Evidence of original antigenic sin in a cohort study of the seasonal influenza vaccine

# Authors

Note for Andreas: when anyone else contributes, I’m happy to add them. But I won’t randomly add people who never saw the manuscript this time.

* W. Zane Billings (ORCID: 0000-0002-0184-6134);
* Yang Ge (0000-0001-5100-0703);
* Amanda L. Skarlupka (0000-0002-3654-9076);
* Ted M. Ross (0000-0003-1947-7469);
* Andreas Handel (0000-0002-4622-1146).

# Author affiliations

* : College of Public Health, University of Georgia, Athens, GA, USA.
* : Center for the Ecology of Infectious Diseases, University of Georgia, Athens, GA, USA.
* : Florida Research & Innovation Center, Cleveland Clinic, Port St. Lucie, FL, USA

: ALS is currently an employee of the National Cancer Institute, National Institutes of Health, Bethesda, Maryland, USA. This work was prepared while ALS was an employee of the University of Georgia, Athens, GA, USA. The opinions expressed here are the authors’ own and do not express the views of the National Cancer Institute, the National Institutes of Health, or the Department of Health and Human Services.

# Author contributions (CRediT)

* ZB: conceptualization, data curation, formal analysis, methodology, software, visualization, writing (original draft preparation), writing (review and editing)
* YG: data curation, methodology, writing (review and editing)
* ALS: data curation, methodology, writing (review and editing)
* TMR: data curation, investigation, resources, writing (review and editing)
* AH: conceptualization, methodology, supervision, writing (original draft preparation), writing (review and editing), funding acquisition

# Funding sources

* ZB: none declared
* YG: ?
* ALS: none declared
* TMR: Eminent Scholar, Georgia Research Alliance
* AH: partial funding from NIH contracts/grants LIST THE NUMBERS HERE.

# Potential target journals

* Journal of Virology or mBio (ASV journals, both published computational antigenic distance research before)
* Influenza and other respiratory diseases (we know Ben is interested in this topic, so presumably the other editors will be)
* eLife (good journal for similar computational research but since they don’t really “publish” anymore it scares me a bit)

# Abstract

**Introduction:**

**Methods:**

**Results:**

**Discussion:**

# Introduction

CDC estimates of vaccine effectiveness of the seasonal influenza vaccine in the United States ranged from to from 2009 through 2024 (CITE THIS). Proposed explanations for the low effectiveness of flu vaccines include

* VE for flu is bad, but we don’t really understand what that means for heterologous immunity.
* We previously used strain-specific effects, but quantifying antigenic distance can let us make predictions and generalize our findings.
* Imprinting/OAS suggests that effects should be nonlinear.
* However, it’s unclear how we should actually measure antigenic distance.
* We did this study where we modeled the effect of antigenic distance using different distance quantification methods.

# Methods

## Data source

The data for our study are from a prospective, ongoing human vaccination cohort study which has been described in detail previously [@nunez2017impact; @carlock2019impact; @abreu2020iga]. Briefly, the study recruited participants at three study sites: Pittsburgh, PA, USA, and Port St. Lucie, FL, USA beginning in the 2013/14 influenza season (approximately September through March (CITE THE CDC THING)) and continuing through the 2016/17 influenza season. Beginning in January 2017, the study moved to Athens, GA, USA. Participants received Fluzone (Sanofi-Pasteur) vaccines and donated two serum samples, one before being vaccinated, and one at a follow-up visit approximately 21 days after the first visit. The study was a prospective, open cohort design where individuals could enroll in multiple years in the study, but were not required to re-enroll in every consecutive year. Individuals under 65 received a standard dose Fluzone vaccination, and individuals aged 65 and older were given the choice between standard dose (SD) and high dose (HD) Fluzone vaccines.

Researchers used each serum sample for a panel of hemagglutination inhibition (HAI) assays to the homologous strains, included in the seasonal vaccine formulation, and a panel of historical, heterologous influenza virus strains. Strains included in the historical panel represented the major clades of circulating influenza viruses. In each season, all prior vaccine strains from 2012 onwards were included in the historical panel. See the Supplement for details on the Fluzone vaccine formulation and for a list of strains used in each season.

For our secondary data analysis, we extracted previously deidentified (non-human subjects) records the 2013/14 through 2017/18 influenza seasons. We included all participants with both pre-vaccination and post-vaccination blood samples in our analysis. Our primary outcome of interest was the reciprocal post-vaccination HAI titer, which we log transformed:

We devided the raw reciprocal titer by before taking the log because the HAI assay had a lower limit of detection (LoD) of , and an upper LoD of . Values below the LoD were coded as 5 in the dataset. After our transformation, values below the LoD had a value of . All observed values in our dataset were below the upper LoD. We used the same outcome definitions defined in our previous work on this dataset (CITE OUR PAPER HERE).

## Antigenic distance calculation

Each HAI assay in our dataset has an associated subtype, vaccine strain, and assay strain. Here, we use the word “subtype” for simplicity to describe both Influenza A subtypes (H1N1 and H3N2), and Influenza B lineages (Victoria-like and Yamagata-like). The vaccine strains associated with an HAI assay are the strains used in the Fluzone vaccine formulation in the season when the serum sample was drawn. Each assay has three or four associated vaccine strains, depending on whether the individual who gave the serum sample received a trivalent or quadrivalent vaccine (see Supplement for details on the vaccine formulations). The assay strain for a vaccine is the strain of the actual virus added to the serum sample during the HAI assay.

We computed the pairwise antigenic distance for all strains used in the dataset (again, see the Supplement for a complete list). We used four different methods to compute the antigenic distance: temporal distance, dominant -Epitope distance, Grantham’s distance, and cartographic distance. For complete details on antigenic distance calculation, see the Supplement. Briefly: temporal distance is the absolute difference in the years of isolation between the two strains. Dominant -Epitope distance is the maximum length-normalized Hamming distance across the five major epitope sites on the HA head. Grantham’s distance is a weighted distance based on biochemical properties that considers how different two differing residues at the same position are. Finally, cartographic distance is the euclidean distance between strains on antigenic cartography map. Antigenic cartography, implemented in Racmacs, uses multidimensional scaling to reduce the matrix of post-vaccination HAI data into fewer dimensions while preserving the euclidean distance as much as possible. (CITE ALL THESE!!!)

For our models, we only considered the antigenic distance between the assay strain and the vaccine strain of the same subtype for a given HAI assay. Some of the assay strains used were influenza B strains isolated before the Victoria/Yamagata lineage divergence. Because our main question was about the antigenic distance, we compared pre-divergence B strains to both the Yamagata and Victoria vaccine strains in our analyses.

## Statistical analyses

## Implementation

# Results

## Data description

Our dataset included 62739 paired (pre-vaccination and post-vaccination) HAI titer measurements drawn from 725 who contributed 1372 person-years to the study across three different study sites. The contributions of paired measurements, person-years, and unique participants from each study site are shown in [Table 1](#tbl-counts). Each person-year represented in the data contributed median 47 paired HAI assays with a minimum contribution of 8 assays to a maximum contribution of 52 assays.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Counts of paired HAI assays, person years, and unique participants contributed by each study site for the duration of the study. Note that the PA and FL study sites operated from September 2013 to December 2016 and the UGA study site began operating in January 2017 (during the 2016/17 influenza season).   |  | Season | | | | |  | | --- | --- | --- | --- | --- | --- | --- | |  | 2013/14 | 2014/15 | 2015/16 | 2016/17 | 2017/18 | Total | | Paired HAI assays |  |  |  |  |  |  | | FL | 2459 | 6597 | 7624 | 6967 | 0 | 23647 | | PA | 2852 | 5434 | 5713 | 5281 | 0 | 19280 | | UGA | 0 | 0 | 0 | 6932 | 12880 | 19812 | | Overall | 5311 | 12031 | 13337 | 19180 | 12880 | 62739 | | Person years |  |  |  |  |  |  | | FL | 60 | 150 | 150 | 138 | 0 | 498 | | PA | 91 | 128 | 117 | 119 | 0 | 455 | | UGA | 0 | 0 | 0 | 148 | 271 | 419 | | Overall | 151 | 278 | 267 | 405 | 271 | 1372 | | New participants |  |  |  |  |  |  | | FL | 60 | 113 | 38 | 31 | 0 | 242 | | PA | 91 | 57 | 2 | 16 | 0 | 166 | | UGA | 0 | 0 | 0 | 148 | 169 | 317 | | Overall | 151 | 170 | 40 | 195 | 169 | 725 | |

A summary of the demographic information for the individuals included in our analysis is shown in [Table 2](#tbl-demographics), and includes information about their reported race/ethnicity, sex assigned at birth, age at first enrollment, and year of birth (see Supplement for detailed coding descriptions). The majority of participants identified their race as White or Caucasian, and were assigned female at birth. All participants from the PA and FL study sites were adults, but the UGA study site also recruited teenagers, and all three study sites included elderly people over 65 years of age. Most participants returned to the study site in at least one subsequent year, contributing more than one person-year of data to the study.

|  |
| --- |
|  |

## Correlation between antigenic distance metrics varies by subtype

[Figure 1](#fig-raw-data-plot) shows a subsample of the observed titer measurements vs. the antigenic distance between the vaccine strain and the assay strain for four different antigenic distance methods. The trendlines are similar for SD and HD across all of the metrics and subtypes – there were never any large divergences between the doses in terms of the overall rolling average. Notably this trendline does not take the effect of age into account, so the effect of being young is conflated with the effect of receiving the standard dose vaccine. The different antigenic distance metrics also have different distributions in the set of observed variables. Some distance metrics (temporal, cartographic) tend to estimate a continuous range of distances, while others (Grantham, p-Epitope) form clusters around a few discrete values.

|  |
| --- |
| Figure 1: Post-vaccination titer on the y-axis vs. antigenic distance from the vaccine strain on the x-axis. Each subplot (facet) shows one combination of a distance metric (changes by rows) and influenza subtype or lineage (changes by columns). The trend line shown is a rolling average with a window of 3. Standard dose observations are indicated by green circles, with a solid green trendline. High dose observations are indicated by orange triangles, which a dashed orange trendline. |

# References