Effect modification of the relationship between immune response and pre-existing immunity by vaccine dose for the seasonal influenza vaccine

Zane Billings, Tzu-Chun Chu

2021-10-23

# 1 Summary/Abstract

# 2 Background and Introduction

While influenza has been well-known for over 100 years and has caused 4 pandemics within that time span, the full scope of the human immune response to influenza viruses remains not fully understood. The difficulty in developing antiviral drugs and vaccines for influenza viruses is partially due to the fast rate of antigenic drift which influenza viruses display. Two major proteins for predicting pathogenicity and immunogenicity of influenza virus are the capsule surface proteins hemagglutinin (HA), which facilitates the viral entry process, and neuraminidase (NA), which facilitates the viral exit process. The genetic sequence encoding the unique properties of variants of these proteins are what give rise to the naming system of influenza A subtypes, e.g. H1N1. Hemagglutinin inhibition (HAI) titer, measured as the lowest of a series of two-fold dilutions at which hemagglutination is observed for a serum sample, is generally accepted by the influenza research community as a correlate of protection from influenza (that is, HAI titer is correlated with the strength of immune response to influenza).

Recent mechanistic modeling of the influenza immune response at the host level (Zarnitsyna, Lavine, Ellebedy, Ahmed, & Antia, 2016) predicts that fold change in antibody levels after vaccination with a seasonal influenza vaccine will exhibit a negative linear relationship with preexisting antibody levels on a log-log scale, with an additional “threshold” or so-called “antigen ceiling” effect–if the host has a high-enough level of preexisting immunity, there will be no response to the influenza vaccine. This “threshold” can be overcome by increasing the dose of the vaccine.

## 2.1 Description of data and data source

The data we will use for this project is from an ongoing cohort study of immune response to the seasonal influenza vaccine conducted by Ted Ross at the School of Veterinary Medicine at the University of Georgia. The study design is a prospective, open cohort where a few hundred participants receive either the FluZone, FluZone HD, or FluMist seasonal influenza vaccine each year since 2014. Patients receive a pre-vaccination and a 28-day vaccination blood draw. Hemagglutination inhibition titers are conducted on serum samples for both the homologous strains of influenza included in the vaccine, as well as a wide panel of historical strains. Influenza A strains which are included are from both H1N1 and H3N2 subtypes, and influenza B strains tested against include pre-divergence historical strains, as well as strains from both the Victoria-like and Yamagata-like lineages. Several other host factors including age, sex, BMI, history of vaccine and comorbidities were also collected. The study design and assay collection methods have been described previously, see (Ross et al., 2014).

The data is provided to the Handel research group through UGA’s CIVR-HRP site, a division of the NIH CIVICs program, and will be accessible by both of us. The data has been partially cleaned already, but is continually in a process of wrangling and refinement, and addressing our specific question will likely involve further cleaning and reshaping of the data. All data are available in a private GitHub repository, but we hope to obtain permission from the Ross group to post a subset of the data publicly as part of the project and the paper which we hope will result from the project. We will ensure that the data is completely anonymized before release and will ensure that all project participants have sufficient training (if necessary) to access the data.

## 2.2 Questions/Hypotheses to be addressed

To that end, we propose the following reach question. Does vaccine dose modify the effect of pre-existing immunity on immune response to the seasonal influenza vaccine?

We hypothesize that dose does modify this relationship, but there are likely interactions with other host covariates, such as age or influenza subtype imprinting probability (Arevalo, McLean, Belongia, & Cobey, 2020), so this relationship is likely not as simple as it appears in the mechanistic models.

# 3 Methods

Several techniques could be used to analyze this relationship, such as: CART models if a tree-based approach is desirable; Unsupervised clustering or trajectory clustering across longitudinal data to identify latent groupings of covariates which could potentially interact to determine outcomes; and Advanced machine learning algorithms such as support vector machines or neural networks to predict an individual’s vaccine response.

However, we hypothesize that this relationship is causal, so prediction algorithms may not be the best fit here. While advanced causal analyses are potentially outside the scope of this analysis, we propose to use generalized linear mixed-effects models (GLMMs), potentially in a Bayesian framework, in order to gain insight about effects on influenza vaccine response.

## 3.1 Data import and cleaning

###############################  
# processing script  
###############################  
  
#this script loads the cleaned data from the CIVIC repo, cleans it with additional steps  
#and saves it as Rds file in the processed\_data folder  
  
#load required packages  
library(here) #for data loading/saving

## Warning: package 'here' was built under R version 4.0.5

library(tidyverse) # for data wrangling

## Warning: package 'tidyverse' was built under R version 4.0.5

## Warning: package 'ggplot2' was built under R version 4.0.5

## Warning: package 'tibble' was built under R version 4.0.5

## Warning: package 'tidyr' was built under R version 4.0.5

## Warning: package 'dplyr' was built under R version 4.0.5

## Warning: package 'stringr' was built under R version 4.0.5

## Warning: package 'forcats' was built under R version 4.0.5

library(lubridate) # working with date and time data   
  
# load data  
dat\_orig <- readRDS(here::here("data","raw\_data","clean\_data.rds"))  
  
  
######################################  
# Data cleaning and wrangling  
######################################  
  
# look up data structure and summary  
str(dat\_orig)  
summary(dat\_orig)  
  
# check percentage of missing values in each column by flu season  
aggregate(dat\_orig[,3:30], by=list(dat\_orig$season), FUN = function(x) { sum(is.na(x))/length(x)\*100 })  
  
# It looks like 'bmi' and 'date\_vaccinated' was not collected in 2014, 2015, and were mostly missing in 2016. Large proportion of 'race2' was missing in 2019 and some were missing in 2020. Need to recode 'race' column.   
  
# check race variables  
table(dat\_orig$race)  
table(dat\_orig$race2)  
table(dat\_orig$race, dat\_orig$race2)  
  
  
white <- c("Polynesian","White","White/ Caucasian","White/Caucaian","White/Caucasian")  
black <- c("Black","Black or African American","Black/African American")  
hisp <- unique(dat\_orig$race)[grep("Hispanic", unique(dat\_orig$race))] # extract values that contain "Hispanic"  
  
dat\_clean <- dat\_orig %>%   
 # create new variable called race4 that contains four groups  
 mutate(race4 = ifelse(race %in% white, "White",  
 ifelse(race %in% black, "Black",   
 ifelse(race %in% hisp, "Hispanic", "Other"))),  
 # create new obesity variable   
 obesity = ifelse(bmi >= 30, "Yes", "No"))   
   
# create days before vaccination variable   
REF\_DATE <- "-09-01"  
  
dat\_clean2 <- dat\_clean %>%  
 tidyr::drop\_na(prevactiter, titerincrease) %>%  
 dplyr::mutate(  
 year = factor(season),  
 days\_before\_vac = lubridate::time\_length(  
 x = ifelse(  
 test = is.na(season),  
 yes = NA,  
 no = lubridate::interval(  
 start = lubridate::ymd(paste0(season, REF\_DATE)),  
 end = date\_vaccinated  
 )  
 ),  
 unit = "day"  
 )  
 )  
  
  
dat\_long <- dat\_clean2 %>%  
 # pivot the data. This makes it easier to plot for all strains.  
 tidyr::pivot\_longer(  
 ends\_with("vaccine\_fullname"),  
 names\_to = "subtype", values\_to = "vaccine\_component"  
 ) %>%  
 dplyr::mutate(  
 subtype = factor(case\_when(  
 startsWith(subtype, "h1n1") ~ "H1N1",  
 startsWith(subtype, "h3n2") ~ "H3N2",  
 startsWith(subtype, "bvic") ~ "B-Vic",  
 startsWith(subtype, "yama") ~ "B-Yam",  
 TRUE ~ "uh-oh"  
 ))  
 )  
  
# Save as RDS  
saveRDS(dat\_clean2, here::here("data","processed\_data","clean\_data.rds"))  
saveRDS(dat\_long, here::here("data","processed\_data", "long\_data.Rds"))

# 4 Results

## 4.1 Part I: Exploratory analysis

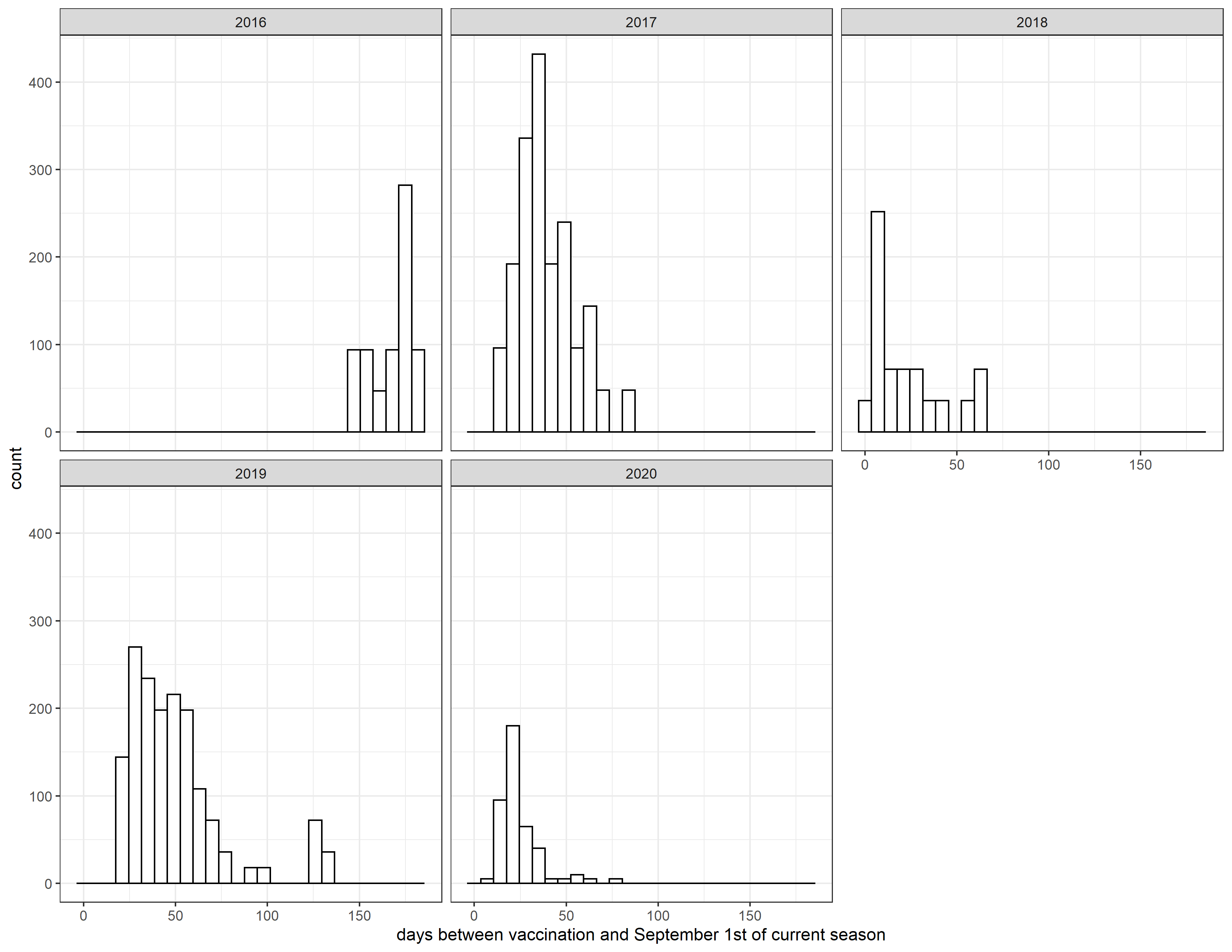
## Warning: package 'gtsummary' was built under R version 4.0.5

## Warning: package 'gt' was built under R version 4.0.5

Table ?? shows a table summarizing the host characteristics stratified by vaccine dose between 2014 and 2019.

Table 1 shows a descriptive summary of the host characteristics by vaccine dose across each flu season. Female, Black and individuals with obesity or prior season vaccination were more likely to receive high dose vaccine.

## 4.2 Distribution of times since vaccination and dose

Figure ?? shows a distribution of time since vaccination. 

# 5 Here the bin size is 7 so the histogram is effectively showing how many elderly people were vaccinated per week.

# 6 Not sure what is going on with the 2016 flu season data so it is probably best to make sure there are not any errors here.

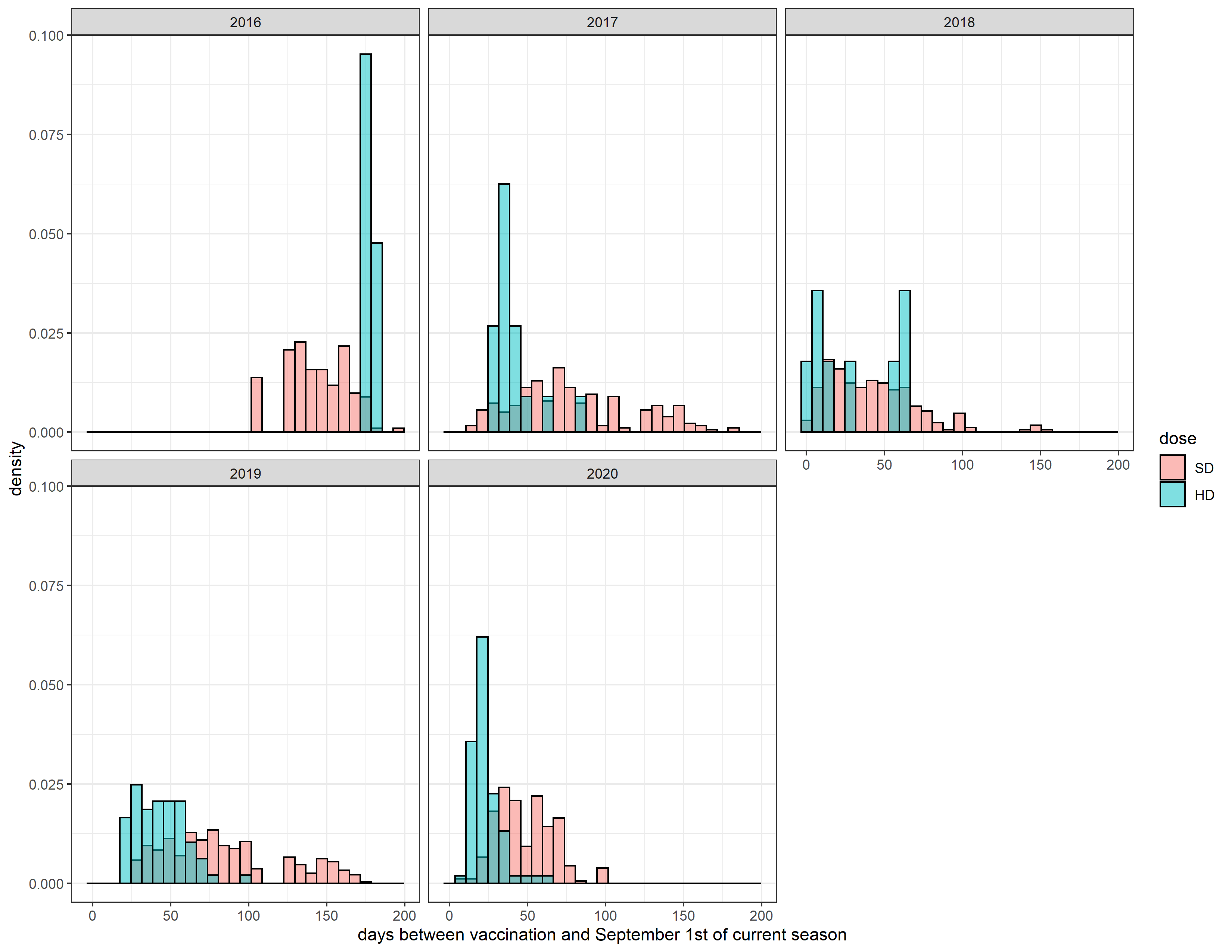
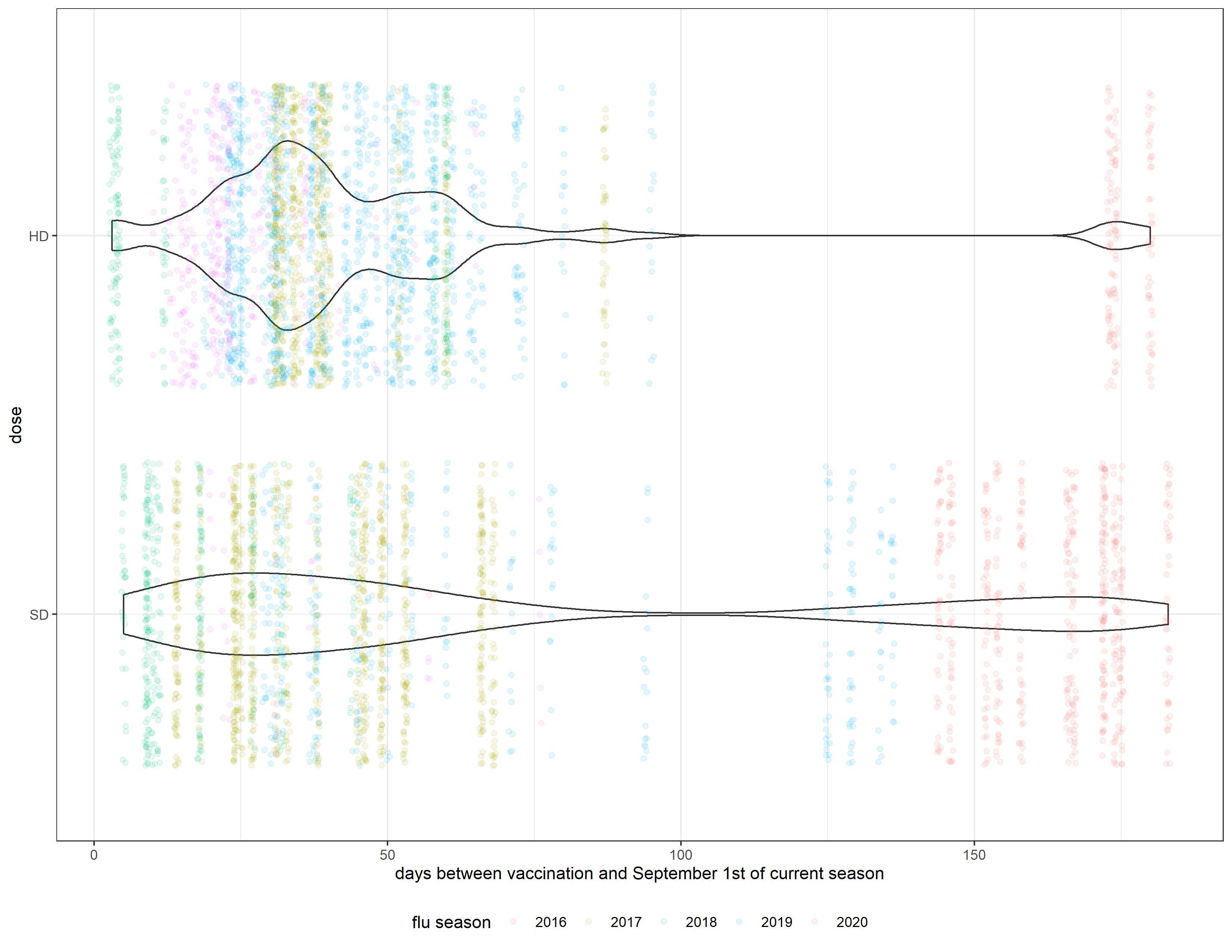
Figure ?? shows a distribution of vaccine dose. 

Figure ?? shows a distribution of times versus vaccine dose. 

## 6.1 Evaluate relationship between main predictors of interest and outcome

### 6.1.1 Outcome: immune reseponse measured by titer increase

### 6.1.2 Main predictors: vaccine dose and other host factors including dose, obesity, prior year vaccination and age

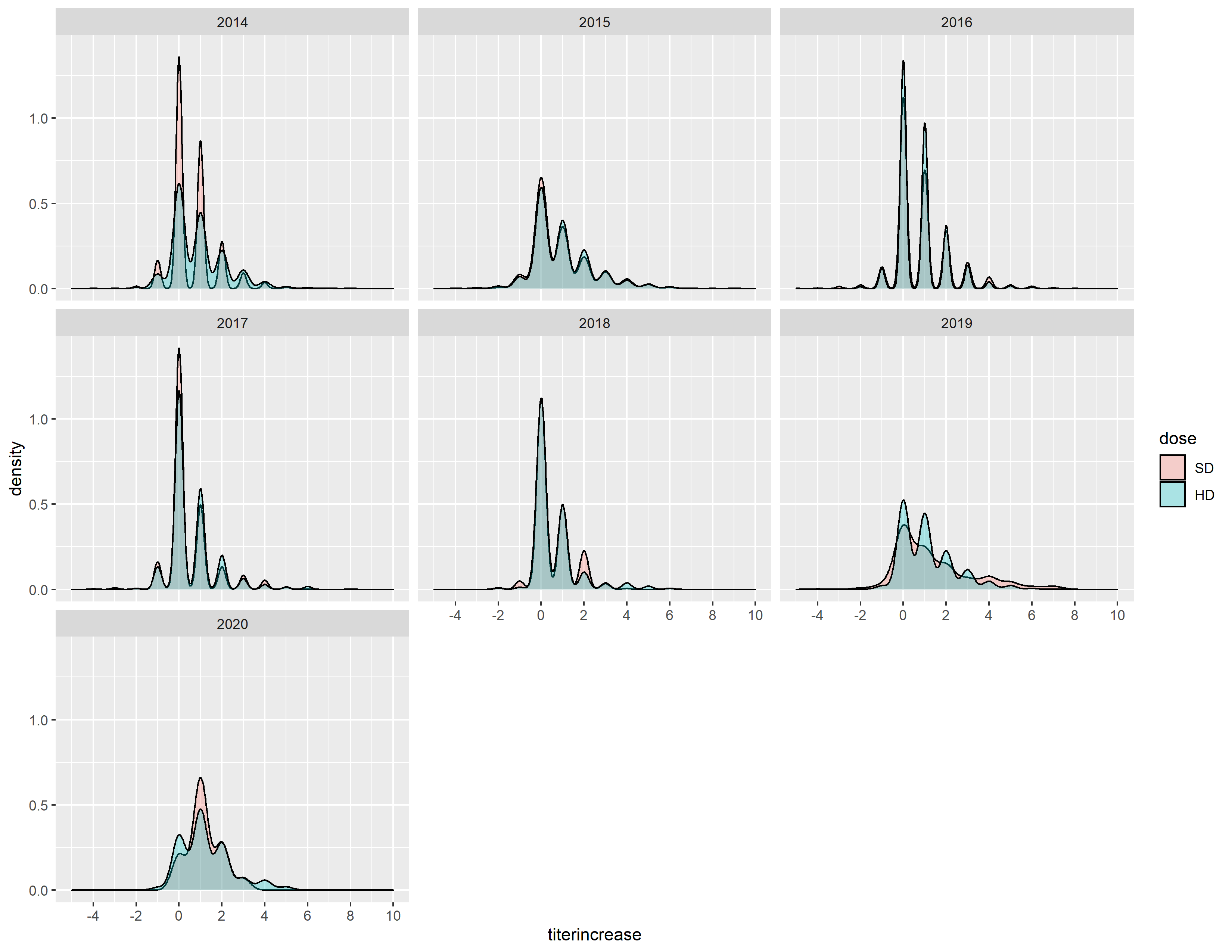
Figure ?? represents a distribution of titer increase by vaccine dose 

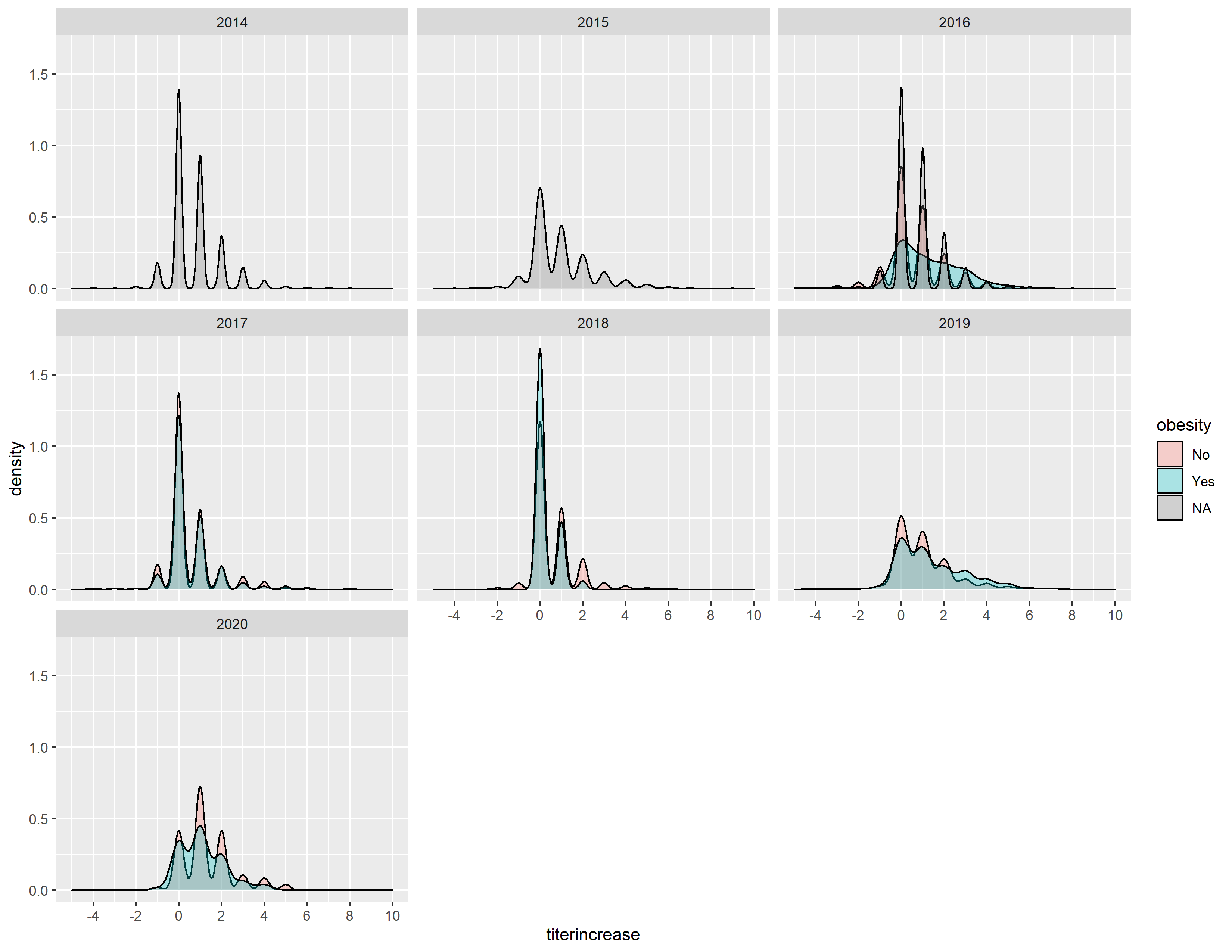
Figure ?? represents a distribution of titer increase by obesity status 

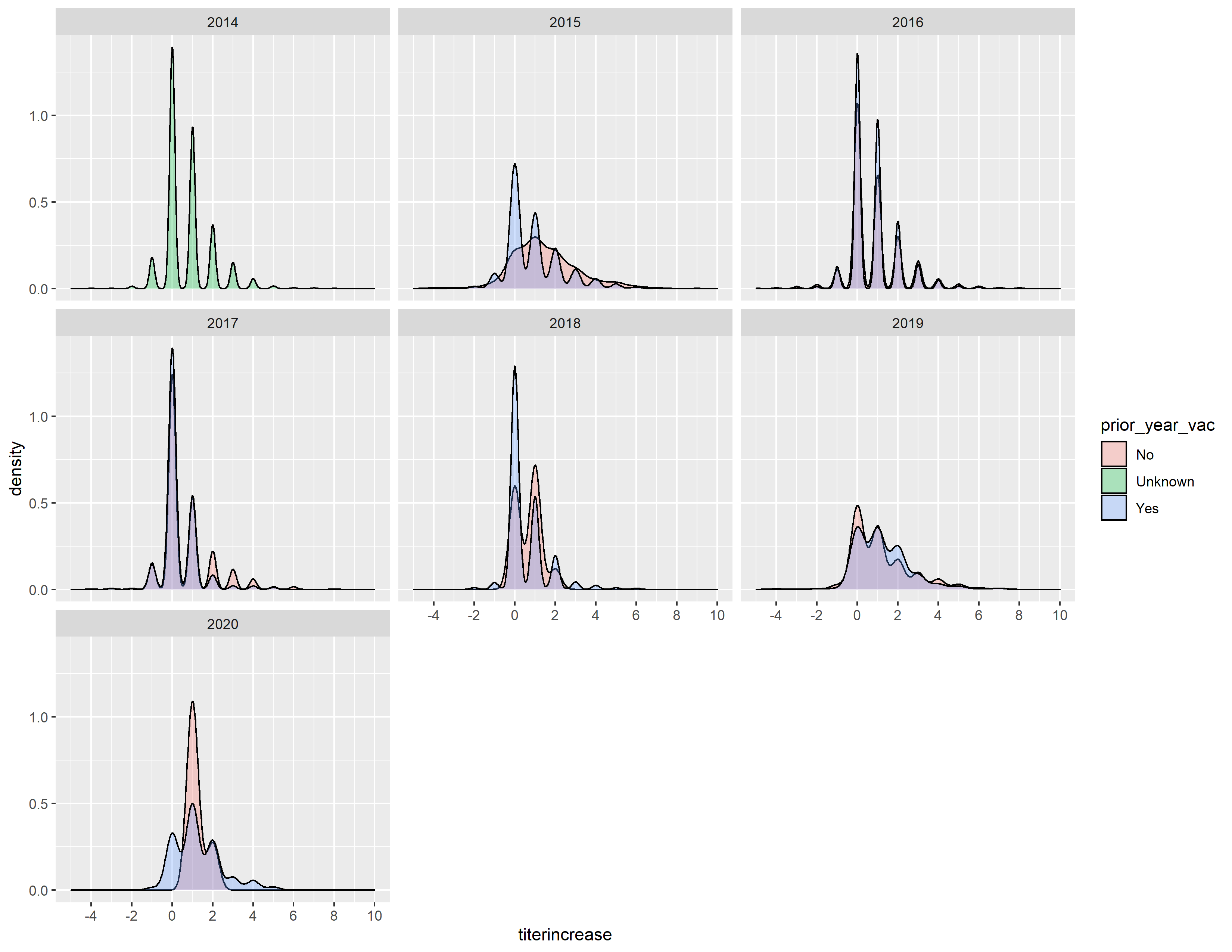
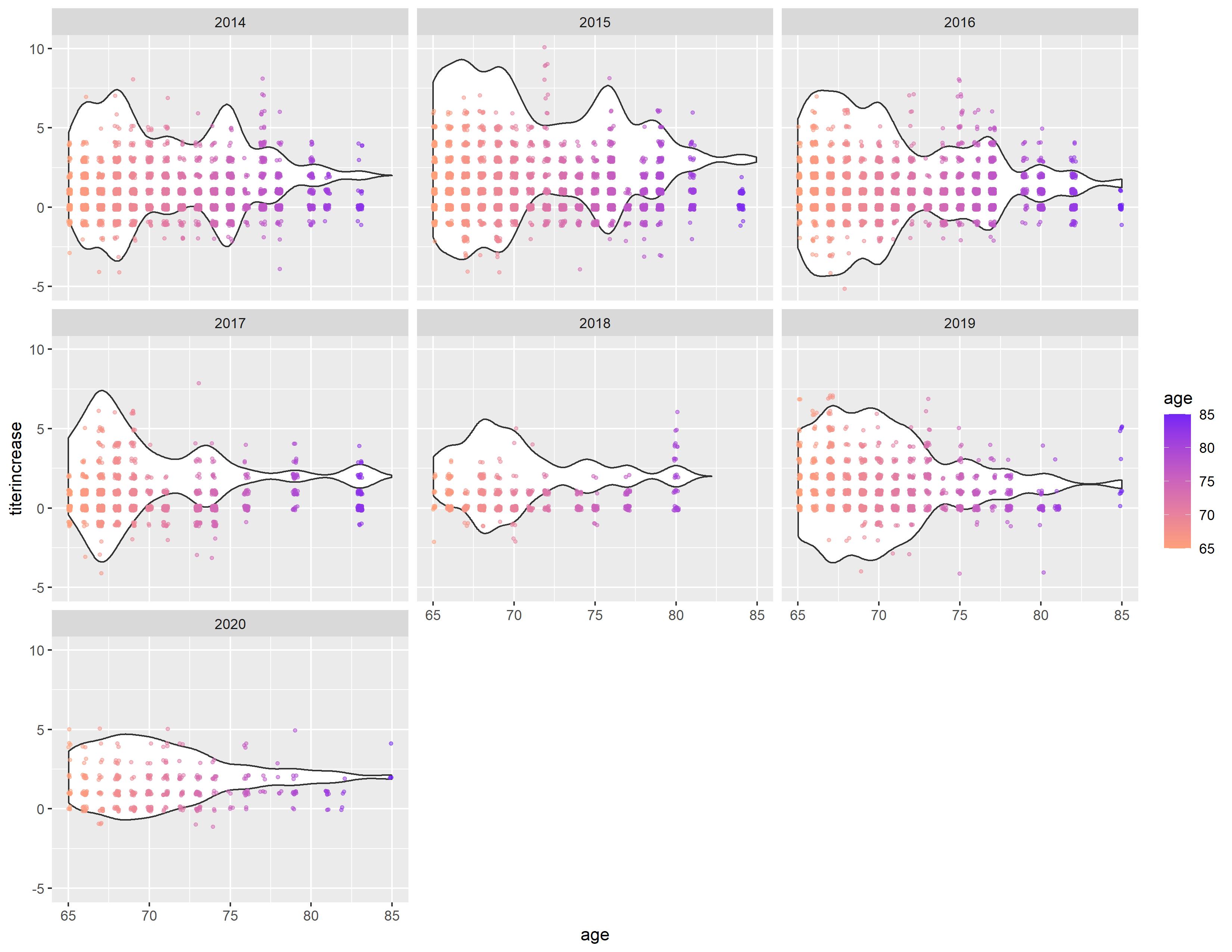
Figure ?? represents a distribution of titer increase by prior vaccination 

Figure ?? represents a distribution of titer increase by age 

## 6.2 Evaluate relationship between pre-existing immunity and immune response stratified by strains, vaccine dose and other host factors

*Use a combination of text/tables/figures to explore and describe your data. You should produce plots or tables or other summary quantities for the most interesting/important quantities in your data. Depending on the total number of variables in your dataset, explore all or some of the others. FIgures produced here might be histograms or density plots, correlation plots, etc. Tables might summarize your data.*

*Continue by creating plots or tables of the outcome(s) of interest and the predictor/exposure/input variables you are most interested in. If your dataset is small, you can do that for all variables. Plots produced here can be scatterplots, boxplots, violinplots, etc. Tables can be simple 2x2 tables or larger ones.*

*To get some further insight into your data, if reasonable you could compute simple statistics (e.g. t-tests, simple regression model with 1 predictor, etc.) to look for associations between your outcome(s) and each individual predictor variable. Though note that unless you pre-specified the outcome and main exposure, any “p<0.05 means statistical significance” interpretation is not valid.*

## 6.3 Full analysis

# 7 Discussion

## 7.1 Summary and Interpretation

## 7.2 Strengths and Limitations

## 7.3 Conclusions

This paper (**???**) discusses types of analyses.

Note that this cited reference will show up at the end of the document, the reference formatting is determined by the CSL file specified in the YAML header. Many more style files for almost any journal [are available](https://www.zotero.org/styles). You also specify the location of your bibtex reference file in the YAML. You can call your reference file anything you like, I just used the generic word references.bib but giving it a more descriptive name is probably better.

# 8 References

Arevalo, P., McLean, H. Q., Belongia, E. A., & Cobey, S. (2020). Earliest infections predict the age distribution of seasonal influenza A cases. *eLife*, *9*. <https://doi.org/10.7554/eLife.50060>

Ross, T. M., Lin, C. J., Nowalk, M. P., Huang, H.-H., Spencer, S. M., Shay, D. K., … Zimmerman, R. K. (2014). Influence of pre-existing hemagglutination inhibition titers against historical influenza strains on antibody response to inactivated trivalent influenza vaccine in adults 50-80 years of age. *Human Vaccines & Immunotherapeutics*, *10*(5), 1195–1203. <https://doi.org/10.4161/hv.28313>

Zarnitsyna, V. I., Lavine, J., Ellebedy, A., Ahmed, R., & Antia, R. (2016). Multi-epitope Models Explain How Pre-existing Antibodies Affect the Generation of Broadly Protective Responses to Influenza. *PLoS Pathogens*, *12*(6), e1005692. <https://doi.org/10.1371/journal.ppat.1005692>