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

**Authors**

* W. Zane Billings (ORCID: 0000-0002-0184-6134);
* Amanda L. Skarlupka (0000-0002-3654-9076);
* Murphy H. John (0009-0009-9591-1117);
* Savannah L. Miller (0000-0003-2231-3510);
* Hayley Hemme (0009-0002-6609-1390);
* Ted M. Ross (0000-0003-1947-7469);
* Andreas Handel (0000-0002-4622-1146)

**Author affiliations**

1. College of Public Health, The University of Georgia, Athens, GA, USA.
2. Center for the Ecology of Infectious Diseases, The University of Georgia, Athens, GA, USA.
3. Vivli, Burlington, MA, USA.
4. Center for Vaccines and Immunology, The University of Georgia, Athens, GA, USA.
5. Florida Research & Innovation Center, Cleveland Clinic, Port St. Lucie, FL, USA.

Corresponding author: WZB (wesley.billings@uga.edu). Alternate corresponding author: AH (ahandel@uga.edu)

**Conflicts of interest**: None of the authors declare any conflicts of interest.

**Funding sources**:

* WZB: none declared
* ALS: none declared
* MHJ: none declared
* SLM: none declared
* HH: none declared
* TMR: funded by the Georgia Research Alliance as an Eminent Scholar
* AH: received partial support from NIH grants/contracts U01AI150747, R01AI170116, and 75N93019C00052.

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Author contributions**:

* WZB: conceptualization, methodology, software, formal analysis, data curation, writing - original draft, writing - review and editing, visualization
* ALS: conceptualization, methodology, software, formal analysis, writing - review and editing
* MJ: data curation, validation, writing - review and editing
* SLM: validation, writing - review and editing
* HH: validation, writing - review and editing
* TMR: investigation, resources, data curation, writing - review and editing
* AH: conceptualization, methodology, writing - review and editing supervision, funding acquisition

**Notes for feedback**

* target journal: We want to submit to PNAS or Proc Soc B, or another general science journal with wide readership. So the manuscript needs to have as few specific technical details as possible. (It probably has a lot of technical details right now because I usually write in a boring and technical way, so I appreciate any suggestions.)
* given the subsampling results, I think we should only talk about AUC in this manuscript. I also think that would make it less confusing. Does anyone object to only talking about the AUC? We can compare it to the seroconversion rate and GMT, and then we don’t have to worry about setting up the magnitude/breadth/total strength stuff with no citations.
* Planned future results: plot how the metrics change over time.
* Planned future result: compare the CA/09 and MI/15 vaccines in UGA only as a case study for using the metrics.
* Planned future result: how does the number of strains and similarity of strains we include in our panels affect the ICC? This can tell us about the “minimum antigenic diversity” needed to get robust answers.
* Right now the ICCs are frequentist with approximate CIs. I didn’t change the stuff in the methods because the bayesian ones are running now, but unfortunately they take longer than the actual models.
* **if you want to be an author on the actual manuscript, please send me your preferred name, affiliation, orcid, and funding information along with your comments.**

# Abstract

**Background:** Developing broadly-protective vaccines is a clear path towards reducing the burden of influenza pandemics and seasonal epidemics. However, measuring the breadth of response of a vaccine or candidate entails estimating the immunogenicity of the vaccine against many different influenza strains, and then combining these measurements. The most common method is to count the proportion of strains to which a vaccine induces a clinically noticeable immune response, which is strongly dependent on the panel of strains used for analysis. Estimates of breadth using this method cannot fairly be compared across different studies using different virus panels. This prevents proper comparison between vaccine candidates. In our study, we develop a novel robust metric for quantifying vaccine breadth across labs and virus panels.

**Methods:** We used data from a prospective, open annual influenza vaccination cohort across three study sites from Fall 2013/14 through Spring 2017/18. Each individual contributed pre-vaccination and post-vaccination serum samples which were used for a panel of hemagglutination inhibition (HAI) assays against many strains of influenza, including the strains used in each season’s vaccine. We computed multiple antigenic distance metrics between each HAI assay strain and the corresponding seasonal vaccine strain

**Results:** First, we calculated antigenic distance metrics using three different methods. We then generated summary antibody landscapes for vaccine response vs. antigenic distance. From these summary landscapes, we computed our novel metrics for vaccine strength and breadth. For comparison, we also computed currently used breadth metrics. We subsampled multiple panels of heterologous strains and sets of individuals from our data to create hypothetical studies, and we computed metrics across these subsampled studies. Finally, we calculated the intraclass correlation (ICC) to measure how much each metric varied across labs with different populations and virus panels.

**Conclusions:** Our novel metrics are much more robust to the panel of viruses chosen by a given lab than currently used metrics for assessing vaccine breadth. Implementing our methods is easy and does not require imputing missing strains from vaccine panels or costly logistics for coordinating which strains are used in vaccine breadth panels, although a combination of approaches would be ideal. Our newly proposed methods allow for fairer comparison of vaccine candidates by robustly accounting for differences in heterologous vaccine panels used to assess breadth.

**Keywords:** influenza, vaccines, heterologous immunity, vaccine breadth, broadly-reactive vaccines, universal influenza vaccine, cohort study, secondary data analysis

# Introduction

Developing a broadly-protective (or “universal”) influenza vaccine is a key goal for reducing the burden of seasonal influenza and increasing pandemic preparedness (1–4). As a given influenza strain spreads, many hosts develop an immune response to that strain, placing selective pressure on the virus. Mutant lineages can acquire antigenic changes that are not recognized by hosts who have previously received vaccines, allowing for vaccine escape. The rapid evolution and vaccine escape patterns of influenza make developing a universal vaccine difficult (5–8). Many new vaccine candidates are under development (9), but the best method for assessing the breadth of a vaccine candidate (how broadly immunogenic a vaccine candidate is) is unclear.

In previous research, the breadth of the response is quantified by measuring vaccine-induced immunogenicity against a panel of historical influenza strains. Among multiple vaccine candidates, the one that induces a clinically meaningful immune response to the highest number of strains is considered the most broadly reactive (10–15). This method is easy to conduct, but the selection of different virus panels across research groups makes the results from different labs hard to compare. Some methodological work has focused on methods for imputing data across different virus panels (16,17), but there is no practical exploration of how well these imputation measures work, and they rely on a low-rank approximation which may be unreliable (18). Requiring many different labs to use exactly the same panel of viral strains (including constant monitoring for adaptations, and ensuring all protocols are the same) is logistically not feasible and would be extremely expensive. Breadth metrics that are robust to the selection of different panels of viral strains would circumvent these logistical issues and allow for fair comparisons of broadly-reactive vaccine candidates across lab groups.

Recent methods have focused on quantitative analysis of individual antibody landscapes, rather than counting the number of seroprotection events across an immune assay panel. HAI is the most common assay used for these panels, but inherent biases of the HAI assay can influence breadth measurements (19–22), so a method for quantifying vaccine breadth should be agnostic to the specific measurement used. Other proposed immunological assays for breadth calculation include stem-binding neutralization assays (23), binding profiles (24), neuraminidase inhibition (25), and a multiplex neutralization assay (26). Regardless of the assay one uses, an antibody landscape is a curve which depicts antibody response on the -axis with respect to some measurement on the -axis which attempts to order the responses by the differences between influenza strains (distance metrics). Just as there is no gold standard for the best immunological assay to use, there is no gold standard for the distance metric, and different metrics can measure different ways for two strains to be different. Previously proposed distance metrics include the year of strain isolation (27,28), genetic differences (29–33), biochemical or biophysical differences (34,35); and distances derived from antigenic cartography (12,36–40). Previous work suggests notable differences between different antigenic distance metrics (41), but how these differences affect breadth quantification is unclear.

In our study, we use of antibody landscapes from influenza vaccine cohort studies as the basis for measuring vaccine breadth. Our metrics are agnostic to the immunological assay and distance metrics used, and are robust to the use of different virus panels, compared to currently-used metrics. We develop methods for creating population-summary antibody landscapes, and metrics derived from the summary antibody landscape which are more robust to differences in virus panels than previously used methods. We also include a case study showing how our methods can be used to assess breadth of response between two different vaccine candidates. Overall, our study contributes to the definition of breadth and the reproducibility of selecting broadly-reactive vaccine candidates by improving the ability to compare results across labs.

# Methods

## Data source

For this study, we performed a secondary data analysis of a subcohort from a prospective, open design ongoing vaccine cohort study. The cohort study has been described previously (42,43), but in brief, individuals were recruited one of two study sites (Port St. Lucie, FL, or Pittsburgh, PA) during the influenza season, donated a pre-vaccination blood draw, received a Fluzone (Sanofi) seasonal influenza vaccine, and returned at a follow-up visit to donate a post-vaccination blood draw at approximately 28 days after vaccination. Each serum sample was used for a panel of hemagglutination inhibition (HAI) assays to the strains in the vaccine, as well as for a panel of historical strains. We considered only the influenza A(H1N1) responses to the standard dose Fluzone vaccine in our study as a proof of concept, but our methods apply equally to any panel of heterologous assays. We calculated descriptive statistics for our cohort, including counts per season and demographic summaries.

We calculated the antigenic distance between the vaccine strain (the A(H1N1) strain used in the vaccine formulation in a given season) and the assay strain (the actual virus strain added to the serum during the HAI assay) for every HAI assay using three different metrics. We calculated the temporal antigenic distance, which is the difference in isolation year between the vaccine strain and the assay strain; the dominant -epitope distance (31), which is the maximum proportion of amino acid differences across the five immunodominant epitope sites on the hemagglutinin head; and the cartographic distance (30). Antigenic cartography employs statistical dimension reduction on a matrix of serological titer data before taking Euclidean distances between strains. We created a two-dimensional cartographic map with Racmacs (44) using pre-immune human sera from all assays in our study sample, and used the lowest stress map from 100 random initializations with 100 L-BFGS optimization rounds each. To fairly compare across the three antigenic distance metrics, which have different units, we min-max normalized all of the antigenic distance values within a given season.

## Vaccine strength estimation framework

Next, we developed a framework for estimating vaccine performance including both the magnitude of the response (response to the homologous vaccine strain) and breadth of response (summary of responses to heterologous strains). Our framework also allows for estimation of the total strength of a vaccine, which combines the magnitude and breadth into a one-number summary of the response induced by vaccination. We developed one metric for each of the magnitude, breadth, and total strength of the response and compared these to the metrics currently used in the literature.

Metrics in the current literature (see Introduction) rely on calculating the geometric mean titer (GMT) or seroconversion rate (SCR) across a panel of historical strains. In this “current” framework, the estimate of the magnitude of the response is the GMT to the homologous strain; the estimate of the breadth of response is the seroconversion rate across all strains in the panel; and the estimate of the total strength of the response is the GMT across all strains. See the Supplement for mathematical details on these quantities.

Our novel robust metrics are derived from a sample summary antibody landscape. That is, rather than consider all of the individual antibody landscapes in our study (there is one per person-year), we fit a model that estimates the summary antibody landscape for our study sample. Any statistical model that can estimate vaccine outcomes as a function of antigenic distance could be used here, but for simplicity we used a linear regression model. In order to fairly weight the contributions of each individual antibody landscape, we modeled the post-vaccination titer as a Gaussian outcome with a global intercept, a fixed effect for antigenic distance, and correlated varying intercepts and antigenic distance effects for each individual. For simplicity, we treated all person-years from the same individual as independent data. The Supplement contains model formulas and details on our regression model implementation.

From the population summary antibody landscape, we can derive estimates of the magnitude, breadth, and total strength of the vaccine response. We estimate the magnitude as the intercept of the linear regression line (or in general, the predicted response at an antigenic distance of zero); the breadth as the antigenic distance value where the summary antibody landscape intercepts the value 1:40 (the HAI titer accepted as a marker for seroconversion and seroprotection); and the total strength as the area under the curve (AUC) of the regression line between antigenic distances of 0 and 1.

## Model implementation and censoring correction

We implemented our summary antibody landscape models in a bayesian framework, which allows us to easily obtain uncertainty estimates for each metric as credible intervals. In order to fairly compare the current metrics with our novel metrics, we also developed models for calculating the current metrics in a bayesian framework to obtain comparable CIs (45–47). We estimated the homologous GMT using a Gaussian regression model for the post-vaccination titer with an intercept only, fit only to titers from the homologous strain — in this model, the estimate of the intercept estimates the post-vaccination GMT. Similarly, we estimated the total strength GMT with a Gaussian regression model for the post-vaccination titer with an intercept only, using data from all assay strains. Finally, we estimated the seroconversion rate using a logistic regression model with an intercept only, where the outcome was seroconversion to a given strain. We defined seroconversion as a post-vaccination titer of at least 1:40 with a 4-fold change or greater between pre- and post-vaccination titers.

Because we estimated all of our metrics using regression models, we can easily apply a correction to the likelihood to adjust for the interval censoring and limit of detection issues inherent to HAI titers (48). The assays in our dataset had a lower limit of detection of 1:10 and an upper limit of detection of 1:20480. We estimated all six of the metrics (the current metrics and our novel metrics) using models with censoring correction and without a censoring correction, since censoring is known to artificially deflate the amount of uncertainty. In order to fairly compare metrics across each study, we min-max normalized each of the metrics within each season. Since all of the metrics have different units, normalization allows us to compare the relative spread of each metric fairly.

We estimated these metrics on the entire dataset for each season to determine if vaccine strength varied seasonally after accounting for the breadth of response, and we calculated the metrics with and without a censoring correction to gain a baseline understanding of how much censoring affects the metrics.

## Robustness analysis

Next, in order to show that our novel metrics are more robust to comparisons across different historical strain panels than currently used metrics, we implemented a subsampling analysis. We created 25 subsampled “studies” which are intended to simulate 25 labs analyzing the same vaccine candidate using different virus panels. For each of the subsampled studies, we randomly sampled 9 heterologous strains from the panel of either 15 or 16 for a given season, and added the homologous strain for that season to get a panel of 10 viruses for that study. We also randomly sampled 100 individuals to “participate” in a given study. We calculated all six metrics on each subsampled study, both with and without a censoring correction.

Once we calculated the metrics for each subsampled study, we used the intraclass correlation (ICC) as a measure of consistency across the subsamples for a given metric. The ICC measures the proportion of variance in a statistic that can be attributed to a specific grouping factor, and ranges from 0 to 1. An ICC close to 0 indicates that the differences between studies contribute very little to the variation in estimated metrics, while an ICC close to 1 indicates that the differences between studies are the major source of variation in estimated metrics. We computed the ICC for each of the six metrics, both with and without a censoring correction.

We performed this robustness analysis for each seasonal cohort in our study sample to determine if the ICC for each metric changed seasonally. We calculated each ICC using a bayesian random effects model which contained only a global intercept and a random intercept for each subsampled study. The outcome in each model was the metric of interest, and we used a Gaussian likelihood distribution. See the Supplement for detailed information on all of our models.

## Implementation

We implemented all of our results in a bayesian framework for two main reasons. First, estimating the vaccine response metrics in a bayesian framework allows us to calculate CIs for each metric without resorting to bootstrapping or other solutions for frequentist confidence intervals that can be unreliable. Second, our study is a secondary data analysis, so any frequentist confidence intervals and p-values we calculate would suffer from inflated false positive rates. Thus, we do not report any p-values in our results.

Our analysis was conducted with R version 4.4.2 (2024-10-31) using RStudio version 2024.12.1.563 (49). Our project was developed as a targets pipeline (50) using renv (51) and here (52) to enhance reproducibility. We deployed our pipeline using the crew package on the University of Georgia’s Sapelo2 computing cluster, which uses the Slurm scheduling software (53,54). We implemented our bayesian models with the brms package (55–57) using the cmdstanr backend (58) and cmdstan version 2.36.0 (59) as the interface to the Stan probabilistic programming language (60,61). We additionally used the tidyverse package suite (62) and the packages qs2 (63), tidybayes (64), and pracma (65) for formal analysis. We created our manuscript and supplementary material using Quarto version 1.6.40 (66) with the packages knitr (67–69) and softbib (70). We made our figures using ggplot2 (71), patchwork (72), and ggdist (73,74) and our tables using flextable (75). The Supplement contains more details on our methodology, including instructions for running our code. We archived our dataset and analysis code on Zenodo (LINK HERE) and GitHub (LINK HERE).

# Results

## Data description and antibody landscape

Our study sample consisted of several thousand pairs of HAI assays ([Table 1](#tbl-counts)), with at least 1000 in each study season. These assay pairs represent homologous and heterologous HAI assay pairs (one pair includes one pre-vaccination and one post-vaccination titer). Every individual enrolling in a given season contributed the same number of HAI assays, and each study site used the same historical panel in a given season ([Table 2](#tbl-people)). Overall, we included person-years of data, representing pairs of HAI assays. While some individuals re-enrolled in the study in successive seasons, at the same study site each time, we treated all person-years as independent for the purpose of our analysis. Thus, is the total number of enrolled person-years, but the total number of enrolled unique individuals was ( from the FL study site, from PA, and from UGA). Summary plots of the sample titer distributions, along with more information on the heterologous strains used each season, are shown in the Supplement.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 1: Count of HAI assay pairs included in our study from each study site during each season. Numbers shown are count (cell percentage). The only season in which all three study sites operated was 2016/2017, and a dash indicates that the study site for that row was not operational during the season indicated by the column.   | Study Site | Season | | | | | | | --- | --- | --- | --- | --- | --- | --- | | 2013/14 | 2014/15 | 2015/16 | 2016/17 | 2017/18 | Overall | | PA | 638 (4%) | 1290 (7%) | 1215 (7%) | 1024 (6%) | — | 4167 (24%) | | FL | 960 (5%) | 2250 (13%) | 1920 (11%) | 1904 (11%) | — | 7034 (40%) | | UGA | — | — | — | 2320 (13%) | 4080 (23%) | 6400 (36%) | | Overall | 1598 (9%) | 3540 (20%) | 3135 (18%) | 5248 (30%) | 4080 (23%) | 17601 (100%) | |
| Table 2: Count of individuals enrolled in our study from each study site during each season. Numbers shown are count (cell percentage). The last row of the table shows the number of assay pairs contributed by each individual who enrolled in the study and completed both study visits. Each individual who enrolled in a particular season contributed the same amount of HAI titers or was excluded from our analysis.   | Study Site | Season | | | | | | | --- | --- | --- | --- | --- | --- | --- | | 2013/14 | 2014/15 | 2015/16 | 2016/17 | 2017/18 | Overall | | PA | 638 (4%) | 1290 (7%) | 1215 (7%) | 1024 (6%) | — | 4167 (24%) | | FL | 960 (5%) | 2250 (13%) | 1920 (11%) | 1904 (11%) | — | 7034 (40%) | | UGA | — | — | — | 2320 (13%) | 4080 (23%) | 6400 (36%) | | Overall | 1598 (9%) | 3540 (20%) | 3135 (18%) | 5248 (30%) | 4080 (23%) | 17601 (100%) | |

[Figure 1](#fig-landscape) shows our summary antibody landscape for the 2016-2017 influenza season, pooling data across the three studies. We show the antibody landscapes for the three different distance metrics, both with and without the censoring correction. Overall, the censoring correction slightly increased the intercept and made the slope more negative (making the line slightly steeper) for all three models, but the choice of antigenic distance metric had a more noticeable impact on the landscapes. The summary landscapes using the cartographic distance were steeper and had a higher intercept than the p-Epitope landscapes, which in turn had steeper slopes and higher intercepts than the temporal distance landscapes.

The differences between the antigenic distance metrics are also noticeable in a qualitatively different ordering and relative placement of certain strains. Using the temporal distance, the CA/09 strain is more similar to the USSR/77 strain than to the SC/18, which is misleading. The CA/09 strain represents the pandemic-like H1N1 lineage which is more genetically and antigenic similar to the SC/18 strain, as reflected in the p-Epitope and cartographic distance orderings respectively. Notably, the p-Epitope distance placed the three distances roughly equidistant while the cartographic distance estimates the CA/09 strain as being more different from either SC/18 or USSR/77 than those strains are to each other, which intuitively reflects the unique deletions that differentiate the CA/09 clade from other H1N1 variants.

The differences between the metrics and the corrected (for censoring) vs. uncorrected landscapes is apparent in the vaccine immunogenicity metrics for this subcohort ([Table 3](#tbl-metrics)). After correcting for censoring, all of the metrics increased (the current metric for breadth, seroconversion rate, which is based on a binary outcome and therefore cannot be corrected for censoring, so it did not change). While many of the corrected and uncorrected metrics had overlapping credible intervals, the intercepts in particular (novel metrics for magnitude) showed a notable increase for all three antigenic distance metrics. The difference between censoring methods shows how failing to properly account for censoring can bias vaccine results towards the null, potentially making vaccine candidates look worse than they are. Focusing only on the metrics with censoring corrections going forward, the novel metrics also varied greatly across the three antigenic distance metrics.

In particular, using the temporal method (the most common method for antibody landscape analysis in the literature) underestimated all three of the metrics, because the temporal method is insufficient for capturing antigenic and genetic differences between virus strains. The p-Epitope metric yielded more optimistic results for breadth and total strength than the cartographic method, but a smaller estimate of the magnitude. From these data alone, we cannot tell which metric is more appropriate, but based on the summary landscape ([Figure 1](#fig-landscape)), the distances covered by our historical panel were different for cartographic and p-Epitope distances. While the cartographic distance had a fairly even coverage of the antigenic distance metric, there was a gap where we had no viruses with a normalized cartographic distance around 0.25. The p-Epitope metric, however, is discrete and lumps many more strains close together. Having more strains with normalized antigenic distances near 1 (which have high leverage in a regression model) may bias the intercept downward, and correspondingly affect the conditional estimate of the slope. In short, we need a panel which evenly covers antigenic space for both cartographic and p-Epitope distances before we can determine which metric produces the “correct” vaccine immunogenicity metrics.

We also noted several differences between the current set of metrics and our novel metrics. While the temporal metric underestimated immunogenicity for all three metrics relative to the current metric set, the p-Epitope metric produced a smaller estimate of the magnitude and higher estimate of the breadth and total strength. Since the novel magnitude estimate of the cartographic distance was similar to the magnitude estimate from the current metric, comparisons between current and novel metrics using the cartographic distance may be the most useful for calibrating our understanding. Both the cartographic and p-Epitope metrics produced more optimistic metrics of breadth and total strength than the current metrics or novel metrics using temporal distance, which likely indicates that using a “real” metric of genetic or antigenic distance allows for less biased estimates of vaccine breadth.

While we show only the results for the 2016/17 season here, the other seasons had similar trends and we include summary antibody landscapes and metric calculations for the other seasons in the Supplement.

|  |
| --- |
| Figure 1: 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. The strain labels show the relative positions of three different strains (CA/09, the homologous strain; SC/18, the 1918 pandemic-like strain; and USSR/77, the most different strain from the vaccine strain in the genetic comparison. Each column of plots shows a different antigenic distance metric (left: cartographic distance; center: p-Epitope sequence distance; right: temporal year-based distance). The two rows show landscapes fitted without a censoring correction (top row) where values lower than the LoD were set to 5, and the bottom row shows landscapes with the likelihood-based censoring correction. |
| Table 3: Current and novel vaccine immunogenicity metrics for the 2016 - 2017 season data, 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. All of the ‘Novel’ metric sets, each computed using a different metric for antigenic distance, are based on the summary antibody landscape in [Figure 1](#fig-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.   | Metric Set | Censoring correction | | | LoD set to 5 | | | | --- | --- | --- | --- | --- | --- | --- | | Magnitude | Breadth | Total Strength | Magnitude | Breadth | Total Strength | | 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) | 5.23 (5.06, 5.43) | 0.50 (0.48, 0.52) | 3.00 (2.87, 3.11) | 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) | |

## Subsampling study results

To show that our novel metrics are more robust to the choice of virus panel than the current metrics for vaccine breadth, we created 25 subsampled labs from each seasonal subcohort in our study sample. Each lab had a virus panel consisting of the homologous vaccine strain for the season and 9 heterologous strains randomly sampled without replacement from the set of strains used in that season. [Figure 2](#fig-subsamples) shows the posterior distribution of each metric across the different subsamples from the 2016/17 subcohort, after correcting for censoring. The results were clearly different between the current metrics, novel metrics using cartographic distance, and novel metrics using p-Epitope distance; as well as across the metrics for magnitude, breadth, and total strength of the vaccine. For magnitude metrics, the current and novel (Cartographic) metrics distributions looked qualitatively similar, although the p-Epitope estimates were notably lower (as we observed in [Figure 1](#fig-landscape)) and there is considerably more between-group variation.

The breadth metric results were striking: the estimate of breadth using the current metric (Seroconversion rate) was very consistent around across all subsamples – this appears to be due to a low overall rate of seroconversion in our dataset. Our novel metric showed a higher overall estimate for both cartographic and p-Epitope antigenic distances, although the results were much more consistent across subsamples for the cartographic distance than the p-epitope distance. Finally, the total strength metrics followed the same trend that we observed in the full cohort ([Table 3](#tbl-metrics)), but the estimates of the novel metrics using the cartographic dataset were the most consistent across subsamples. These qualitative results are supported by the intraclass correlation (ICC) that we calculated for each of the 9 immunogenicity metrics ([Table 4](#tbl-iccs)).

The current metric set had the lowest ICC among the three groups we compared for the magnitude estimate, but the highest ICC for the breadth and total strength estimates, indicating variability in the current metrics for breath and total strength was dominated by between-subsample variation. The novel metric set using the p-epitope antigenic distance had high ICCs for all three metrics. Finally, the novel metrics using the cartographic antigenic distance had a comparable ICC to the current metric set for the breadth metrics, and the lowest ICC for breadth and total strength. In particular, the ICC for total strength using the cartographic distance was much lower than the current metrics, indicating that substantially less variability in our novel metrics, when the cartographic distance is used for estimation, is due to variation across subsamples. The results for the other seasons were similar and are shown in the Supplement.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 4: 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.   | Metric Set | Magnitude | Breadth | Total Strength | | --- | --- | --- | --- | | Current | 0.55 (0.36, 0.65) | 0.79 (0.68, 0.85) | 0.88 (0.82, 0.93) | | Novel (Cartographic) | 0.59 (0.43, 0.70) | 0.53 (0.38, 0.67) | 0.54 (0.39, 0.66) | | Novel (p-Epitope) | 0.79 (0.69, 0.85) | 0.75 (0.62, 0.82) | 0.78 (0.66, 0.86) | |
| Figure 2: Estimated immunogenicity metrics for each simulated lab drawn from the 2016/17 subcohort data. We computed 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. |

# Discussion

Our novel methods for evaluating broadly-reactive vaccine candidates are more robust to differences in virus panels than currently used metrics. In addition, our incorporation of antigenic distance allows our metrics to be used for further analyses which provide more information about the vaccine-induced immune response than metrics based only on geometric mean titers and seroconversion rates. We found that current methods for evaluating breadth and the total strength of the vaccine response (incorporating breadth) underestimate the true breadth of vaccine response. Methods based on the seroconversion rate across strains and methods that use year of isolation (or difference between years of isolation) as a proxy for antigenic distance are the most common methods used for analyzing breadth in the literature, but neither works well.

Our metrics appear to be more robust and informative than the most common methods in current use, but rely strongly on the choice of an antigenic distance metric. We previously found that many antigenic distance metrics can be used to make similar predictions about vaccine response (CITE AIM 2PREPRINT WHEN WE GET IT ONLINE), which seem to contradict these results. However, the metrics we calculate here do not need to account for many sources of intraindivdual variability and confounders, and uncertainty is only induced by the uncertainty in estimating model parameters from the data, so we have a much higher resolution with which to see differences in antigenic distance metrics. The major issue with choosing the correct antigenic distance metric is that there is no ontologically correct choice – different antigenic distance metrics serve different purposes. Our analysis on only A(H1N1) strains suggests that cartographic distance works well for estimating our metrics, likely because it is continuous (rather than discrete with a limited number of possible values like p-epitope), and because the historical panel of viruses used in our study had good coverage in cartographic distance space. However, there are many issues with HAI assays (19) and results might vary across cohorts or with A(H3N2) viruses where HAI assays can be more troublesome. Since cartographic distance is widely accepted, we feel that recommending the use of cartographic distance for calculating our robust metrics is not overly hasty, although there is to date no major comparison of cartographic distances computed from ferret and human sera, or differences in cartographic maps across human cohorts.

We also do not assess the major alternative strategy for solving the problem of differences in virus panels. Methods for imputing systematically missing strain data across different panels also seem promising (16,17), although we have not attempted to evaluate them in our subsampling context. Our robust metrics can be used alongside imputation methods, in situations where imputating titers across strain panels is reliable. A future multi-center study using the same influenza strains, vaccine candidates, and protocols (preferably with a high-throughput multiplex assay) at all study sites is necessary to truly evaluate the robustness of any procedure for measuring response breadth. Such a future study should also plan to collect sera, store serum samples long term, and run HAI assays for “future” strains that will evolve after the serum samples are collected. Since we cannot feasibly run protection or efficacy trials to assess vaccine breadth, analyzing which vaccines induce a strong response against these “future” strains will be the best surrogate to fairly assess any metric that estimates vaccine breadth.

In short, our novel metrics can be used to quantify breadth of vaccine response, and compare broadly reactive vaccine candidates more robustly across labs using different virus panels. Understanding the best antigenic distance method to use for quantification is difficult, but using cartographic distance still provides researchers with much more information than current methods, despite the limitations of cartography. We hope that in the future our strategy for analyzing antibody landscapes to understand breadth of vaccine response can be combined with other promising methods and study designs to further the development of a universal influenza vaccine.

# Acknowledgments

We thank Michael Carlock (Cleveland Clinic, Port St. Lucie, FL, USA) for assistance with obtaining data. In addition, we thank the developer of targets, William Michael Landau (Eli Lilly and Company, Indianapolis, IN, USA) for assistance with developing our computational pipeline.

This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia’s Office of the Vice President for Research and Office of the Vice President for Information Technology.

# References

1. Kim H, Webster RG, Webby RJ. [Influenza Virus: Dealing with a Drifting and Shifting Pathogen](https://doi.org/10.1089/vim.2017.0141). Viral Immunology. 2018 Mar;31(2):174–83.

2. Erbelding EJ, Post DJ, Stemmy EJ, Roberts PC, Augustine AD, Ferguson S, et al. [A Universal Influenza Vaccine: The Strategic Plan for the National Institute of Allergy and Infectious Diseases](https://doi.org/10.1093/infdis/jiy103). The Journal of Infectious Diseases. 2018 Jul;218(3):347–54.

3. Wei CJ, Crank MC, Shiver J, Graham BS, Mascola JR, Nabel GJ. [Next-generation influenza vaccines: Opportunities and challenges](https://doi.org/10.1038/s41573-019-0056-x). Nature Reviews Drug Discovery. 2020 Apr;19(4):239–52.

4. Vogel OA, Manicassamy B. [Broadly Protective Strategies Against Influenza Viruses: Universal Vaccines and Therapeutics](https://doi.org/10.3389/fmicb.2020.00135). Frontiers in Microbiology. 2020;11:135.

5. Krammer F, Smith GJD, Fouchier RAM, Peiris M, Kedzierska K, Doherty PC, et al. [Influenza](https://doi.org/10.1038/s41572-018-0002-y). Nature Reviews Disease Primers. 2018 Jun;4(1):1–21.

6. Koelle K, Cobey S, Grenfell B, Pascual M. [Epochal Evolution Shapes the Phylodynamics of Interpandemic Influenza A (H3N2) in Humans](https://doi.org/10.1126/science.1132745). Science. 2006 Dec;314(5807):1898–903.

7. Zinder D, Bedford T, Gupta S, Pascual M. [The Roles of Competition and Mutation in Shaping Antigenic and Genetic Diversity in Influenza](https://doi.org/10.1371/journal.ppat.1003104). PLOS Pathogens. 2013 Jan;9(1):e1003104.

8. Oidtman RJ, Arevalo P, Bi Q, McGough L, Russo CJ, Cruz DV, et al. [Influenza immune escape under heterogeneous host immune histories](https://doi.org/10.1016/j.tim.2021.05.009). Trends in microbiology. 2021 Dec;29(12):1072–82.

9. Taaffe J, Ostrowsky JT, Mott J, Goldin S, Friede M, Gsell P, et al. [Advancing influenza vaccines: A review of next-generation candidates and their potential for global health impact](https://doi.org/10.1016/j.vaccine.2024.126408). Vaccine. 2024 Dec;42(26):126408.

10. Allen JD, Ross TM. [Evaluation of Next-Generation H3 Influenza Vaccines in Ferrets Pre-Immune to Historical H3N2 Viruses](https://doi.org/10.3389/fimmu.2021.707339). Frontiers in Immunology. 2021;12:707339.

11. Dugan HL, Guthmiller JJ, Arevalo P, Huang M, Chen YQ, Neu KE, et al. [Preexisting immunity shapes distinct antibody landscapes after influenza virus infection and vaccination in humans](https://doi.org/10.1126/scitranslmed.abd3601). Science Translational Medicine. 2020 Dec;12(573):eabd3601.

12. 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;

13. Li ZN, Liu F, Gross FL, Kim L, Ferdinands J, Carney P, et al. [Antibody landscape analysis following influenza vaccination and natural infection in humans with a high-throughput multiplex influenza antibody detection assay](https://doi.org/10.1128/mbio.02808-20). Subbarao K, editor. mBio. 2021-02-23, 2021-02;12(1).

14. Jang H, Ross TM. [Hemagglutination Inhibition (HAI) antibody landscapes after vaccination with H7Nx virus like particles](https://doi.org/10.1371/journal.pone.0246613). PloS One. 2021;16(3):e0246613.

15. 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.

16. Huang L, Li X, Guo P, Yao Y, Liao B, Zhang W, et al. [Matrix completion with side information and its applications in predicting the antigenicity of influenza viruses](https://doi.org/10.1093/bioinformatics/btx390). Bioinformatics. 2017 Oct;33(20):3195–201.

17. Einav T, Cleary B. [Extrapolating missing antibody-virus measurements across serological studies](https://doi.org/10.1016/j.cels.2022.06.001). Cell Systems. 2022 Jul;13(7):561–573.e5.

18. Arhami O, Rohani P. [TOPOLOW: A MAPPING ALGORITHM FOR ANTIGENIC CROSS-REACTIVITY AND BINDING AFFINITY ASSAY RESULTS](https://doi.org/10.1101/2025.02.09.637307). Bioinformatics; 2025.

19. Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, et al. [Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift](https://doi.org/10.1126/science.1178258). Science (New York, NY). 2009 Oct;326(5953):734–6.

20. Hensley SE. [Challenges of selecting seasonal influenza vaccine strains for humans with diverse pre-exposure histories](https://doi.org/10.1016/j.coviro.2014.07.007). Current Opinion in Virology. 2014 Oct;8:85–9.

21. Zacour M, Ward BJ, Brewer A, Tang P, Boivin G, Li Y, et al. [Standardization of hemagglutination inhibition assay for influenza serology allows for high reproducibility between laboratories](https://doi.org/10.1128/cvi.00613-15). Hodinka RL, editor. Clinical and Vaccine Immunology. 2016 Mar;23(3):236–42.

22. Waldock J, Zheng L, Remarque EJ, Civet A, Hu B, Jalloh SL, et al. [Assay harmonization and use of biological standards to improve the reproducibility of the hemagglutination inhibition assay: A FLUCOP collaborative study](https://doi.org/10.1128/msphere.00567-21). Pasetti MF, editor. mSphere. 2021-08-25, 2021-08;6(4):10.1128/msphere.00567–21.

23. Einav T, Creanga A, Andrews SF, McDermott AB, Kanekiyo M. [Harnessing low dimensionality to visualize the antibody–virus landscape for influenza](https://doi.org/10.1038/s43588-022-00375-1). Nature Computational Science. 2023 Feb;3(2):164–73.

24. Azulay A, Cohen-Lavi L, Friedman LM, McGargill MA, Hertz T. [Mapping antibody footprints using binding profiles](https://doi.org/10.1016/j.crmeth.2023.100566). Cell Reports Methods. 2023 Aug;3(8):100566.

25. Catani JPP, Smet A, Ysenbaert T, Vuylsteke M, Bottu G, Mathys J, et al. [The antigenic landscape of human influenza N2 neuraminidases from 2009 until 2017](https://doi.org/10.7554/eLife.90782). Neher RA, Diamond B, editors. eLife. 2024 May;12:RP90782.

26. Loes AN, Tarabi RAL, Huddleston J, Touyon L, Wong SS, Cheng SMS, et al. [High-throughput sequencing-based neutralization assay reveals how repeated vaccinations impact titers to recent human H1N1 influenza strains](https://doi.org/10.1128/jvi.00689-24). Journal of Virology. 2024 Sep;98(10):e00689–24.

27. 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.

28. 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.

29. Nachbagauer R, Choi A, Hirsh A, Margine I, Iida S, Barrera A, et al. [Defining the antibody cross-reactome directed against the influenza virus surface glycoproteins](https://doi.org/10.1038/ni.3684). Nature Immunology. 2017-02-13, 2017-02;18(4):464–73.

30. 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.

31. 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.

32. Anderson CS, McCall PR, Stern HA, Yang H, Topham DJ. [Antigenic cartography of H1N1 influenza viruses using sequence-based antigenic distance calculation](https://doi.org/10.1186/s12859-018-2042-4). BMC bioinformatics. 2018 Feb;19(1):51.

33. Forghani M, Khachay M. [Convolutional Neural Network Based Approach to In Silico Non-Anticipating Prediction of Antigenic Distance for Influenza Virus](https://doi.org/10.3390/v12091019). Viruses. 2020 Sep;12(9):1019.

34. Grantham R. [Amino Acid Difference Formula to Help Explain Protein Evolution](https://doi.org/10.1126/science.185.4154.862). Science. 1974 Sep;185(4154):862–4.

35. Dang CC, Le QS, Gascuel O, Le VS. [FLU, an amino acid substitution model for influenza proteins](https://doi.org/10.1186/1471-2148-10-99). BMC Evolutionary Biology. 2010 Apr;10(1):99.

36. Fonville JM, Wilks SH, James SL, Fox A, Ventresca M, Aban M, et al. [Antibody landscapes after influenza virus infection or vaccination](https://doi.org/10.1126/science.1256427). Science (New York, NY). 2014 Nov;346(6212):996–1000.

37. 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.

38. Kucharski AJ, Lessler J, Cummings DAT, Riley S. [Timescales of influenza A/H3N2 antibody dynamics](https://doi.org/10.1371/journal.pbio.2004974). PLOS Biology. 2018 Aug;16(8):e2004974.

39. Hay JA, Laurie K, White M, Riley S. [Characterising antibody kinetics from multiple influenza infection and vaccination events in ferrets](https://doi.org/10.1371/journal.pcbi.1007294). PLoS Computational Biology. 2019 Aug;15(8):e1007294.

40. Wang Y, Tang CY, Wan XF. [Antigenic characterization of influenza and SARS-CoV-2 viruses](https://doi.org/10.1007/s00216-021-03806-6). Analytical and Bioanalytical Chemistry. 2021-12-14, 2021-12;414(9):2841–81.

41. Bedford T, Suchard MA, Lemey P, Dudas G, Gregory V, Hay AJ, et al. [Integrating influenza antigenic dynamics with molecular evolution](https://doi.org/10.7554/eLife.01914). eLife. 2014;3:e01914.

42. Nuñez IA, Carlock MA, Allen JD, Owino SO, Moehling KK, Nowalk P, et al. [Impact of age and pre-existing influenza immune responses in humans receiving split inactivated influenza vaccine on the induction of the breadth of antibodies to influenza A strains](https://doi.org/10.1371/journal.pone.0185666). PLOS ONE. 2017 Nov;12(11):e0185666.

43. Carlock MA, Allen JD, Hanley HB, Ross TM. [Longitudinal assessment of human antibody binding to hemagglutinin elicited by split-inactivated influenza vaccination over six consecutive seasons](https://doi.org/10.1371/journal.pone.0301157). PLOS ONE. 2024 Jun;19(6):e0301157.

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

45. McElreath R. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. 2nd ed. Boca Raton: Taylor and Francis, CRC Press; 2020. (CRC Texts in Statistical Science).

46. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models. Cambridge ; New York: Cambridge University Press; 2007. (Analytical methods for social research).

47. Gelman A, Vehtari A, Simpson D, Margossian CC, Carpenter B, Yao Y, et al. Bayesian Workflow [Internet]. arXiv; 2020 [cited 2025 Apr 15]. Available from: <https://arxiv.org/abs/2011.01808>

48. Breen R. [Regression Models: Censored, Sample Selected, or Truncated Data](https://doi.org/10.4135/9781412985611). SAGE Publications, Inc.; 1996.

49. Posit team. RStudio: Integrated development environment for R. Boston, MA: Posit Software, PBC; 2025.

50. Landau WM. The targets r package: A dynamic make-like function-oriented pipeline toolkit for reproducibility and high-performance computing. Journal of Open Source Software [Internet]. 2021;6(57):2959. Available from: <https://doi.org/10.21105/joss.02959>

51. Ushey K, Wickham H. Renv: Project environments [Internet]. 2025. Available from: <https://CRAN.R-project.org/package=renv>

52. Müller K. Here: A simpler way to find your files [Internet]. 2020. Available from: <https://CRAN.R-project.org/package=here>

53. Yoo AB, Jette MA, Grondona M. [SLURM: Simple Linux Utility for Resource Management](https://doi.org/10.1007/10968987_3). In: Feitelson D, Rudolph L, Schwiegelshohn U, editors. Job Scheduling Strategies for Parallel Processing. Berlin, Heidelberg: Springer; 2003. p. 44–60.

54. Jette MA, Wickberg T. [Architecture of the Slurm Workload Manager](https://doi.org/10.1007/978-3-031-43943-8_1). In: Klusáček D, Corbalán J, Rodrigo GP, editors. Job Scheduling Strategies for Parallel Processing. Cham: Springer Nature Switzerland; 2023. p. 3–23.

55. Bürkner PC. [brms: An R package for Bayesian multilevel models using Stan](https://doi.org/10.18637/jss.v080.i01). Journal of Statistical Software. 2017;80(1):1–28.

56. Bürkner PC. [Advanced Bayesian multilevel modeling with the R package brms](https://doi.org/10.32614/RJ-2018-017). The R Journal. 2018;10(1):395–411.

57. Bürkner PC. [Bayesian item response modeling in R with brms and Stan](https://doi.org/10.18637/jss.v100.i05). Journal of Statistical Software. 2021;100(5):1–54.

58. Gabry J, Češnovar R, Johnson A, Bronder S. Cmdstanr: R interface to ’CmdStan’ [Internet]. 2024. Available from: <https://github.com/stan-dev/cmdstanr>

59. Stan Development Team. CmdStan User’s Guide. 2025.

60. Stan Development Team. Stan User’s Guide. 2025.

61. Stan Development Team. Stan Reference Manual. 2025.

62. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. [Welcome to the tidyverse](https://doi.org/10.21105/joss.01686). Journal of Open Source Software. 2019;4(43):1686.

63. Ching T. qs2: Efficient serialization of r objects [Internet]. 2024. Available from: <https://CRAN.R-project.org/package=qs2>

64. Kay M. tidybayes: Tidy data and geoms for Bayesian models [Internet]. 2023. Available from: <http://mjskay.github.io/tidybayes/>

65. Borchers HW. Pracma: Practical numerical math functions [Internet]. 2023. Available from: <https://CRAN.R-project.org/package=pracma>

66. Allaire J, Dervieux C. Quarto: R interface to ’quarto’ markdown publishing system [Internet]. 2024. Available from: <https://github.com/quarto-dev/quarto-r>

67. Xie Y. Knitr: A general-purpose package for dynamic report generation in r [Internet]. 2024. Available from: <https://yihui.org/knitr/>

68. Xie Y. Dynamic documents with R and knitr [Internet]. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC; 2015. Available from: <https://yihui.org/knitr/>

69. Xie Y. Knitr: A comprehensive tool for reproducible research in R. In: Stodden V, Leisch F, Peng RD, editors. Implementing reproducible computational research. Chapman; Hall/CRC; 2014.

70. Arel-Bundock V, McCrain J. [Software citations in political science.](https://doi.org/10.1017/S1049096523000239) PS: Political Science & Politics. 2023;April:1–4.

71. Wickham H. ggplot2: Elegant graphics for data analysis [Internet]. Springer-Verlag New York; 2016. Available from: <https://ggplot2.tidyverse.org>

72. Pedersen TL. Patchwork: The composer of plots [Internet]. 2024. Available from: <https://CRAN.R-project.org/package=patchwork>

73. Kay M. [ggdist: Visualizations of distributions and uncertainty in the grammar of graphics](https://doi.org/10.1109/TVCG.2023.3327195). IEEE Transactions on Visualization and Computer Graphics. 2024;30(1):414–24.

74. Kay M. ggdist: Visualizations of distributions and uncertainty [Internet]. 2024. Available from: <https://mjskay.github.io/ggdist/>

75. Gohel D, Skintzos P. Flextable: Functions for tabular reporting [Internet]. 2024. Available from: <https://CRAN.R-project.org/package=flextable>