Feasibility of proposed dissertation aims

Zane Billings

2022-10-13

This document is based on my dissertation ideas list. This is a compilation of exploratory data analyses with some writeup to determine if specific aims would be feasible for my dissertation.

Our exploratory data analyses use vaccine cohort data collected by Ted Ross at the University of Georgia. Data are from the UGAFluVac cohort (2016 – present) as well as two previous studies conducted in Pennsylvania and Florida. All studies collected longitudinal responses to a seasonal influenza vaccine from an open cohort, and took serum samples at the day of vaccination and 21 or 28 days post-vaccination. Serum samples were used for HAI assays against the vaccine strain as well as a panel of heterologous historical strains, with the panel being updated as the vaccine recommendations changed.

If possible, we plan to incorporate additional data from several other studies, with data accessible as part of the DIVERsity grant collaboration.

Table of contents

1 Quantify the effect of preexisting immunity on heterologous humoral immune response after recieving a seasonal influenza vaccine.

- **Hypothesis**: Individuals with higher antibody titers prior to vaccination will boost less than naive individuals, but the magnitude of this difference will decrease as antigenic distance between the vaccine strain and test strain increases.
- **Significance**: Quantifying the patterns of immunity to heterologous strains is vital for evaluating the breadth of vaccine response.
- **Innovation**: Previous studies use simple measures of antigenic distance, such as year between strain isolation, and poor metrics of breadth (e.g. proportion of strains an individual seroconverts to, which is extremely sensitive to the panel of strains tested); or strains are treated as categorical (e.g. Yang's dissertation), which is a flexible modeling framework but may not be extensible to other strains. Incorporating sequence-based measures of antigenic distance provides a method of modeling immune response that may be generalizable to strains which were not explicitly tested.

1.1 Background

For a review of major work on the topic in the context of influenza, see [COLD SPRING PAPER] and [OIDTMAN REVIEW]. In short, the mechanistic models used in [VERONIKA'S PAPER] predict that fold change in antibody titer will be negatively associated with (log) prior titer, with an "antigen ceiling," or threshold, effect at high values of preexisting titer. That is, individuals will sufficiently high pre-vaccination titers will have rapidly diminishing returns (or even no boosting at all) from vaccination. Furthermore, the intercept of the response curve will be shifted upwards as vaccine dose increases, but the slope of the relationship will not change.

While the original paper tested the mechanistic model against some real data, the model only makes predictions for homologous responses (i.e., immune response to the strain that was used in the vaccine). However, activation of partially cross-protective memory responses to antigenically drifted influenza strains is a key component in determining the degree of protection conferred by vaccination. Notably, while protection against cross-reactive strains is generally desirable, activation of a memory response can result in the production of weakly neutralizing antibodies which provide little protection, but block the formation of novel, stronger antibody

species by steric hindrance (the Zarnitsyna paper calls this mechanism "epitope masking"). Assessing the degree to which the immune response to heterologous strains follows this pattern would provide information about the breadth of the influenza immune response. Furthermore, based on [AMANDA'S WORK ON AG DISTANCE], quantifying how the effect of preexisting immunity varies as antigenic distance between the vaccine strain and the assay strain decreases (and therefore, cross-protection decreases) would provide a flexible modeling approach that has the potential to be less dependent on the strains selected for testing.

If antigenic distance were the only feature of a given viral strain important in understanding whether a vaccine will provide protection (this may not be the case), models trained on strains with a wide range of antigenic distances could produce predictions for any strain based on the antigenic distance with the vaccine strain, not just strains used in training.

1.2 Exploratory data analysis

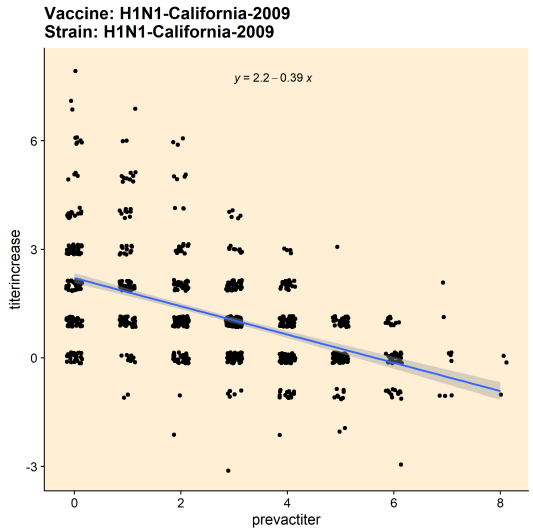
Initially, we plotted the fold change in antibody titer vs. the pre-vaccination antibody titer for each vaccine strain / test strain combination represented in the data. Each of these plots has one data point per person-year, and does not account for repeated measurements of the same person, but provides a rough equivalent of the plots shown in [ZARNITSYNA PAPER]. Figure 1 shows one of these linear models. In the homologous case (left) we see a strong negative correlation, matching the predictions made by the mechanistic model. However, for distant strains (right) we observe much weaker correlations. The plots shown only show examples for two strains.

We then computed all of the linear models of this form, obtaining the slope for each vaccine/test strain combination. This slope coefficient represents the expected difference in fold change for a 1 log-unit increase in prevaccination antibody titer – so a negative coefficient matches the predictions of the mechanistic model. When we plot these slopes against a crude measure of antigenic distance, as shown in Figure 2, (the difference in isolation years of the two strains), we see a general positive trend, indicating that **as strains become more antigenically distant, the effect of pre-existing titer diminishes**, matching our hypothesis.

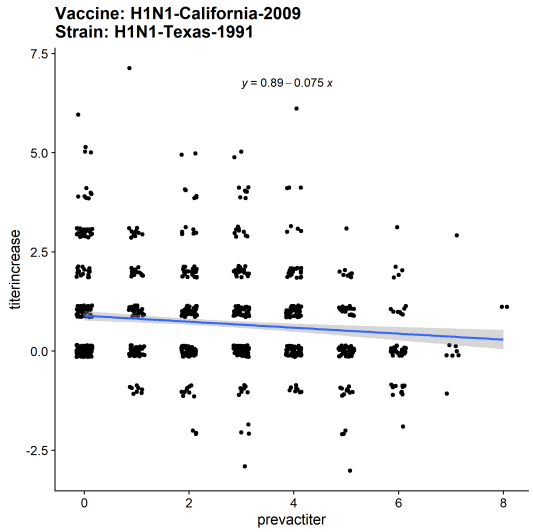
Furthermore, if we stratify by the subtype of influenza in this plot (Figure 3), we see different relationships for each subtype/lineage. As expected, protection from A(H3N2) influenza diminishes the quickest, which tracks with A(H3N2) strains evolving faster than the other subtypes shown.

1.3 Analytical plan

- Incorporate several measures of antigenic distance calculated by Amanda.
- Fit multilevel Bayesian models with the ability to control for potential confounds, to isolate the effect of antigenic distance.
- Sensitivity analysis to determine the impact of measured strains, if possible.



(a) Homologous response: the test strain is the same as the vaccine strain.



(b) Heterologous response: the test strain used in the HAI assay is not the same as the strain used in the vaccine.

Figure 1: Fold change ('titerincrease') vs. log pre-vaccination HAI titer ('prevactiter') for two test strains.

- Post-hoc analysis of differences between antigenic distance metrics, given that all other measurements remain the same.
- This could potentially include modifications to Veronika's mechanistic model to incorporate heterologous responses (or more likely a separate aim).

1.4 Feasibility assessment

Amanda has already computed the antigenic distance metrics for Ted's cohort data. If we intend to incorporate DIVERsity data, that will entail a substantial amount of new data cleaning, but this would be common across all dissertation aims. Furthermore, if other studies use different virus panels for assessing HAI data, the antigenic distance metrics for those would need to be computed.

If we want to incorporate any data other than HAI responses, there would be a significant learning curve for me. However, I feel confident in my ability to fit models to answer this question using Ted's cohort data with Amanda's calculated antigenic distance measurements. The last remaining question here would be the structure of the model, including what other variables should be included, and the parametric form for the inclusion of antigenic distance. If only using Ted's HAI data would be satisfactory (as an extension or alternative to Yang's model where strains are treated as categorical variables), I think this aim is completely feasible.

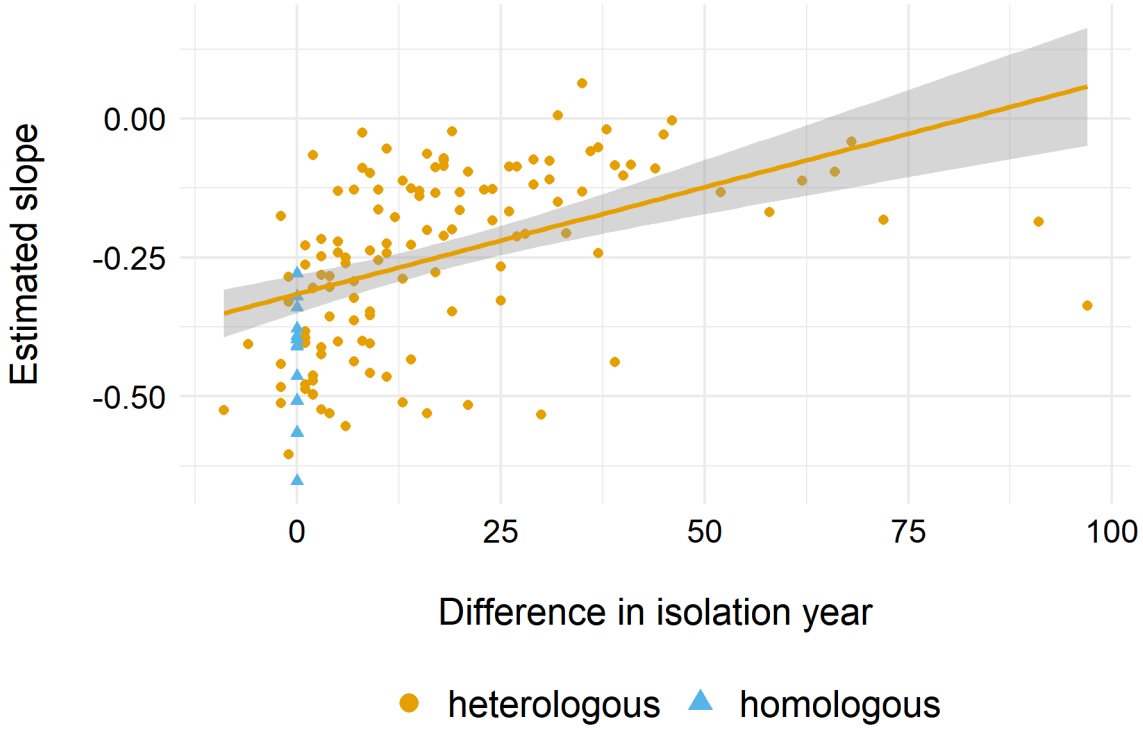


Figure 2: Estimated slopes for each combination of test strain and vaccine strain present in the data. Overall, as the calendar time between the emergence of the two strains increases, the estimate slope increases. Therefore, preexisting immunity has less of an impact on vaccine boosting for strains which are far from the vaccine in calendar time. Calendar time is only a rough approximation of antigenic similarity, but we expect better measurements to show the same general pattern.

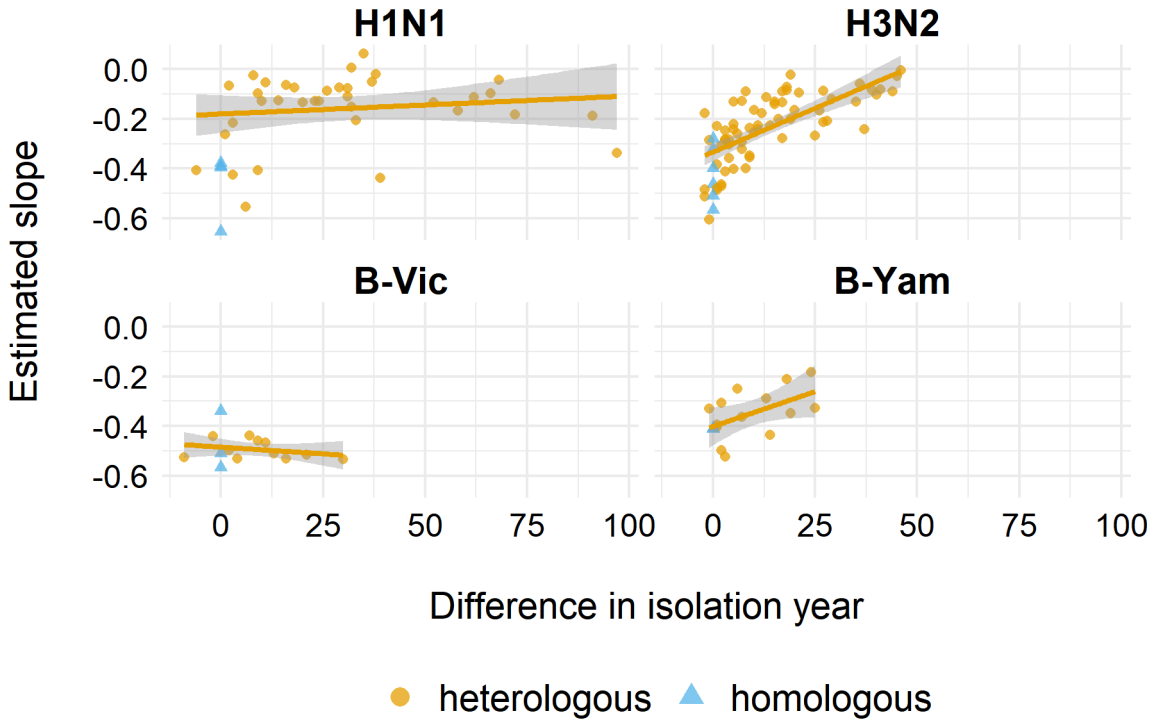


Figure 3: The estimate slopes vs. difference in isolation year, but this time stratified by subtype (for influenza A strains) or lineage (for influenza B strains). Influenza A(H3N2) and B (Yamagata-like lineage) show the strongest trends. The A(H1N1) subtype shows only a moderate trend, with several outliers, which could potentially be due to the inclusion of both 2009 pandemic-like strains and pre-pandemic strains. Interestingly, B (Yamag) lineage shows the opposite trend, although the effect is weak. Notably, divergences in the B strains are quite recent and it is thought that cross-reactivity is significantly higher among influenza B strains due to the comparatively slower rate of antigenic drift.

My estimated time scale for a first draft (working only on this project) would be between 1 – 3 months depending on any issues with combining the antigenic distance data from Amanda with Ted's cohort data, and resolving any remaining data cleaning issues with the data.

2 Determine how demographic information, including indicators of high influenza risk, modulate response to the seasonal influenza vaccine.

3 Quantify differences between vaccine doses and types using compiled data from the DIVERsity grant.

- Would consider age, sex, race (if possible), comorbidities (if possible), and prior vaccination history.

4 Determine whether boosting and waning patterns in longitudinal immune responses are associated with pre-existing immunity and host factors.

5 Quantify the direct effect of immunosenesence on waning immunity.

- Mediation analysis where birth year -> age -> immune response? Not quite sure about this one, probably not feasible.

6 Determine if viral characteristics other than antigenic distance are important for predicting/modeling immune response. E.g. Justin Bahl's T_{eff} sequence data.

7 References