |  |  |  | | Female | 97 (76) | 170 (72) | 156 (75) | 221 (67) | 142 (56) | 786 (68) | | Male | 31 (24) | 66 (28) | 53 (25) | 107 (33) | 113 (44) | 370 (32) | | Race/Ethnicity, n (%) |  |  |  |  |  |  | | White | 89 (70) | 158 (67) | 147 (70) | 235 (72) | 203 (80) | 832 (72) | | Black or African American | 24 (19) | 39 (17) | 35 (17) | 43 (13) | 17 (7) | 158 (14) | | Other | 13 (10) | 17 (7) | 13 (6) | 31 (9) | 25 (10) | 99 (9) | | Hispanic or Latino | 2 (2) | 21 (9) | 13 (6) | 18 (5) | 10 (4) | 64 (6) | | Unknown | 0 (0) | 1 (0) | 1 (0) | 1 (0) | 0 (0) | 3 (0) | | Age (years), Median (Min - Max) | 58 (23 - 81) | 57 (20 - 83) | 56 (21 - 84) | 44 (18 - 82) | 25 (12 - 83) | 48 (12 - 84) | | Birth year, Median (Min - Max) | 1956 (1932 - 1990) | 1957 (1932 - 1994) | 1960 (1932 - 1994) | 1972 (1934 - 1998) | 1991 (1934 - 2006) | 1968 (1932 - 2006) | | Study site, n (%) |  |  |  |  |  |  | | FL | 60 (47) | 150 (64) | 128 (61) | 119 (36) | 0 (0) | 457 (40) | | PA | 68 (53) | 86 (36) | 81 (39) | 64 (20) | 0 (0) | 299 (26) | | UGA | 0 (0) | 0 (0) | 0 (0) | 145 (44) | 255 (100) | 400 (35) | |

Participant demographics for each seasonal cohort are shown in [Table 3](#tbl-demographics), including sex assigned at birth, self-reported race/ethnicity, age in years, birth year, and the study site an individual enrolled at. The majority of participants in our study were white and female. Demographics were fairly consistent across the seasonal cohorts with the exception of age. The UGA study site which began recruiting patients in 2016/2017 recruited younger patients, which was not a primary focus of the other two study sites, so the 2016/17 and 2017/18 cohorts have more young people than the preceding cohorts.

## 3.3 Summary antibody landscapes for all seasons

Here, we include the summary antibody landscapes for all of the seasons in our study. The general trends were the same across all landscapes, so we chose to report the 2016/2017 landscape in the main text due to the large number of participants.

* [Figure 1](#fig-landscape-13) shows the summary landscape for the 2013/14 seasonal cohort.
* [Figure 2](#fig-landscape-14) shows the summary landscape for the 2014/15 seasonal cohort.
* [Figure 3](#fig-landscape-15) shows the summary landscape for the 2015/16 seasonal cohort.
* [Figure 4](#fig-landscape-16) shows the summary landscape for the 2016/17 seasonal cohort. This figure is also included in the main text but it is reproduced here for easier comparison.
* [Figure 5](#fig-landscape-17) shows the summary landscape for the 2017/18 seasonal cohort.

|  |
| --- |
| Figure 1: Raw data and summary antibody landscapes for the 2013 - 2014 influenza season. Each point shows the post-vaccination HAI titer to a specific strain with a specified normalized antigenic distance from the vaccine strain (CA/09 in 2013/14). The dashed line and envelope show the mean and 95% credible interval (CrI) of the posterior summary antibody landscape. |

|  |
| --- |
| Figure 2: Raw data and summary antibody landscapes for the 2014 - 2015 influenza season. Each point shows the post-vaccination HAI titer to a specific strain with a specified normalized antigenic distance from the vaccine strain (CA/09 in 2014/15). The dashed line and envelope show the mean and 95% credible interval (CrI) of the posterior summary antibody landscape. |

|  |
| --- |
| Figure 3: Raw data and summary antibody landscapes for the 2014 - 2015 influenza season. Each point shows the post-vaccination HAI titer to a specific strain with a specified normalized antigenic distance from the vaccine strain (CA/09 in 2014/15). The dashed line and envelope show the mean and 95% credible interval (CrI) of the posterior summary antibody landscape. |

|  |
| --- |
| Figure 4: Raw data and summary antibody landscapes for the 2016 - 2017 influenza season. Each point shows the post-vaccination HAI titer to a specific strain with a specified normalized antigenic distance from the vaccine strain (CA/09 in 2016/17). The dashed line and envelope show the mean and 95% credible interval (CrI) of the posterior summary antibody landscape. |

|  |
| --- |
| Figure 5: Raw data and summary antibody landscapes for the 2017 - 2018 influenza season. Each point shows the post-vaccination HAI titer to a specific strain with a specified normalized antigenic distance from the vaccine strain (MI/15 in 2017/18). The dashed line and envelope show the mean and 95% credible interval (CrI) of the posterior summary antibody landscape. |

## 3.4 Summary landscape metrics for all seasons

[Table 4](#tbl-metrics-all) shows the summary metrics for all of the seasonal cohorts. The trends for the other seasons are similar to the 2016/17 results, but we include the other seasons here for completeness. In general the metrics are overall more optimistic in the 2016/2017 and 2017/18 cohorts, but this is because those cohorts had much more young people relative to the total sample size for that season. We elected to present these results in the main text because the signal is consistent with the other seasons, but easier to see. We chose the 2016/17 cohort for the main results because it used the same vaccine as the older cohorts and contains a mix of individuals from both, so it is easier to compare to the older cohorts directly.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: Current and novel vaccine immunogenicity metrics for each season, shown both with and without the censoring correction. The ‘Current’ metric set uses the homologous GMT for magnitude, the seroconversion rate across the historical panel for breadth, and the GMT across all strains for the total strength. The ‘Novel’ metric sets are based on the corresponding summary landscape. The novel metrics are regression line intercept for magnitude, proportion of the line above a titer of 1:40 for breadth, and area under the curve for total strength. All metrics were derived from bayesian regression models and numbers shown are the posterior mean and 95% CrI.   | Season | Metric Set | Censoring correction | | | LoD set to 5 | | | | --- | --- | --- | --- | --- | --- | --- | --- | | Magnitude | Breadth | Total Strength | Magnitude | Breadth | Total Strength | | 2013 - 2014 | Current | 3.45 (3.15, 3.75) | 0.09 (0.07, 0.10) | 2.09 (1.96, 2.21) | 3.04 (2.75, 3.29) | 0.09 (0.07, 0.10) | 2.28 (2.20, 2.36) | | Novel (Cartographic) | 3.65 (3.30, 4.00) | 0.25 (0.15, 0.37) | 2.35 (2.09, 2.62) | 3.21 (2.94, 3.45) | 0.13 (0.00, 0.26) | 2.43 (2.26, 2.60) | | Novel (p-Epitope) | 2.71 (2.39, 3.04) | 0.00 (0.00, 0.00) | 2.32 (2.08, 2.57) | 2.56 (2.35, 2.81) | 0.00 (0.00, 0.00) | 2.39 (2.23, 2.54) | | Novel (Temporal) | 2.29 (1.98, 2.58) | 0.00 (0.00, 0.00) | 1.99 (1.72, 2.29) | 2.41 (2.21, 2.61) | 0.00 (0.00, 0.00) | 2.20 (2.03, 2.37) | | 2014 - 2015 | Current | 3.61 (3.40, 3.82) | 0.08 (0.07, 0.08) | 2.26 (2.18, 2.34) | 3.21 (3.03, 3.41) | 0.08 (0.07, 0.08) | 2.36 (2.31, 2.41) | | Novel (Cartographic) | 3.88 (3.62, 4.11) | 0.32 (0.26, 0.39) | 2.51 (2.35, 2.67) | 3.38 (3.20, 3.56) | 0.22 (0.15, 0.31) | 2.52 (2.41, 2.63) | | Novel (p-Epitope) | 2.84 (2.59, 3.07) | 0.01 (0.00, 0.07) | 2.46 (2.29, 2.62) | 2.70 (2.52, 2.87) | 0.00 (0.00, 0.00) | 2.48 (2.37, 2.59) | | Novel (Temporal) | 2.31 (2.12, 2.49) | 0.00 (0.00, 0.00) | 2.22 (2.06, 2.40) | 2.44 (2.32, 2.56) | 0.00 (0.00, 0.00) | 2.32 (2.21, 2.42) | | 2015 - 2016 | Current | 3.93 (3.73, 4.13) | 0.07 (0.06, 0.08) | 1.73 (1.64, 1.82) | 3.45 (3.26, 3.66) | 0.07 (0.06, 0.08) | 2.09 (2.04, 2.14) | | Novel (Cartographic) | 4.69 (4.42, 4.90) | 0.34 (0.31, 0.38) | 2.21 (2.05, 2.37) | 3.88 (3.69, 4.07) | 0.29 (0.25, 0.34) | 2.37 (2.26, 2.47) | | Novel (p-Epitope) | 2.99 (2.72, 3.27) | 0.03 (0.00, 0.14) | 2.19 (2.00, 2.37) | 2.84 (2.67, 3.01) | 0.00 (0.00, 0.00) | 2.36 (2.25, 2.48) | | Novel (Temporal) | 2.29 (2.12, 2.47) | 0.00 (0.00, 0.00) | 1.41 (1.20, 1.60) | 2.35 (2.24, 2.46) | 0.00 (0.00, 0.00) | 1.94 (1.84, 2.04) | | 2016 - 2017 | Current | 5.22 (5.07, 5.39) | 0.10 (0.09, 0.11) | 2.64 (2.58, 2.70) | 4.72 (4.55, 4.88) | 0.10 (0.09, 0.11) | 2.57 (2.52, 2.61) | | Novel (Cartographic) | 5.24 (5.06, 5.44) | 0.50 (0.48, 0.52) | 3.00 (2.88, 3.12) | 4.61 (4.44, 4.78) | 0.45 (0.43, 0.48) | 2.83 (2.73, 2.92) | | Novel (p-Epitope) | 4.52 (4.36, 4.69) | 0.60 (0.55, 0.64) | 3.25 (3.13, 3.37) | 4.05 (3.90, 4.20) | 0.52 (0.47, 0.56) | 3.04 (2.94, 3.14) | | Novel (Temporal) | 3.29 (3.17, 3.43) | 0.14 (0.08, 0.19) | 2.19 (2.05, 2.34) | 3.00 (2.89, 3.11) | 0.02 (0.00, 0.07) | 2.28 (2.19, 2.37) | | 2017 - 2018 | Current | 5.57 (5.40, 5.75) | 0.10 (0.09, 0.11) | 2.14 (2.06, 2.23) | 5.07 (4.90, 5.26) | 0.10 (0.09, 0.11) | 2.35 (2.30, 2.40) | | Novel (Cartographic) | 5.95 (5.77, 6.14) | 0.51 (0.50, 0.53) | 3.06 (2.94, 3.18) | 5.18 (5.00, 5.34) | 0.50 (0.48, 0.51) | 2.98 (2.89, 3.07) | | Novel (p-Epitope) | 5.57 (5.37, 5.76) | 0.59 (0.57, 0.62) | 3.41 (3.29, 3.53) | 4.95 (4.78, 5.13) | 0.58 (0.55, 0.60) | 3.26 (3.17, 3.36) | | Novel (Temporal) | 3.19 (3.04, 3.33) | 0.07 (0.02, 0.11) | 1.72 (1.58, 1.86) | 3.03 (2.93, 3.12) | 0.02 (0.00, 0.07) | 2.07 (2.00, 2.15) | |

## 3.5 Subsample metrics plot for all seasons

We also repeated the subsampling analysis using data from the other cohorts. Again, because the signal is stronger in the 2016/17 and 2017/18 cohorts we presented those results in the main text. The results for the other seasons showed a similar pattern, and we include them here for completeness.

* [Figure 6](#fig-subsamples-13) shows the posterior distribution of the vaccine metrics for all of the subsamples from the 2013/14 cohort data.
* [Figure 7](#fig-subsamples-14) shows the posterior distribution of the vaccine metrics for all of the subsamples from the 2014/15 cohort data.
* [Figure 8](#fig-subsamples-15) shows the posterior distribution of the vaccine metrics for all of the subsamples from the 2015/16 cohort data.
* [Figure 9](#fig-subsamples-16) shows the posterior distribution of the vaccine metrics for all of the subsamples from the 2016/17 cohort data. This figure is included in the main text but reproduced here for easier comparisons.
* [Figure 10](#fig-subsamples-17) shows the posterior distribution of the vaccine metrics for all of the subsamples from the 2017/18 cohort data.

All of these figures show the current metrics for magnitude, breadth, and total strength, and the novel metrics usingt both the cartographic distance and the p-Epitope distance for each subsample. The black circles show samples from the posterior distribution of each metric. The red dotted line shows the overall mean metric estimate across the subsample, and the red x for each subsample shows the mean metric estimate for that subsample. In general, metrics with lower ICCs (less variation explained by subsample grouping) will have group means that are more similar to the overall mean. We show only 1000 posterior samples for each subsample/metric to avoid unnecessary overplotting.

|  |
| --- |
| Figure 6: Estimated immunogenicity metrics for each simulated lab drawn from the 2013/14 subcohort data. |

|  |
| --- |
| Figure 7: Estimated immunogenicity metrics for each simulated lab drawn from the 2014/15 subcohort data. |

|  |
| --- |
| Figure 8: Estimated immunogenicity metrics for each simulated lab drawn from the 2015/16 subcohort data. |

|  |
| --- |
| Figure 9: Estimated immunogenicity metrics for each simulated lab drawn from the 2016/17 subcohort data. |

|  |
| --- |
| Figure 10: Estimated immunogenicity metrics for each simulated lab drawn from the 2017/18 subcohort data. |

## 3.6 Intraclass correlation (ICC) analysis for all seasons

To quantify the results in the subsampling figures, we also include the ICC estimates for each of the seasonal cohorts. Yet again, the results were similar to the 2016/17 cohort with some outliers, but we include them here for completeness ([Table 5](#tbl-iccs-all)).

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 5: Intraclass correlation coefficients (ICCs) for consistency across the subsampled studies. Each number shown is the posterior mean and 95% CrI for the ICC, which we calculated as the between-groups variance for subgroups divided by the between-groups variance for subgroups plus the residual error variance. An ICC closer to zero indicates that little of the variance in metric estimates is due to variability across subsamples, while an ICC closer to one indicates that variability across subsamples makes up the majority of the variation.   | Season | Metric Set | Magnitude | Breadth | Total Strength | | --- | --- | --- | --- | --- | | 2013 - 2014 | Current | 0.16 (0.09, 0.24) | 0.04 (0.00, 0.42) | 0.55 (0.39, 0.79) | | Novel (Cartographic) | 0.37 (0.24, 0.52) | 0.30 (0.20, 0.45) | 0.33 (0.05, 0.52) | | Novel (p-Epitope) | 0.65 (0.52, 0.78) | 0.58 (0.25, 0.74) | 0.52 (0.40, 0.65) | | 2014 - 2015 | Current | 0.43 (0.31, 0.61) | 0.00 (0.00, 0.00) | 0.89 (0.81, 0.95) | | Novel (Cartographic) | 0.52 (0.39, 0.67) | 0.35 (0.00, 0.55) | 0.66 (0.52, 0.78) | | Novel (p-Epitope) | 0.63 (0.37, 0.81) | 0.59 (0.45, 0.75) | 0.68 (0.57, 0.83) | | 2015 - 2016 | Current | 0.31 (0.21, 0.43) | 0.61 (0.00, 0.92) | 0.81 (0.68, 0.93) | | Novel (Cartographic) | 0.53 (0.00, 0.80) | 0.39 (0.00, 0.71) | 0.42 (0.00, 0.68) | | Novel (p-Epitope) | 0.73 (0.58, 0.87) | 0.72 (0.60, 0.82) | 0.70 (0.53, 0.85) | | 2016 - 2017 | Current | 0.50 (0.31, 0.69) | 0.22 (0.00, 0.90) | 0.86 (0.73, 0.95) | | Novel (Cartographic) | 0.61 (0.48, 0.76) | 0.32 (0.00, 0.76) | 0.56 (0.42, 0.69) | | Novel (p-Epitope) | 0.78 (0.68, 0.90) | 0.47 (0.08, 0.85) | 0.72 (0.50, 0.87) | | 2017 - 2018 | Current | 0.31 (0.21, 0.45) | 0.31 (0.00, 0.91) | 0.63 (0.00, 0.96) | | Novel (Cartographic) | 0.76 (0.56, 0.89) | 0.52 (0.00, 0.79) | 0.81 (0.70, 0.90) | | Novel (p-Epitope) | 0.81 (0.67, 0.90) | 0.28 (0.00, 0.68) | 0.58 (0.37, 0.83) | |

# 4. References

1. Boyoglu-Barnum S, Ellis D, Gillespie RA, Hutchinson GB, Park YJ, Moin SM, et al. [Quadrivalent influenza nanoparticle vaccines induce broad protection](https://doi.org/10.1038/s41586-021-03365-x). Nature. 2021-03-24, 2021-03;592(7855):623–8.

2. Hinojosa M, Shepard SS, Chung JR, King JP, McLean HQ, Flannery B, et al. [Impact of immune priming, vaccination, and infection on influenza A(H3N2) antibody landscapes in children](https://doi.org/10.1093/infdis/jiaa665). The Journal of Infectious Diseases. 2020-10-22, 2020-10;

3. Auladell M, Phuong HVM, Mai LTQ, Tseng YY, Carolan L, Wilks S, et al. [Influenza virus infection history shapes antibody responses to influenza vaccination](https://doi.org/10.1038/s41591-022-01690-w). Nature Medicine. 2022 Feb;28(2):363–72.

4. Yang B, Lessler J, Zhu H, Jiang CQ, Read JM, Hay JA, et al. [Life course exposures continually shape antibody profiles and risk of seroconversion to influenza](https://doi.org/10.1371/journal.ppat.1008635). PLOS Pathogens. 2020 Jul;16(7):e1008635.

5. Gupta V, Earl DJ, Deem MW. [Quantifying influenza vaccine efficacy and antigenic distance](https://doi.org/10.1016/j.vaccine.2006.01.010). Vaccine. 2006 May;24(18):3881–8.

6. Burke DF, Smith DJ. [A Recommended Numbering Scheme for Influenza A HA Subtypes](https://doi.org/10.1371/journal.pone.0112302). PLoS ONE. 2014 Nov;9(11):e112302.

7. Hamming RW. [Error detecting and error correcting codes](https://doi.org/10.1002/j.1538-7305.1950.tb00463.x). The Bell System Technical Journal. 1950 Apr;29(2):147–60.

8. Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, et al. [Mapping the Antigenic and Genetic Evolution of Influenza Virus](https://doi.org/10.1126/science.1097211). Science. 2004 Jul;305(5682):371–6.

9. Fonville JM, Fraaij PLA, de Mutsert G, Wilks SH, van Beek R, Fouchier RAM, et al. [Antigenic maps of influenza A(H3N2) produced with human antisera obtained after primary infection](https://doi.org/10.1093/infdis/jiv367). Journal of Infectious Diseases. 2015-07-03, 2015-07;213(1):31–8.

10. Wilks S. Racmacs: Antigenic cartography macros. 2024.

11. Lewandowski D, Kurowicka D, Joe H. [Generating random correlation matrices based on vines and extended onion method](https://doi.org/10.1016/j.jmva.2009.04.008). Journal of Multivariate Analysis. 2009 Oct;100(9):1989–2001.