Problem Set 4

Santiago Taborga

2025-03-23

Summary

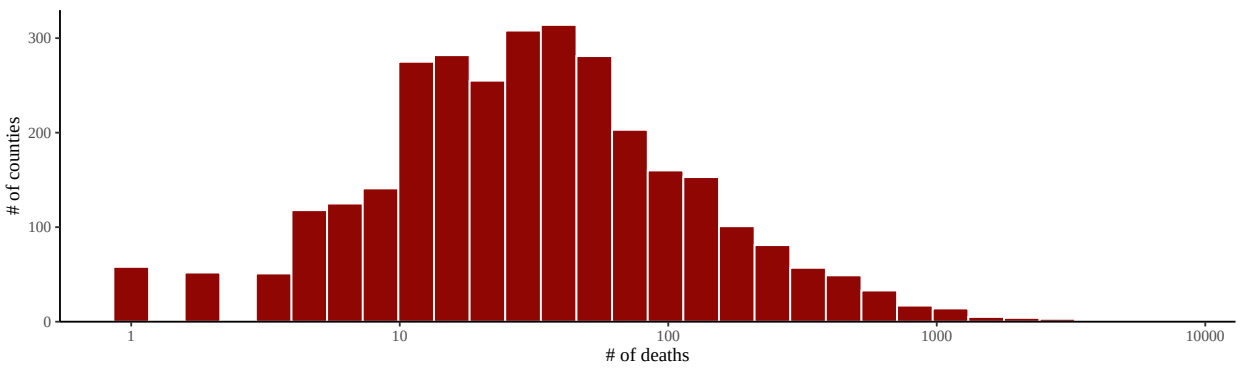
This report examines the effects of various interventions, including mask usage and vaccination, on COVID-19 deaths in U.S. counties between 2021 and 2022. Data from multiple sources were used, including COVID-19 death counts, mask usage surveys, vaccination rates, and the Social Vulnerability Index (SVI). Three fixed-effects linear models were developed to assess the impact of mask usage, vaccination rates, and their combined effects on COVID-19 mortality. The analysis found that both mask usage and vaccination rates are associated with a reduction in COVID-19 deaths. Specifically, increased mask usage was linked to a decrease of approximately 113 deaths per unit increase, while higher vaccination rates were associated with about 1.12 fewer deaths per unit increase. The combined model, which includes both interventions, showed a weaker but still significant negative relationship with COVID-19 deaths. Vulnerable populations, as indicated by the SVI, experienced higher mortality across all models. These findings suggest that both practices are effective in reducing deaths, with the most socially vulnerable populations benefiting less from these interventions.

Data

The data used in this paper is retrieved from three sources. The first looks at the number of deaths due to COVID-19 at the county level in the United States between 2021 and 2022 due to COVID-19. The second is a survey conducted in July 2020 which looks at how often people reported using masks at the county level. The third is CDC vaccination rates and the SVI, which contains information of socially vulnerable populations at the county level compiled in 2023.

The cleaned dataset includes county level statistics for the total population number, the number of deaths due to COVID-19, the percent of people who reported always wearing a mask, the percent of people who are fully vaccinated, the SVI index, and a scale by population of COVID-19 related deaths.

- COVID deaths

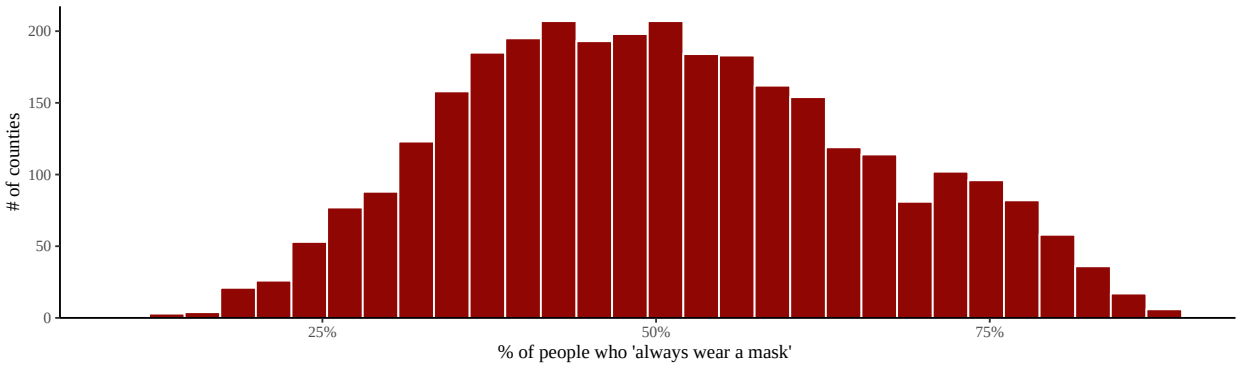


In between 2021 and 2022, the number of COVID deaths at the national level shows that 99.17 percent of counties experienced less than 1,000 deaths. The mean number of deaths across counties is 84.14, with a median of 29, which indicates highly right-skewed data. Thereby, there were some counties that have experienced higher than average deaths, though this indicator does not factor in total population numbers, which might be creating this skewness.

The most populated county that falls very close to the mean number of deaths is Champaign, Illinois with a total population of 209,689 people. The most populated county that is within the median number of deaths is Gallatin, Montana with a total population of 114,434 people.

The county with the most deaths nationwide is Los Angeles, California with a total population of 10,039,107 people and 7,034 deaths.

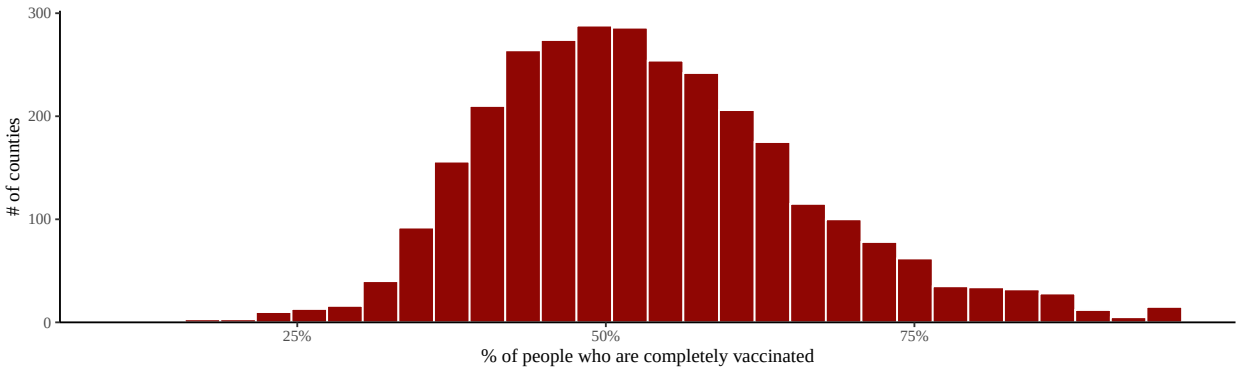
- Percentage "always" masking



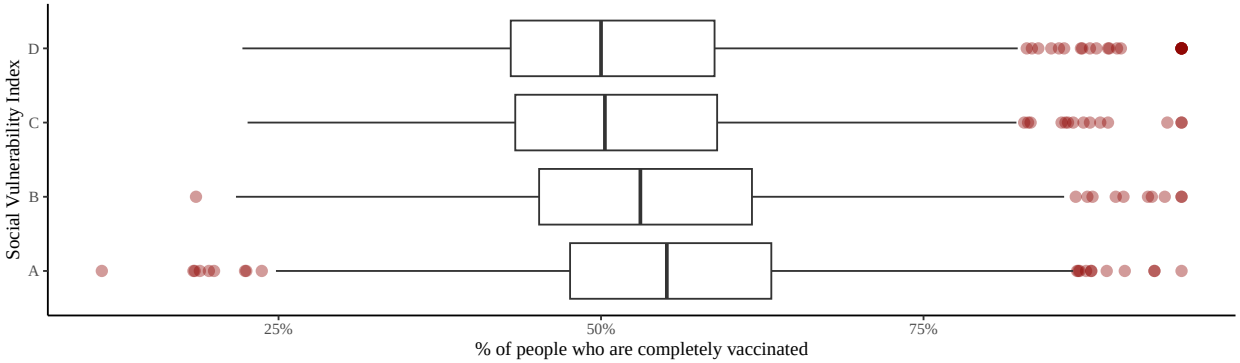
The distribution of counties in which people reportedly 'always wear a mask' is normally distributed. The mean score of people always wearing a mask across counties is 50.77 percent, with a median of 49.7 percent. Around 49.71 percent of counties fall within the first and third quartiles, which are attributed to scores of 39.3 percent and 61.3 percent respectively.

The county with the lowest score for this variable is Valley, Montana. Only 11.5 percent of the people there report always wearing a mask. The county with the highest score is Inyo, California. About 88.9 percent of the people there report always wearing a mask.

- Rates of vaccination



The numbers show that the percent of people across counties who are fully vaccinated is slightly skewed to the right. 54.93 percent of counties report that at least half their population have been fully vaccinated. The mean county has 53.43 percent of their population fully vaccinated, with a mean of 52.1 percent.



The boxplot above illustrates the distribution of counties based on their reported efforts to fully vaccinate individuals based on the SVI Counties that were awarded a score of 'A' had a mean score of 55.75 and a median score of 55.10. Counties that received a valuation of 'B' scored a mean of 54.03 and a median of 53.05. Counties with a 'C' displayed a mean of 52.05 and a median of 50.30. The 'D'-scoring counties were estimated to have a mean score of 52.0 and a median of 50.0.

Below is a table of the highest and lowest performers in fully vaccinating their people:

% fully vaccinated	State	County
95.0	Arizona	Apache
95.0	Arizona	Santa Cruz
95.0	California	Imperial
95.0	Colorado	San Juan
95.0	Georgia	Chattahoochee
95.0	Kansas	Geary
95.0	New Mexico	McKinley
95.0	Texas	Brooks
95.0	Texas	Irion
95.0	Texas	Maverick
95.0	Texas	Presidio
95.0	Texas	Starr
95.0	Texas	Webb
95.0	Wyoming	Teton
11.3	North Dakota	Slope

Analysis

This section focuses on understanding the effects of different practices in reducing COVID-19-related deaths. To achieve this, three fixed effects linear models were constructed: "Masks", "Vaccination", and "Masks and Vaccination". The "Masks" model estimates the relationship between always wearing a mask and COVID-19 deaths, controlling for factors such as population size and the SVI. The "Vaccination" model assesses the impact of vaccination rates on COVID-19 mortality, controlling for the same factors. Finally, the "Masks

	Number of deaths		
	Masks	Vaccination	Masks and Vaccination
Mask Usage	-112.887***		-83.296***
	(10.212)		(10.834)
Population	-0.000***	-0.000***	
	(0.000)	(0.000)	
SVI - B	11.415***	9.968***	9.581***
	(2.884)	(2.878)	(2.859)
SVI - C	15.142***	14.399***	13.128***
	(3.111)	(3.103)	(3.080)
SVI - D	19.811***	18.353***	17.491***
	(3.401)	(3.393)	(3.367)
Vaccine Completion		-1.123***	-0.923***
		(0.093)	(0.099)
Num.Obs.	3049	3049	3049

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

and Vaccination" model combines both interventions to analyze how their combined effects contribute to reducing COVID-19 deaths.

In the Masks model, the coefficient for always wearing a mask (Mask Usage) is -112.89, indicating that increased mask usage is associated with a reduction in COVID-19 deaths by about 113 deaths for each unit increase in constant mask usage, holding other factors constant.

In the Vaccination model, the coefficient for being completely vaccinated (Vaccine Completion) is -1.12 holding other factors constant, indicating that higher vaccination rates are associated with a reduction in COVID-19 deaths by about 1.12 deaths per unit increase in vaccination.

Finally, the Masks and Vaccination model combines both interventions. Here, the coefficient for always.mask is -83.3, which still indicates a negative relationship with COVID-19 deaths, but the effect is less pronounced compared to the separate Masks model. The vax.complete coefficient is -0.92, also showing a negative relationship with COVID-19 deaths, but again weaker than in the Vaccination model.

Given that all numbers are statistically significant at the 99 percent significance level, these results suggest that both interventions—individually and in combination—have notable negative effects on COVID-19 deaths, with more vulnerable populations (as measured by the SVI) consistently experiencing higher mortality.

Appendix: Replication code

```
# Setup -----
knitr::opts_chunk$set(
  echo = FALSE,
  eval = TRUE,
  fig.align = 'center',
  message = FALSE,
  warning = FALSE
)

# packages
  library(tidyverse)
  library(scales)
  library(lfe)
  library(modelsummary)
  library(gt)
  library(data.table)
  library(kableExtra)

# graph theme
  theme_personalized = theme_classic() +
    theme(
      axis.title.x = element_text(size = 11),
      axis.title.y = element_text(size = 11),
      legend.text = element_text(size = 11),
      text = element_text(family = "Times")
    )

# create the dataset -----
## 2021-2022 deaths (BIG FILES)
  covid <-
    data.table::fread('https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties-2022.
    filter(!is.na(fips), state != 'Puerto Rico') %>%
    select(fips, county, state, date, deaths) %>%
    group_by(fips, county, state) %>%
    summarise(deaths = max(deaths, na.rm = T) - min(deaths, na.rm = T))

## estimated mask usage from July 2020 survey
  mask <-
    read_csv('https://raw.githubusercontent.com/nytimes/covid-19-data/master/mask-use/mask-use-by-county
    mutate(
      fips = as.integer(COUNTYFP),
      always.mask = ALWAYS, #always masking
      .keep = 'none'
    ) # for merging

## prep CDC data from directory
  vax <-
    read_csv('cdc vax mar1.csv') %>%
    filter( # drop unknown/incomplete/questionable reports
      FIPS != 'UNK',
      Recip_State != 'VI',
```

```
      Completeness_pct > 0,
      !is.na(Administered_Dose1_Recip)
    ) %>%
    mutate(
      fips = as.integer(FIPS),
      population = Census2019,
      vax.complete = Series_Complete_Pop_Pct, # percent vaxd
      svi.index = SVI_CTGY, # social vulnerability index
      .keep = 'none'
    )

## merge
  covid <-
    left_join(covid, mask) %>%
    left_join(vax) %>%
    mutate(deaths.scaled = deaths / population * 100000) %>%
    ungroup() # scale by population

  rm(mask, vax)
# COVID deaths nationally -----
covid %>%
  ggplot(aes(x = (1 + deaths))) +
  geom_histogram(color = 'white', fill = '#900603') +
  scale_x_log10() + # to mitigate skew
  scale_y_continuous(
    expand = expansion(mult = c(0,0.05))
    ) +
  labs(x = "# of deaths",
       y = "# of counties"
       ) +
  theme_personalized
# Mask usage -----
  covid %>%
    ggplot(aes(x = always.mask)) +
    geom_histogram(color = 'white', fill = '#900603') +
    scale_x_continuous(labels = scales::percent_format(scale = 100)) +
    scale_y_continuous(
      expand = expansion(mult = c(0,0.05))
      ) +
    labs(x = "% of people who 'always wear a mask'",
         y = "# of counties"
    ) +
    theme_personalized
# Rates of vaccination -------------
## Histogram
covid %>%
    ggplot(aes(x = vax.complete)) +
    geom_histogram(color = 'white', fill = '#900603') +
    scale_x_continuous(
      labels = scales::percent_format(scale = 1),
      breaks = seq(0, 75, by = 25)) +
    scale_y_continuous(
      expand = expansion(mult = c(0,0.05))
```

```
    ) +
    labs(x = "% of people who are completely vaccinated",
         y = "# of counties"
    ) +
    theme_personalized
## Boxplot
covid %>%
    filter(!is.na(svi.index)) %>%
    ggplot(aes(x = vax.complete, y = svi.index)) +
    geom_boxplot(outlier.shape = 16, outlier.size = 3, outlier.color = "#900603", alpha = 0.4) +
    scale_x_continuous(
      labels = scales::percent_format(scale = 1)) +
    labs(y = "Social Vulnerability Index",
         x = "% of people who are completely vaccinated"
         ) +
    theme_personalized
## find high/low counties
  covid %>%
    select(vax.complete, state, county) %>%
    filter(vax.complete %in% c(min(vax.complete, na.rm = T),
                               max(vax.complete, na.rm = T))) %>%
    arrange(desc(vax.complete)) %>%
    knitr::kable(
      col.names = c("% fully vaccinated", "State", "County"),
      booktabs = TRUE, align = "c"
    ) %>%
    kable_styling(latex_options = "striped")
# Impact on 2022 COVID deaths -----
## regression estimates
  mods <-
    list(
      "Masks" = felm(deaths.scaled ~ always.mask + population + svi.index | state, data = covid),
      "Vaccination" = felm(deaths.scaled ~ vax.complete + population + svi.index | state, data = covid)
      "Masks and Vaccination" = felm(deaths.scaled ~ always.mask + vax.complete + svi.index | state, da
    )

## regression table
  table <- modelsummary(
    mods,
    coef_order = c("always.mask", "vax.complete", "population", "svi.indexB", "svi.indexC", "svi.indexD
    coef_rename = c(
      "always.mask" = "Mask Usage",
      "vax.complete" = "Vaccine Completion",
      "population" = "Population",
      "svi.indexB" = "SVI - B",
      "svi.indexC" = "SVI - C",
      "svi.indexD" = "SVI - D"
    ),
    gof_map = c('nobs'),
    output = 'gt',
    star = TRUE
  )
```

```r
### Add header above the table
  table %>%
    tab_spanner(
      label = "Number of deaths",
      columns = c(`Masks`, `Vaccination`, `Masks and Vaccination`)
    )
```