PM 566 Final Project: Analysing Vaccine Preventable Disease Trends in California, by County (2001-2022)

University of Southern California, Fall 2024

Kendall Dimson

Introduction

Long since the rollout of the polio vaccine, there have been a variety of vaccines created in the 1900s to present-day that have been influential in the epidemiological shift, with vaccine efficiency as one of the main modes of primary prevention in infectious disease epidemiology. However, despite continuous efforts to vaccinate, many Vaccine-Preventable diseases are prevalent to this day that threaten the public health system. As a result, in the modern era there have been several vaccine-preventable epidemics that have occurred. According to the Center for Disease Control and Prevention, epidemics are defined as the occurrence of more cases of disease than expected in a given area or among a specific group of people over a particular period of time. There are various factors that can influence the occurrence of these epidemics, including vaccination rates and vaccine hesitancy.

This project serves to conduct a comprehensive exploratory analysis into vaccine-preventable diseases in California. Various research questions include: Where did the previous vaccine-preventable disease epidemics occur in California? What is the normalized distribution in each county of the vaccine-preventable disease with the most cumulative incidence? Is there an association between patterns of vaccine hesitancy and vaccine-preventable disease incidence across counties? Is there an association between patterns of childhood immunization rates and vaccine-preventable disease incidence across counties?

Methods

The dataset, vaccine, was extracted through the California Health and Human Services Agency, with data collected from the California Department of Public Health. The data includes the variables: ID, Disease (the disease reported), County (the location the disease was reported from), Year (the year that the disease was reported to CDPH), and count (the cumulative

number of cases reported of a given disease in a given year, from the specific county). The dataset includes fifteen Disease categories:

Diphtheria, Invasive Meningococcal Disease, Measles, Pertussis, Tetanus, Mumps, Rubella, Hepatitis A, Acute Hepatitis B, Acute Hepatitis C, Varicella Hospitalizations, Varicella Hospitalization/Death, Chronic Hepatitis B, Haemophilus influenzae, Perinatal Hepatitis B, and Rubella (Congenital Syndrome).

To evaluate vaccine hesitancy in association to vaccine-preventable disease incidence, the dataset "Public Use Microdata Area with estimated COVID-19 vaccine hesitancies" was extracted from the Center of Disease Control and Prevention, based on the Health and Human Services' Household Pulse Survey conducted from May 26 - June 7, 2021. The measure was based on the survey question, "Once a vaccine to prevent COVID-19 is available to you, would you…get a vaccine?" Options for the survey included: 1) "definitely get a vaccine" 2) "probably get a vaccine" 3) "unsure" 4) "probably not get a vaccine" 5) "definitely not get a vaccine."

From there, the dataset measures the percentage of individuals in all U.S. counties that selected "probably not" and "definitely not," defining vaccine hesitancy. The dataset also includes race/ethnicity percentage distribution in each Public Use Microdata Area. For this project's analysis, the dataset is stratified to California counties.

In the data cleaning and wrangling step, counties that recorded 0 cases of a disease were removed. Variety of subsetted datasets were created, subsetting into "california" and "county" datasets since there are some categories in the "County" variable that have a summation number of diseases in California. For each, the filter() function subset. The kable() function was used to print a table of the total number of vaccine preventable diseases reported in the entirety of the dataset. The arrange() function was used to print the top five vaccine-reportable diseases with highest cumulative incidence, as well as to print the top five counties of each disease. Ggplot barplots were used to visualize trends in incidence counts. To analyze the distribution of pertussis cases by counties, I used the leaflet package to code an interactive bubble map, normalized by population density.

To visually analyze vaccine hesitancy association , leaflet heat maps were utilized.

Results

Table 1: Table 1. Total Cases of Vaccine-Preventable Diseases Reported in California, 2001-2022

Disease	Total Cases
Pertussis	57219
Hepatitis B, Chronic	4462
Hepatitis A	3035
Invasive Meningococcal Disease	2718

Disease	Total Cases
Hepatitis B, Acute	1349
Mumps	1108
Hepatitis C, Acute	1050
Measles	521
Varicella Hospitalizations	406
Varicella Hospitalization/Death	93
Tetanus	86
Rubella	9
Diphtheria	2

There were a total of 72,058 incident vaccine preventable cases recorded. The top five in total cases reported were Pertussis (n=57,219), Hepatitis B, chronic (n=4462) and acute (n=1349), Invasive Meningococcal Disease (n=2718), and Hepatitis A (n=3035).

Conclusion and Summary