Worldwide Bayesian Causal Impact Analysis of Vaccine Administration on Deaths and Cases Associated with COVID-19: A BigData Analysis of 145 Countries

2021-11-15

Abstract

Policy makers and mainstream news anchors have promised the public that the COVID-19 vaccine rollout worldwide would reduce symptoms, and thereby cases and deaths associated with COVID-19. While this vaccine rollout is still in progress, there is a large amount of public data available that permits an analysis of the effect of the vaccine rollout on COVID-19 related cases and deaths. Has this public policy treatment produced the desired effect?

One manner to respond to this question can begin by implementing a Bayesian causal analysis comparing both pre- and post-treatment periods. This study analyzed publicly available COVID-19 data from OWID (⊕Hannah Ritchie and Roser 2020) utlizing the R package *CausalImpact* (⊕Brodersen et al. 2015) to determine the causal effect of the administration of vaccines on two dependent variables that

have been measured cumulatively throughout the pandemic: total deaths per million ($y1$) and total cases per million ($y2$). After eliminating all results from countries with $p > 0.05$, there were 128 countries for $y1$ and 103 countries for $y2$ to analyze in this fashion, comprising 145 unique countries in total (avg. $p < 0.004$).

Results indicate that the *treatment* (vaccine administration) has a strong and statistically significant propensity to causally increase the values in either $y1$ or $y2$ over and above what would have been expected with no treatment. $y1$ showed an increase/decrease ratio of (+115/−13), which means 89.84% of statistically significant countries showed an increase in total deaths per million associated with COVID-19 due directly to the causal impact of treatment initiation. $y2$ showed an increase/decrease ratio of (+105/−16) which means 86.78% of statistically significant countries showed an increase in total cases per million of COVID-19 due directly to the causal impact of treatment initiation. Causal impacts of the treatment on $y1$ ranges from −19% to +19015% with an average causal impact of +463.13%. Causal impacts of the treatment on $y2$ ranges from −46% to +12240% with an average causal impact of +260.88%. Hypothesis 1 Null can be rejected for a large majority of countries.

This study subsequently performed correlational analyses on the causal impact results, whose effect variables can be represented as $y1.E$ and $y2.E$ respectively,

with the independent numeric variables of: *days elapsed since vaccine rollout began ($n1$), total vaccination doses per hundred ($n2$), total vaccine brands/types in use ($n3$)* and the independent categorical variables *continent ($c1$), country ($c2$), vaccine variety ($c3$).* All categorical variables showed statistically significant (avg. p: < 0.001) postive Wilcoxon signed rank values ($y1.E\ V$:[$c1$ 3.04; $c2$: 8.35; $c3$: 7.22] and $y2.E\ V$:[$c1$ 3.04; $c2$: 8.33; $c3$: 7.19]). This demonstrates that the distribution of $y1.E$ and $y2.E$ was non-uniform among categories. The Spearman correlation between $n2$ and $y2.E$ was the only numerical variable that showed statistically significant results ($y2.E \sim n2$: ρ: 0.34 CI95%[0.14, 0.51], p: 4.91e-04). This low positive correlation signifies that countries with higher vaccination rates do not have lower values for $y2.E$, slightly the opposite in fact. Still, the specifics of the reasons behind these differences between countries, continents, and vaccine types is inconclusive and should be studied further as more data become available. Hypothesis 2 Null can be rejected for $c1$, $c2$, $c3$ and $n2$ and cannot be rejected for $n1$, and $n3$.

The statistically significant and overwhelmingly positive causal impact after vaccine deployment on the dependent variables total deaths and total cases per million should be highly worrisome for policy makers. They indicate a marked increase in both COVID-19 related cases and death due directly to a vaccine deployment that was originally sold to the public as the "key to gain back our

freedoms." The effect of vaccines on *total cases per million* and its low positive association with *total vaccinations per hundred* signifies a limited impact of vaccines on lowering COVID-19 associated cases. These results should encourage local policy makers to make policy decisions based on data, not narrative, and based on local conditions, not global or national mandates. These results should also encourage policy makers to begin looking for other avenues out of the pandemic aside from mass vaccination campaigns.

Some variables that could be included in future analyses might include vaccine lot by country, the degree of prevalence of previous antibodies against SARS-CoV or SARS-CoV-2 in the population before vaccine administration begins, and the Causal Impact of ivermectin on the same variables used in this study.

"Untruth naturally afflicts historical information. There are various reasons that make this unavoidable. One of them is partisanship for opinions and schools. If the soul is impartial in receiving information, it devotes to that information the share of critical investigation the information deserves, and its truth or untruth thus becomes clear. However, if the soul is infected with partisanship for a particular opinion or sect, it accepts without a moment's hesitation the information that is agreeable to it. Prejudice and partisanship obscure the critical faculty and preclude critical investigation. The result is that falsehoods are accepted and transmitted" (⊕Muhammad ibn Khaldun al-Hadrami 1379, 1–2).

— Ibn Khaldun, 1379 A.D.

Background

Policy makers and mainstream news anchors have promised the public that the vaccine deployment worldwide would reduce symptoms, and thereby cases and deaths. While this vaccine deployment is still in progress, there is a large amount of public data available that permits an analysis of the effect of the vaccine deployment on COVID related cases and deaths. Has public policy treatment produced the desired effect? Responding to this question can begin by implementing a Bayesian Causal analysis comparing both pre and post treatment periods.

This is an important question to respond to due to the vaccine mandates that are being

implemented worldwide. People have a right to know if this broad public policy is achieving the desired results. While there are arguments to be made on both sides of this debate, the question of whether a deployment of COVID-19 gene therapy injections cause less death or cases from the virus in any significant way is a testable hypothesis given public data that is now available. With the debates raging over the effectiveness, legality, and ethics of these vaccine mandates, a way to continually monitor the effect between vaccine deployment and worldwide COVID-19 associated death and case rates seems an important contribution to this ongoing discussion.

Some previous work has been done on the correlation between vaccination percentage rates or vaccine type and new cases and deaths (⊕Subramanian and Kumar 2021), however this work has yet to prove conclusive or has only looked at correlation in a limited number of countries (⊕Alhinai and Elsidig 2021). Indeed, the correlation numbers are a clue where to look, but they do not determine chronological order and therefore do not determine causation. It could just as well be that high total deaths per million are associated with high vaccination rates simply because those countries with higher death rates may have had a more frightened population ready to take vaccinations, or it could be that countries with high vaccination rates also had high rates of recording new cases and deaths as "COVID-19 associated," even when there may have been various other comorbidites present in the individual (⊕Disease Control and Prevention 2020). These factors make correlation an important metric, but more of a preliminary

clue in the dark of where to look for causation. To find causation on the scale of BigData analysis we must focus on the causal impact of the effects before and after vaccine administration on as many countries in the world as possible, this study looks at 145 countries.

Research Question 1

Specific: Does the 'beginning of COVID-19 gene therapy injections' (X) have any statistically significant causal effect in decreasing or increasing *total deaths per million* ($y1$) or *total cases per million* ($y2$) associated with COVID-19?

Simplified: Does vaccine deployment cause less or more COVID-19 associated cases or death?

Hypothesis 1

H0: X has no statistically significant[1] causal effect on Yx.

HA: X has a statistically significant[2] causal effect on Yx.

Research Question 2

Specific: Do the statistically significant results of the causal effect on *total deaths per million* ($y1.E$) or *total cases per million* ($y2.E$) correlate with any of the following independent variables:

- numerical variables:
  - $n1 = $ *days elapsed since vaccine administration*
  - $n2 = $ *total vaccinations per*

hundred

- $n3 = $ *total vaccine types/brands in use*

- categorical variables:
  - $c1 = $ *continent*
  - $c2 = $ *country*
  - $c3 = $ *vaccine type*

Simplified: Are the results of Research Question 1 associated with either of the following variables? :

1. the length of time vaccines have been in use in a country

2. the number of vaccines they have administered

3. the number of brands/types of vaccines they have administered

4. the continent where the vaccines were administered

5. the country where the vaccines were administered

6. the types of vaccines they have administered

Hypothesis 2

H0: X has no statistically significant[3] correlation with $yx.E$.

HA: X has a statistically significant[4] correlation with $yx.E$.

Methods

The methods and code to reproduce this study are as follows:

Obtain up to date COVID-19 data from Our World in Data (⊕Hannah Ritchie and Roser 2020)

It is necessary to download the appropriate data sets including:

1. The data from ⊕Hannah Ritchie and Roser (2020)[5] is updated daily and can be downloaded and converted to a data.frame using the R code in this paper, or can downloaded directly from here as a *.csv*: https://covid.ourworldindata.org/data/owid-covid-data.csv.

2. It was also necessary to convert the data on vaccine types in use by each country (⊕wikipedia 2021)[6] to a *.csv* format, this was done with standard spreadsheet techniques in Google Sheets. This data set I have made publicly available here: https://docs.google.com/spreadsheets/d/1egKoaLyAt_9JoWKqr8uZDxw3xIj9J-Fu33S9sMRL4XQ/edit?usp=sharing

```
# Folder where data is stored, **CHANGE THIS AS NECESSARY**
data_folder <- file.path("./indices[USED]/")

#*** UNCOMMENT FOR NEW DATA ***
# Download most recent OWID COVID-19 data set
#url <- "https://covid.ourworldindata.org/data/owid-covid-data.csv"
#name <- "owid-covid-data.csv"
#download.file(url = url, destfile = paste0(data_folder,name))

## Load the reports ##
# Set directory to location of CSV files first
setwd(data_folder)
# Read files without csv
```

```
filenames <- gsub("\\.csv$","", list.files(pattern="\\.csv$"))
# Load all files
for(i in filenames){
  assign(i, read.csv(paste(i, ".csv", sep=""),fileEncoding="latin1"))
}
```

Clean data and merge datasets

Once all the data is obtained it can be 'cleaned'
for faster data processing. This involves removing
unnecessary variables from the data sets and
renaming columns so all variable names are in
agreement. This can be done with the R code
included in the supplmentary material for this
report.

Run Causal Analysis for all dependent variables

The R package *CausalImpact* (⊕Brodersen et al.
2015) utilizes a robust series of Bayesian
calculations along with predictor data sets to
determine the likely trajectory of a trend line had
a particular intervention not occurred and then
calculates the difference between that projected
trend line and the real data line. The authors of
the package summarized this impressive set of
calculations and its improvements upon previous
methods as a,

> "…method [that] generalises the
> widely used difference-in-differences
> approach to the time-series setting by
> explicitly modelling the counterfactual
> of a time series observed both before
> and after the intervention. It improves
> on existing methods in two respects: it
> provides a fully Bayesian time-series
> estimate for the effect; and it uses
> model averaging to construct the most
> appropriate synthetic control for

modelling the counter factual"
(⊕Brodersen et al. 2015, 247).

Effectively, this allows us to look at the past 12–16 months (each country is slightly different) before vaccine administration began, this is called the *pre-intervention period*, and utilize that data to project where $y1$ (*total deaths per million*) and $y2$ (*total cases per million*) would have been had the intervention of X (*vaccine administration*) not occurred, what the authors call a "counterfactual" (⊕Brodersen et al. 2015, 248–49). Utilizing this estimated counterfactual and the confidence level associated with that estimation we can then compare it with the actual data available and see if there is any difference. If the projected estimation is higher than the actual results, it will appear as a negative impact, while if the projected estimation is lower than the actual results, it will appear as a positive impact.

Another aspect of the *CausalImpact* package is the ability to add control variables that are combined into a "synthetic control" (⊕Brodersen et al. 2015) that closely mimic the trend line, but that are exogenous (i.e. have not been affected by the intervention); this allows for even more accurate predictions and normally a lower p–value and a lower standard deviation. To obtain control variable data sets the authors of *CausalImpact* recommend utilizing similar data that did not receive the same treatment, such as a different region or country, any trend line that can closely mimic or even be constant alongside the trend line one is testing (e.g. similar data from an area with no intervention, price charts, stock prices, temperature records, etc.). As there are virtually

zero countries that have been unaffected by the vaccines, finding a set of control variables is difficult, though not impossible. Ultimately, this study chose to utilize the data of four countries in Africa (Burkina Faso, Chad, DRC, South Sudan) that were chosen specifically for their low average severity indices since vaccine administration began (i.e. low levels of mandatory mask wearing, social distancing, crowd limitations, travel restrictions, etc.) and for their low vaccination rates. These countries in the author's estimation best represent a "natural" progression of the virus with limited vaccine intervention on par with most other nations, making them the least likely to display problems of endogeneity while still acting as valid control groups.

These carefully selected control variables as a synthetic control combined with the formula presented by ⊕Brodersen et al. (2015), which utilizes the data pre-intervention as part of its equation to predict the counterfactual, means that the other synthetic control effectively being utilized internally in the equation is a dynamic calculation for each country based on its own individual experience of COVID–19's impacts on total deaths and cases per million *sans* vaccines for, as mentioned above, approximately 12–16 months.

As ⊕Brodersen et al. (2015) explained,

> "The approach described in this paper inherits three main characteristics from the state–space paradigm. First, it allows us to flexibly accommodate different kinds of assumptions about the latent

state and emission processes underlying the observed data, including local trends and seasonality. Second, we use a fully Bayesian approach to inferring the temporal evolution of counterfactual activity and incremental impact. One advantage of this is the flexibility with which posterior inferences can be summarised. Third, we use a regression component that precludes a rigid commitment to a particular set of controls by integrating out our posterior uncertainty about the influence of each predictor as well as our uncertainty about which predictors to include in the first place, which avoids overfitting" (⊕Brodersen et al. 2015, 251)

In other words, by utilizing the data for total deaths and cases per million from before vaccines existed and combining that with a synthetic control of countries largely non–participatory in the COVID-19 vaccine program, the R package *CausalImpact* is able to produce a high degree of certainty in the results.[7] The control countries' average severity indices since vaccine administration and vaccination rates are listed in Table 1.

This analysis was carried out on all countries in the data set that met the following criteria: (1) more than five observations (i.e. dates), (2) a longer pre intervention period than post intervention period, and (3) no NAs in the observations necessary for calculation.

```
# CAUSAL ANALYSIS --------------------------------------------------------
# Variables for Function
```

```r
# Select data frame
df0 <- df_owid
# Create a countries list
countries <- c(unique(df_owid$ISO))
# Folder where plots will be saved
plots.loc <- file.path("./plots/causalImpact/totalCases/run6/")

## DEPENDENT VARIABLES

################################################################################
# Here you must declare your variables. It is necessary to re-declare variables
# for each y variable that you intend to test. The only variable necessary to
# change is where it says ** CHANGE THIS **, otherwise leave the code as is.
################################################################################

# Declare function name and variables
impactReports <- function(df0, countries, plots.loc) {

# Begin for loop
for (i in countries) {

# Declare Data Frame to Use with Time Series
df <- df0
df <- dplyr::filter(df, ISO == i) # Use ISO codes to loop every country

# Assure the Date column is in the correct format
dates <- as.character(df$date) # change to text
df$date = as.Date(dates, format="%Y-%m-%d") # change to date

# Treatment - Independent Variable
vacc <- dplyr::select(df,date,total_vaccinations_per_hundred) # INDEPENDENT VARIAB
vacc <- unique(vacc) # Make sure there are no duplicate observations
vacc <- na.omit(vacc) # Make sure there are no observations with NA

# Dependent Variable

## Dependent Variable ** CHANGE THIS **
df <- dplyr::select(df,date,total_cases_per_million)

df <- unique(df) # only choose unique values
df <- na.omit(df) # remove all

## Dependent Variable ** CHANGE THIS **
dfy <- df$total_cases_per_million

# Obtain length of Dependent Variable for time-series calculations
dfyN <- as.numeric(length(dfy))
```

```r
# Add in data sets that will be used for creating the Synthetic Control data
# This paper ultimately used the four countries mentioned, however the research
# process also included analyzing the results of other potential control candidate
# Thus, included here are four other countries that were analyzed in this fashion,
# their conclusion ultimately not being used because of lower confidence levels an
# higher standards of deviation suggesting a less reliable synthetic control model
# This data is included here for full transparency and for any other researchers t
# may like to analyze these other potential control countries and their results.

# NOT INCLUDED IN THIS PAPER, BUT INCLUDED IN FULL DATASET
#dfy1 <-  BDI$total_deaths_per_million[1:dfyN] # Burundi
#dfy3 <-  HTI$total_deaths_per_million[1:dfyN]# Haiti
#dfy5 <-  YEM$total_deaths_per_million[1:dfyN] # Yemen
#dfy7 <-  TZA$total_deaths_per_million[1:dfyN] # Tanzania

# SELECTED CONTROL COUNTRIES FOR USE IN CAUSAL IMPACT ANALYSIS
dfy2 <-  COD$total_cases_per_million[1:dfyN] # DRC
dfy4 <-  SSD$total_cases_per_million[1:dfyN] # South Sudan
dfy6 <-  TCD$total_cases_per_million[1:dfyN] # Chad
dfy8 <-  BFA$total_cases_per_million[1:dfyN] # Burkina Faso

# Combine data frames of control countries
dfy <- cbind(dfy,dfy2,dfy4,
             dfy6,dfy8)

# Remove any NA values
dfy <- na.omit(dfy)

# Obtain new length of data frame of the Dependent Variable
dfyN2 <- as.numeric(nrow(dfy))

# Create time-series from this dataframe
dfx <- df$date[1:dfyN2]  # Time Series

# Make sure all data has more than 5 observations
if (length(dfx) < 5 | nrow(dfy) < 5 | nrow(vacc) < 5) {

   # Announce stop and move to next if logic unfulfilled
cat(paste0("###### stopping ",i," at step 2 #######"))
  next
} else {

#Declare Data Frame Variables
df.ts <- dfx # Dates as characters
df.score <-dfy  # Y
df.time <- df$d # Day number
```

```r
#Calculate length of data frame
days <- as.Date(max(df.ts)) - as.Date(min(df.ts)) # Find length of data frame
days <- as.numeric(days) # Make the string a number

#Declare time points from data
time.points <- try(seq.Date(as.Date(min(df.ts)), by = 1, length.out = days))

#Declare time series variables
SCORE.Y <- ts(df.score) # Dependent Variable
TIME.C <- ts(df.time) # Time Series

#Bind them into test groups
test <- try(zoo(cbind(SCORE.Y, TIME.C), time.points))

# Announce start of next step
cat(paste0("###### starting ",i," step 3 #######"))

df <- na.omit(vacc) # Remove any NAs

if (anyNA(df) == TRUE) {

cat(paste0("###### stopping ",i," at step 4 #######"))
  next
  } else {

# Treatment Period
treatmentS <- (min(df$date)) # start of treatment period (i.e. first vaccines admi
treatmentE <- max(df$date) # end of treatment period (i.e. ongoing)

# Pre-Treatment Period
preperS <- min(dfx) # start of pre-treatment period
preperE <- treatmentS - 1 # end of pre-treatment period
# Post-Treatment Period
postperS <- (treatmentS) # start of treatment period (i.e. first vaccines administ
postperE <- treatmentE # end of treatment period
# Use when you need an exact date
pre.period <- c(preperS,preperE) # declare pre period start and end dates
post.period <- c(postperS,postperE) # declare post period start and end dates

# assure pre-period is longer than post-period
if ((preperE - preperS) < (postperE - postperS) | anyNA(test) == TRUE) {

  # Stop and move to next country if logic unfulfilled
  cat(paste0("###### stopping ",i," at step 5 #######"))
  next
```

```r
  } else {
    cat(paste0("###### calculating ",i,"#######")) # Announce beginning of calculation
    impact <- try(CausalImpact(test, pre.period, post.period)) # Calculate Impact
    Sys.sleep(1)
    x <- "Total Vaccinations Per Million" # CHANGE ONLY IF YOU CHANGE X VARIABLE
    y <- "Total Cases Per Million" # ** CHANGE THIS  FOR EACH Y VARIABLE**
    cat(paste0("###### plotting ",i,"#######")) # Announce beginning of plotting
    try(print(plot(impact) +      # Print the plot
      ggplot2::labs(
      title = paste0(i, # Title the plot
                   ": Causal Impact Plot",
                   " | ",
                   x,
                   " effect on ",
                   y),
      caption =
    paste("Source: Data collected from OWID, analyzed and plotted by Kyle Beattie usin
          date())
    ) +
      theme_stata())) # Choose a theme
    Sys.sleep(1) # Let plot be created
    cat(paste0("###### saving plots and reports ",i," #######")) # Announce saving of
    setwd(plots.loc) # Set the drive to the plots folder
    try(savePlotAsImage(paste0(i,      # Save Plot As TITLE
                          "_:_Causal_Impact_Plot",
                          "_|_",
                          x,
                          "_effect_on_",
                          y,
                          "_Report.png"),
                 format = "png",
                 width =  1920, # set ratio sizes
                 height = 1080))
    Sys.sleep(1) # Let plot be saved
    try(summary(impact, "report")) # Produce Summary Report
    try(summary(impact)) # Produce Summary Report 2
    report<<-try(capture.output(summary(impact, "report"))) # Capture Report
    report2<<-try(capture.output(summary(impact)))
    try(cat(report, file = paste0(i,      # Write report to txt
                          ": Causal Impact Plot",
                          " | ",
                          x,
                          " effect on ",
                          y,"Report.txt"), append = TRUE))

    try(stargazer(report2,
            summary = FALSE,
```

```
          type = "text",
          out = paste0(plots.loc,i,        # Write report to txt
                          ": Causal Impact Plot",
                          " | ",
                          x,
                          " effect on ",
                          y,"Report2.txt")))

  #   }
      }
    }
  }
 }
}


# RUN FUNCTION -------------------------------------------------------

# Run this function to produce all the plots and reports above, use try() to pass
try(impactReports(df0,countries,plots.loc), silent = TRUE)
```

Extract Causal Analysis results and merge with main dataset

After completing the causal analysis on all countries and printing a report and plot for each country, the data was filtered and cleaned by removing all results with $p > 0.05$ and turning all causal impact results into both whole numbers (ex. $\gamma_1.effect\_percent$ for Japan = 48%) and decimals (ex. $\gamma_1.effect\_dec$ for Japan = 0.48), both representing the percentage of positive or negative impact in two different forms. This data was then integrated back with the original data frame for further analyses and data presentation.

```
# ANALYZE CAUSAL IMPACT OUTPUT -------------------------------------------


########################### FUNCTION VARIABLES ###########################
# loc_i = the location of text files for import (as file.path)
# loc_o = the location to output csv and txt files (as file.path)
# name = name for csv and txt documents (as character spaces ok)
# y = dependent variable (as character no spaces)
##########################################################################
```

```r
### Extract necessary data from the Causal Impact Analysis Reports ###

dat.extr.CausalReports <- function(loc_i,loc_o,name,y) {

  ## Data location
  data_folder <- file.path(loc_i)
  # Set directory to location of CSV files first
  setwd(data_folder)
  # Read files without csv
  filenames <- gsub("\\.txt$","", list.files(pattern="\\.txt$"))
  # Create data frame first
  df <- data.frame(matrix(ncol = 2, nrow = 0))
  # Load all files
  for(i in filenames){
    df0 <- assign(i, readtext(paste(i, ".txt", sep="")))
    df <- rbind(df,df0)
  }

  ## ISO
  # Extract first three characters to obtain ISO
  df$ISO <- substr(df$doc_id, 1, 3)

  ## p-value
  # Extract all three digit values
  df$p <- str_extract(df$text,"p = 0.\\d{3}")
   # Extract all two digit values
  df$p2 <- str_extract(df$text, "p = 0.\\d{2}")
  # Replace NAs in three digit values with two digit values
  df$p[is.na(df$p)] <- df$p2[is.na(df$p)]
  # Remove two digit column
  df <- dplyr::select(df, -p2)
  # Remove p = sign and convert to numeric
  df$p <- gsub("p = ", "", df$p) %>% as.numeric()
  # Remove all non-statistically significant observations
  df <- filter(df, p < 0.05 )

  ## % Effect
  # separate this part of the text
  df$effect <- str_extract(df$text, "the response variable showed .*?%")
  # remove this part leaving only the %
  df$effect <- gsub("the response variable","",df$effect)
  # remove this text or that
  df$effect <- gsub("^( showed an increase of | showed a decrease of )", "", df$e
  # combine text to make an easy readout for the public
  df$effect_txt <- paste(df$ISO,": Vaccine effect on",y,df$effect)

  ## Interval
```

```r
  # Extract only % numbers between []
  df$interval <- str_extract(df$text, "\\[(\\+|\\-)\\d*%.*?\\]")

  ## Effect as Decimal
  # Remove all symbols and convert to numeric
  df$effect_dec <- gsub("%","",df$effect) %>% as.numeric()
  # Turn percentage to decimal
  df$effect_dec <- df$effect_dec / 100

  ## Merge the effect changes and p values with a new data frame
  # Select only necessary columns
  df5 <- dplyr::select(df,ISO,p,effect,effect_dec)
  setnames(df5, # Change columns names for the main data
           c("p","effect","effect_dec"),
           c(paste0(y,".p"),paste0(y,".effect"),paste0(y,".effect_dec")))

# Write Data Frame to CSV, change data frame, path, and name variables as necessary
write.csv(df5,
          paste0(loc_o,name,date(),".csv"),
          na = "",
          fileEncoding = "UTF-8")

# Write statistical table to txt
stargazer(df5,                      # Export txt
          summary = FALSE,
          type = "text",
          out = paste0(loc_o,name,date(),".txt"))
 }


# RUN FUNCTION
# Variables
loc_i <- file.path("./plots/causalImpact/totalCases/run6/")
loc_o <- file.path("./plots/causalImpact/data_output/")
name <- "CausalAnalysis-vaccines-by-totalCases"
y <- "y2"

# FUNCTION
dat.extr.CausalReports(loc_i,loc_o,name,y)

# Analyze output
df <- read.csv("./plots/causalImpact/data_output/CausalAnalysis-
               vaccines-by-totalCasesTue Oct  28 11:50:13 2021.csv") %>%
  as.data.frame()
cntP <- str_count(df$y2.effect,"\\+") # count number of positive occurrences
count(cntP) # results
```

Correlation Analysis of Causal Analysis Results

After integrating the results of the Causal Impact Analysis with the original data frame, Research Question 2 was addressed through follow-up correlational analyses. These were calculated with the resulting dependent variables from the Causal Impact Analysis ($y1.E$: $y1.effect\_percent$ = effect of vaccine intervention on total deaths per million) and ($y2.E$: $y2.effect\_percent$ = effect of vaccine intervention on total cases per million) utilizing *ggstatsplot* (⊕Patil 2021) for all variables listed in Research Question 2.

Plot Results

Scatter plot and correlational matrix results were used to analyze the statistical significance of the correlation between the dependent variables and the independent numerical variables. The scatter plots and correlational matrices include the ρ (Spearman) scores and their respective p-values. Dot plots were utilized to show the distribution of the dependent variables with the independent categorical variables, they include the Wilcoxon signed rank value and respective p-value.

Materials

The data used in the following analysis comes from the two data sources mentioned above as well as the R packages that are listed at the end of this report. These results were produced using RStudio Version 1.4.1027 (⊕RStudio Team 2020).

Results

After eliminating all results from countries with $p > 0.05$, there were 128 countries for $y1$ and 103 countries for $y2$ to analyze in this fashion (avg. *p-value* < 0.004), 145 unique countries in total. Results indicate that the *treatment* (vaccine administration) has a strong and statistically significant propensity to causally increase the values in either $y1$ or $y2$ over and above what would have been expected with no treatment. $y1$ showed an increase/decrease ratio of (+115/-13), which means 89.84% of statistically significant countries showed an increase in total deaths per million associated with COVID-19 due directly to the causal impact of treatment initiation. $y2$ showed an increase/decrease ratio of (+105/-16) which means 86.78% of statistically significant countries showed an increase in total cases per million of COVID-19 due directly to the causal impact of treatment initiation. Causal impacts of the treatment on $y1$ ranges from -19% to +19015% with an average causal impact of +463.13%. Causal impacts of the treatment on $y2$ ranges from -46% to +12240% with an average causal impact of +260.88%. Hypothesis 1 Null can be rejected for a large majority of countries.

Causal Impact Results

The results of the *CausalImpact* package were produced as both figures and as automatic report summaries for each country. All figures are included as attached PDFs to this report and all figures can be requested in high quality 1920x1280 .png files from the author free of charge. All report summaries are included as .txt files in the supplementary data to this report.
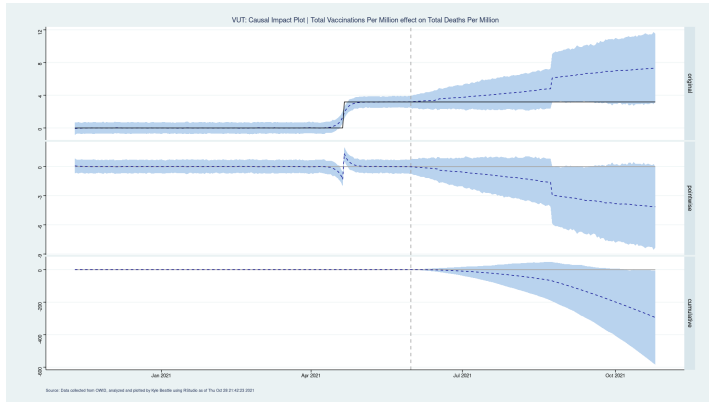
The following figures and report summaries represent a sample of more than 150 figures produced by this code and data. These examples highlight a variety of countries with decreases, average increases, and substantial increases in deaths and cases associated with COVID–19 as a causal impact of vaccine deployment. Included below select figures is also an example of the report summary that was produced by *CausalImpact*, these report summaries assist in explaining how to interpret each figure from a statistical perspective.

To read these figures one should analyze the three different graphs that are labeled on the right y axis as *Original*, *Pointwise*, and *Cumulative*. The *Original* graph represents the actual recorded data as the solid black line; the dashed blue line represents the counterfactual, the predicted trend line had the intervention of vaccine deployment not occurred; and the light blue fill around the counterfactual shows the degree of potential statistical variance, less light blue fill signifies a more accurate counterfactual. The moment of vaccine deployment varies between countries and is represented by the vertical gray dashed line. The *Pointwise* graph shows all of the positive or negative causal impacts by calculating the difference between the counterfactual and the recorded data. Finally, the *Cumulative* graph sums all of the positive or negative causal impacts since the intervention began to show an upward (positive), downward (negative) or neutral (near zero) causal impact.

γ1.E: Total Causal Impact from Vaccine Administration on Total Deaths Per Million

Average Decreases

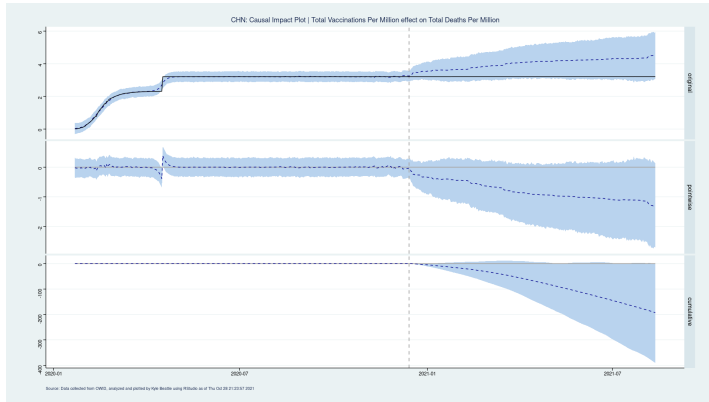**Vanuatu: –39% Vaccine Causal Impact on Total Deaths Per Million**

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 28 01:42:23 2021

Analysis report {*CausalImpact* (⊕Brodersen et al. 2015)}: During the post–intervention period, the response variable had an average value of approx. 3.18. By contrast, in the absence of an intervention, we would have expected an average response of 5.18. The 95% interval of this counterfactual prediction is [3.22, 7.16]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is –2.00 with a 95% interval of [–3.98, –0.038]. For a discussion of the significance of this effect, see below. Summing up the individual data points during the post–intervention period (which can only sometimes be meaningfully interpreted), the respons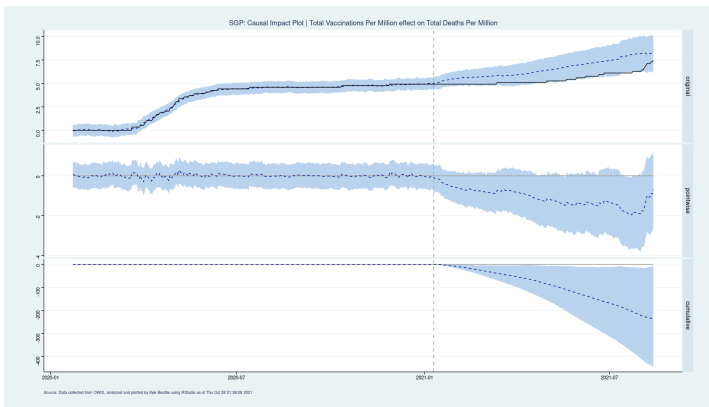e variable had an overall value of 467.46. By contrast, had the intervention not taken place, we would have expected a sum of 761.71. The 95% interval of this prediction is [473.06, 1051.98]. The above results are given in terms of absolute numbers. In relative terms, the response variable showed a decrease of –39%. The 95%

interval of this percentage is [–77%, –1%]. This means that the negative effect observed during the intervention period is statistically significant. If the experimenter had expected a positive effect, it is recommended to double-check whether anomalies in the control variables may have caused an overly optimistic expectation of what should have happened in the response variable in the absence of the intervention. The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability $p = 0.025$). This means the causal effect can be considered statistically significant.

China: –20% Vaccine Causal Impact on Total Deaths Per Million

CHN: Causal Impact Plot | Total Vaccinations Per Million effect on Total Deaths Per Million

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 28 21:23:57 2021

Analysis report {*CausalImpact* (⊕Brodersen et al. 2015)}: During the post–intervention period, the response variable had an average value of approx. 3.21. In the absence of an intervention, we would have expected an average response of 4.01. The 95% interval of this counterfactual prediction is [3.21, 4.83]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is –0.80 with a 95% interval of [–1.62, 0.0011]. For a discussion of the significance of this effect, see below. Summing up the individual data points during the post–intervention period (which can only sometimes be meaningfully interpreted), the response variable had an overall value of 773.57. Had the intervention not taken place, we would have expected a sum of 966.48. The 95% interval of this prediction is [773.30, 1164.21]. The above results are given in terms of absolute numbers. In relative terms, the response variable showed a decrease of –20%. The 95% interval of this percentage is [–40%, +0%]. This means that, although it may look as though the intervention has exerted a negative effect on the response variable when considering the intervention period as a whole, this effect is not statistically

significant, and so cannot be meaningfully interpreted. The apparent effect could be the result of random fluctuations that are unrelated to the intervention. This is often the case when the intervention period is very long and includes much of the time when the effect has already worn off. It can also be the case when the intervention period is too short to distinguish the signal from the noise. Finally, failing to find a significant effect can happen when there are not enough control variables or when these variables do not correlate well with the response variable during the learning period. The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability $p = 0.026$). This means the causal effect can be considered statistically significant.

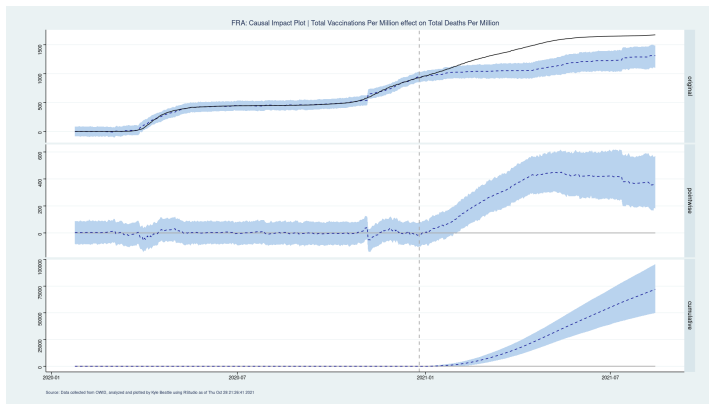**New Zealand: –19% Vaccine Causal Impact on Total Deaths Per Million**

NZL: Causal Impact Plot | Total Vaccinations Per Million effect on Total Deaths Per Million

Source: Data collected from OWID, analysed and plotted by Kier Beutler using RStudio as of Thu Oct 28 01:34:02 2021
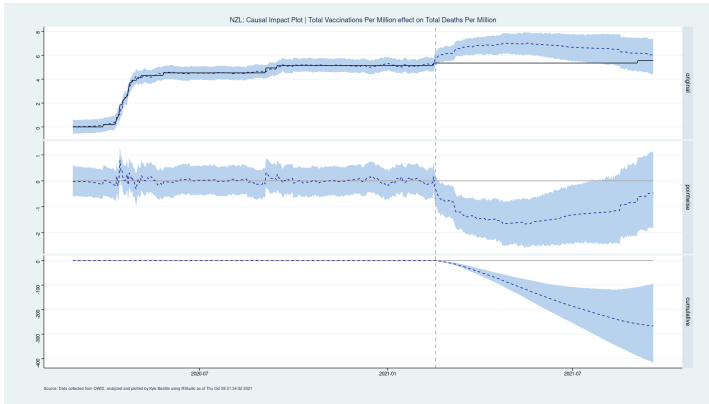
Singapore: –17% Vaccine Causal Impact on Total Deaths Per Million

SGP: Causal Impact Plot | Total Vaccinations Per Million effect on Total Deaths Per Million

Source: Data collected from OWID, analysed and plotted by Kier Beutler using RStudio as of Thu Oct 28 01:38:08 2021

Average Increases
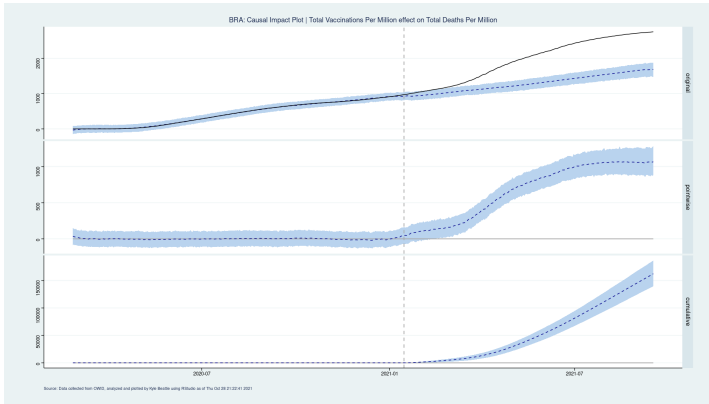
France: +28% Vaccine Causal Impact on Total Deaths Per Million

Analysis report {*CausalImpact* (⊕Brodersen et al. 2015)}: During the post–intervention period, the response variable had an average value of approx. 1.43K. By contrast, in the absence of an intervention, we would have expected an average response of 1.12K. The 95% interval of this counterfactual prediction is [1.02K, 1.22K]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is 0.31K with a 95% interval of [0.22K, 0.41K]. For a discussion of the significance of this effect, see below. Summing up the individual data points during the post–intervention period (which can only sometimes be meaningfully interpreted), the response variable had an overall value of 330.53K. By contrast, had the intervention not taken place, we would have expected a sum of 258.58K. The 95% interval of this prediction is [234.98K, 280.70K]. The above results are given in terms of absolute numbers. In relative terms, the response variable showed an increase of +28%. The 95% interval of this percentage is [+19%, +37%]. This means that the positive effect observed during the intervention period is statistically significant

and unlikely to be due to random fluctuations. It should be noted, however, that the question of whether this increase also bears substantive significance can only be answered by comparing the absolute effect (0.31K) to the original goal of the underlying intervention. The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability $p = 0.001$). This means the causal effect can be considered statistically significant.

Finland: +35% Vaccine Causal Impact on Total Deaths Per Million
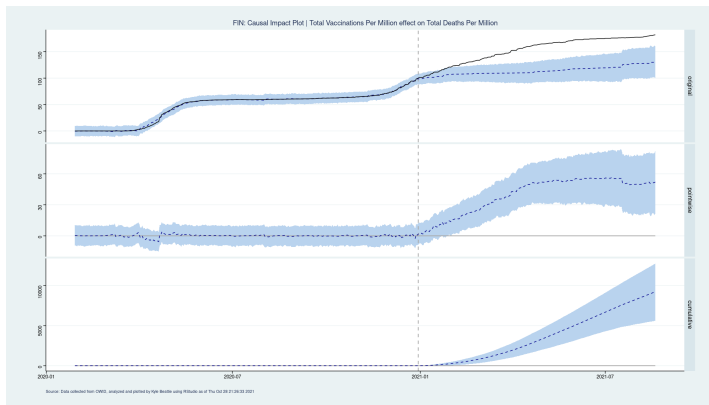
Analysis report {*CausalImpact* (⊕Brodersen et al. 2015)}: During the post–intervention period, the response variable had an average value of approx. 153.55. By contrast, in the absence of an intervention, we would have expected an average response of 113.69. The 95% interval of this counterfactual prediction is [98.50, 129.29]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is 39.85 with a 95% interval of [24.26, 55.05]. For a discussion of the significance of this effect, see below. Summing up the individual data points during the post–intervention period (which can only sometimes be meaningfully interpreted), the response variable had an overall value of 35.62K. By contrast, had the intervention not taken place, we would have expected a sum of 26.38K. The 95% interval of this prediction is [22.85K, 29.99K]. The above results are given in terms of absolute numbers. In relative terms, the response variable showed an increase of +35%. The 95% interval of this percentage is [+21%, +48%]. This means that the positive effect observed during the intervention period is statistically significant and unlikely to be due to random fluctuations. It should be

noted, however, that the question of whether this increase also bears substantive significance can only be answered by comparing the absolute effect (39.85) to the original goal of the underlying intervention. The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability $p = 0.001$). This means the causal effect can be considered statistically significant.
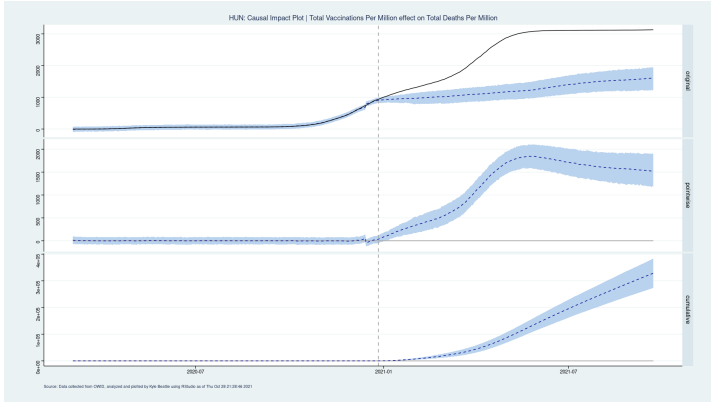
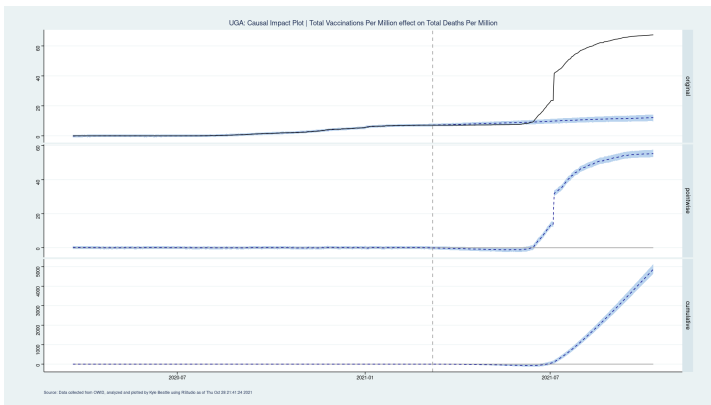## Brazil: +52% Vaccine Causal Impact on Total Deaths Per Million

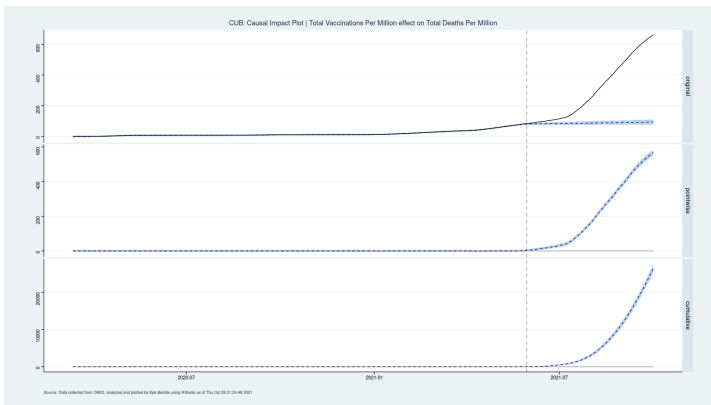## Lebanon: +74% Vaccine Causal Impact on Total Deaths Per Million
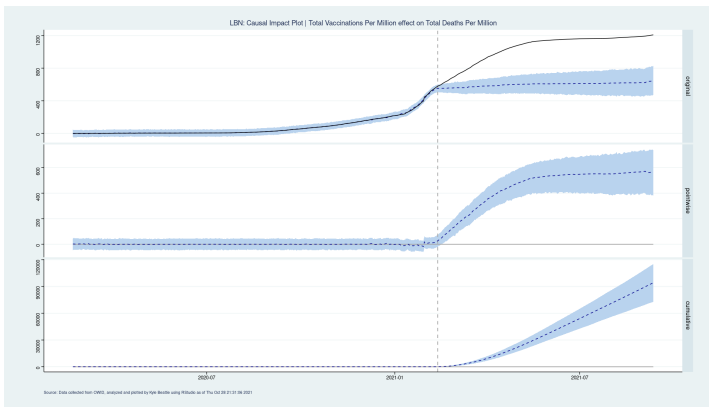
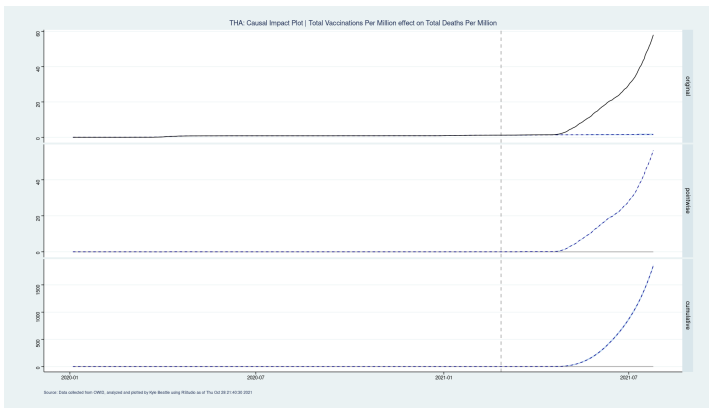**Hungary: +99% Vaccine Causal Impact on Total Deaths Per Million**

Uganda: +235% Vaccine Causal Impact on Total Deaths Per Million

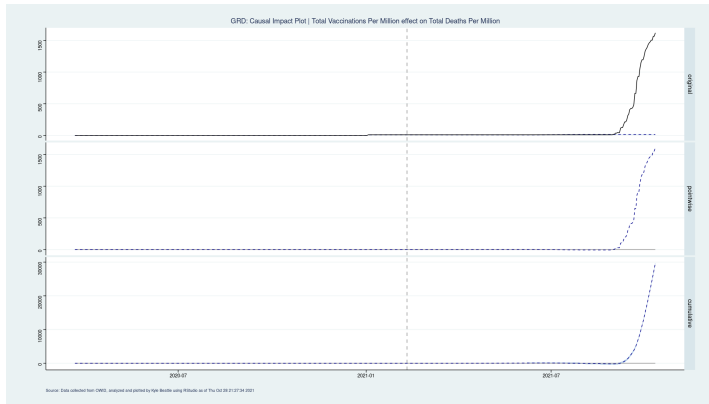#### Cuba: +245% Vaccine Causal Impact on Total Deaths Per Million

Thailand: +887% Vaccine Causal Impact on Total Deaths Per Million

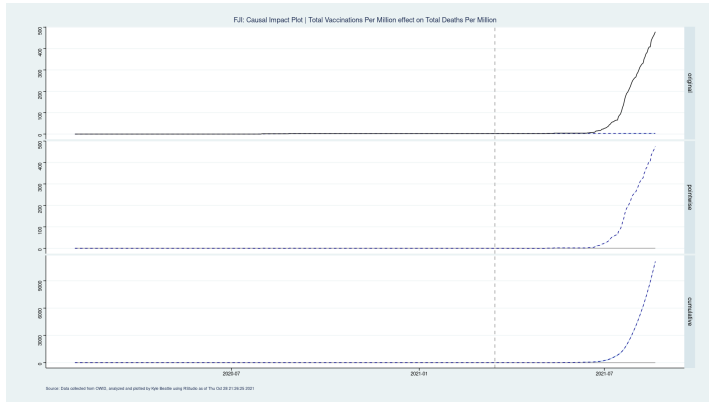THA: Causal Impact Plot | Total Vaccinations Per Million effect on Total Deaths Per Million

Source: Data collected from OWID, analyzed and plotted by Kjel Beattie using RStudio as of Thu Oct 08 21:40:26 2021

Elevated Increases

Grenada: +1180% Vaccine Causal Impact on Total Deaths Per Million

GRD: Causal Impact Plot | Total Vaccinations Per Million effect on Total Deaths Per Million

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 06 01:37:34 2021

Analysis report {*CausalImpact* (⊕Brodersen et al. 2015)}: During the post–intervention period, the response variable had an average value of approx. 132.32. By contrast, in the absence of an intervention, we would have expected an average response of 10.34. The 95% interval of this counterfactual prediction is [8.98, 11.85]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is 121.99 with a 95% interval of [120.47, 123.35]. For a discussion of the significance of this effect, see below. Summing up the individual data points during the post–intervention period (which can only sometimes be meaningfully interpreted), the response variable had an overall value of 32.15K. By contrast, had the intervention not taken place, we would have expected a sum of 2.51K. The 95% interval of this prediction is [2.18K, 2.88K]. The above results are given in terms of absolute numbers. In relative terms, the response variable showed an increase of +1180%. The 95% interval of this percentage is [+1166%, +1193%]. This means that the positive effect observed during the intervention period is statistically significant

and unlikely to be due to random fluctuations. It should be noted, however, that the question of whether this increase also bears substantive significance can only be answered by comparing the absolute effect (121.99) to the original goal of the underlying intervention. The probability of obtaining this effect by chance is very small (Bayesian one–sided tail–area probability $p = 0.001$). This means the causal effect can be considered statistically significant.
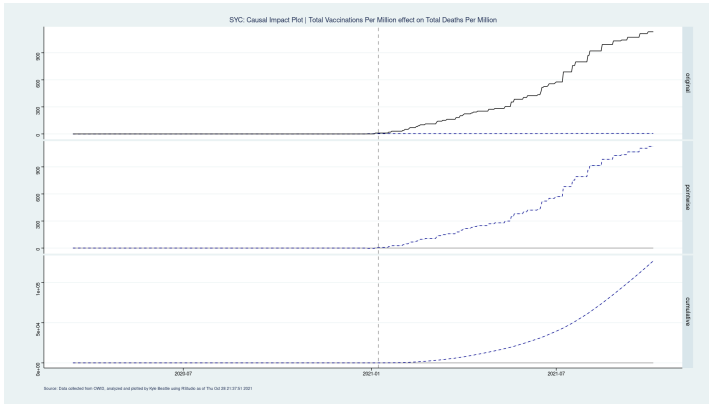
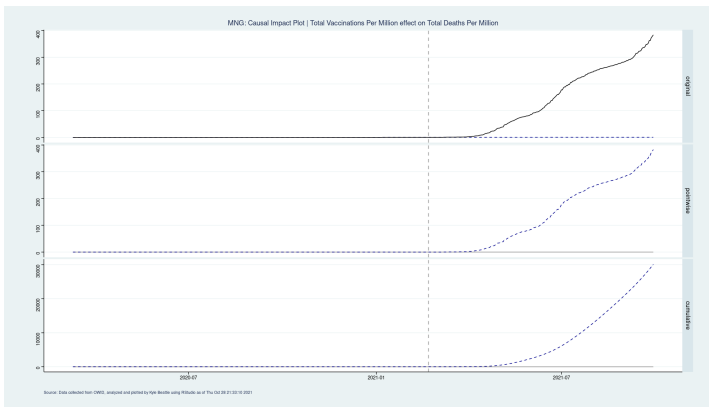# Fiji: +2499% Vaccine Causal Impact on Total Deaths Per Million

Fiji: Causal Impact Plot | Total Vaccinations Per Million effect on Total Deaths Per Million

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 28 21:09:25 2021

Analysis report {*CausalImpact* (⊕Brodersen et al. 2015)}: During the post–intervention period, the response variable had an average value of approx. 73.75. By contrast, in the absence of an intervention, we would have expected an average response of 2.84. The 95% interval of this counterfactual prediction is [2.19, 3.56]. Subtracting this prediction from the observed response yields an estimate of the causal effect the intervention had on the response variable. This effect is 70.92 with a 95% interval of [70.20, 71.56]. For a discussion of the significance of this effect, see below. Summing up the individual data points during the post–intervention period (which can only sometimes be meaningfully interpreted), the response variable had an overall value of 11.58K. By contrast, had the intervention not taken place, we would have expected a sum of 0.45K. The 95% interval of this prediction is [0.34K, 0.56K]. The above results are given in terms of absolute numbers. In relative terms, the response variable showed an increase of +2499%. The 95% interval of this percentage is [+2473%, +2522%]. This means that the positive effect observed during the intervention period is statistically significant and unlikely to be due to random fluctuations. It

should be noted, however, that the question of whether this increase also bears substantive significance can only be answered by comparing the absolute effect (70.92) to the original goal of the underlying intervention. The probability of obtaining this effect by chance is very small (Bayesian one-sided tail-area probability $p = 0.001$). This means the causal effect can be considered statistically significant.

Seychelles: +10680% Vaccine Causal Impact on Total Deaths Per Million

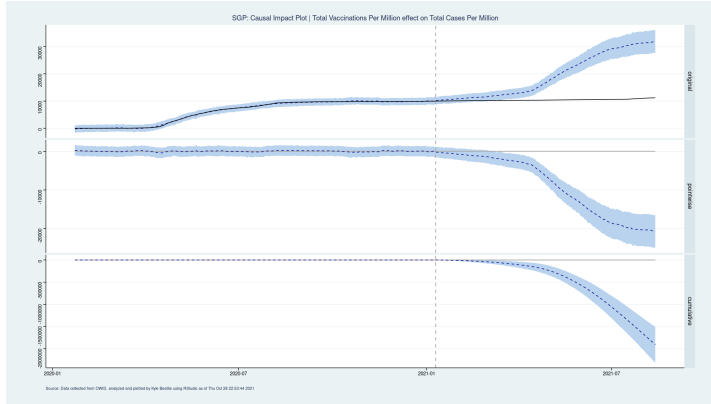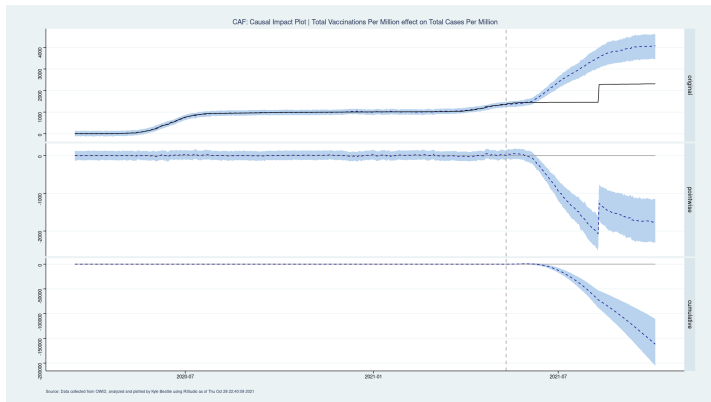## Mongolia: +19015% Vaccine Causal Impact on Total Deaths Per Million

γ2.E: Total Causal Impact from Vaccine Administration on Total Cases Per Million
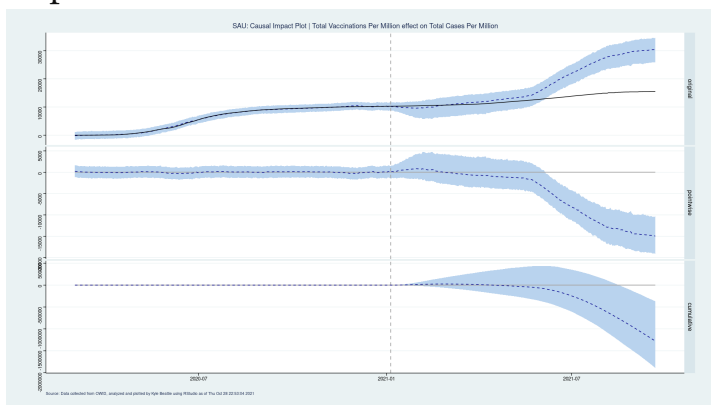
Average Decreases

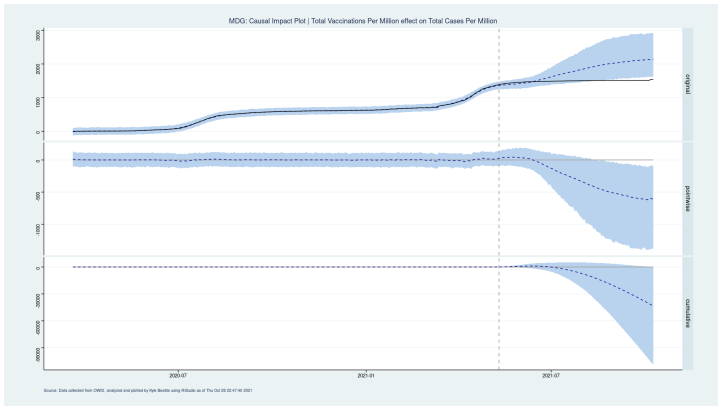### Singapore: –46% Vaccine Causal Impact on Total Cases Per Million

Central African Republic: –38% Vaccine Causal Impact on Total Cases Per Million

Saudi Arabia: –28% Vaccine Causal Impact on Total Cases Per Million

Madagascar: –16% Vaccine Causal Impact on Total Cases Per Million

MDG: Causal Impact Plot | Total Vaccinations Per Million effect on Total Cases Per Million

Average Increases

Russia: +22% Vaccine Causal Impact on Total Cases Per Million

United States: +38% Vaccine Causal Impact on Total Cases Per Million

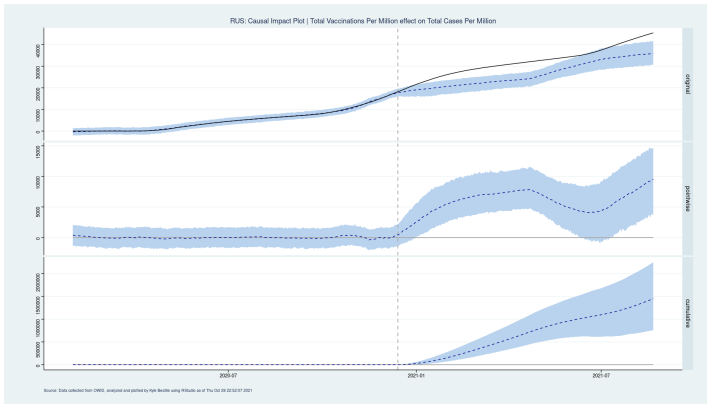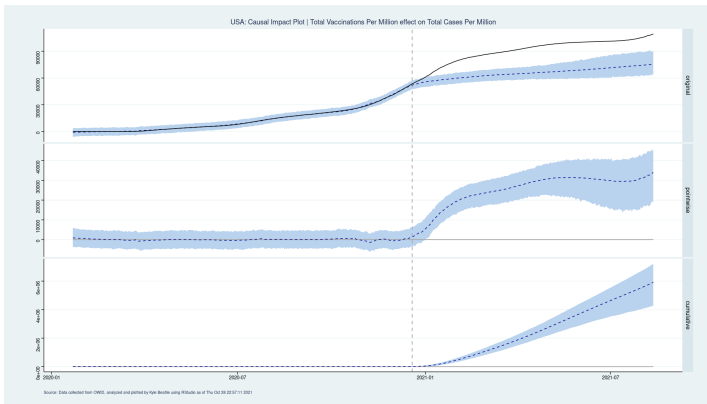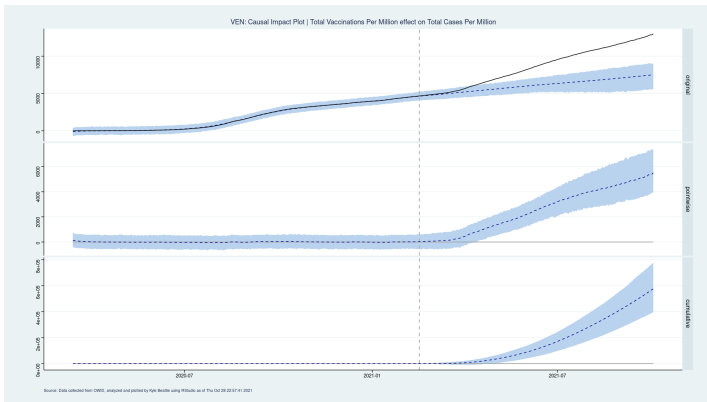#### Venezuela: +41% Vaccine Causal Impact on Total Cases Per Million

India: +74% Vaccine Causal Impact on Total Cases Per Million

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 28 22:45:12 2021

Phillipines: +101% Vaccine Causal Impact on Total Cases Per Million

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 28 22:51:01 2021

Sri Lanka: +170% Vaccine Causal Impact on Total Cases Per Million

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 28 22:54:47 2021

Grenada: +309% Vaccine Causal Impact on Total Cases Per Million

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 28 22:44:01 2021

Uruguay: +390% Vaccine Causal Impact on Total Cases Per Million

Taiwan: +475% Vaccine Causal Impact on Total Cases Per Million
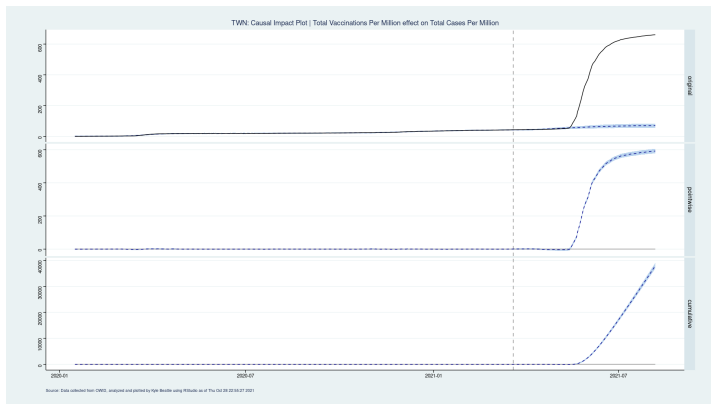
Elevated Increases

Timor-Leste: +839% Vaccine Causal Impact on Total Cases Per Million
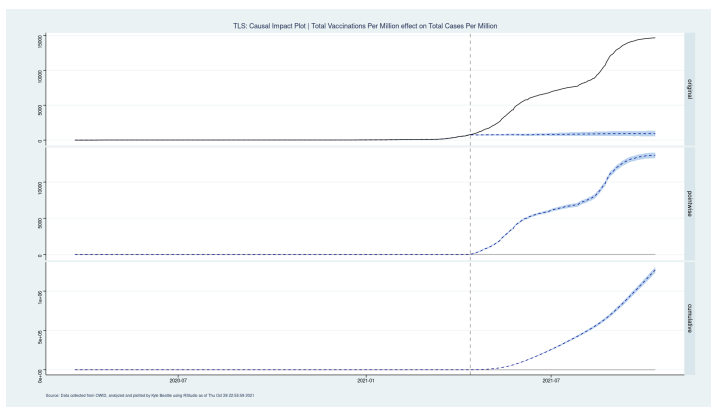
Vietnam: +1099% Vaccine Causal Impact on Total Cases Per Million
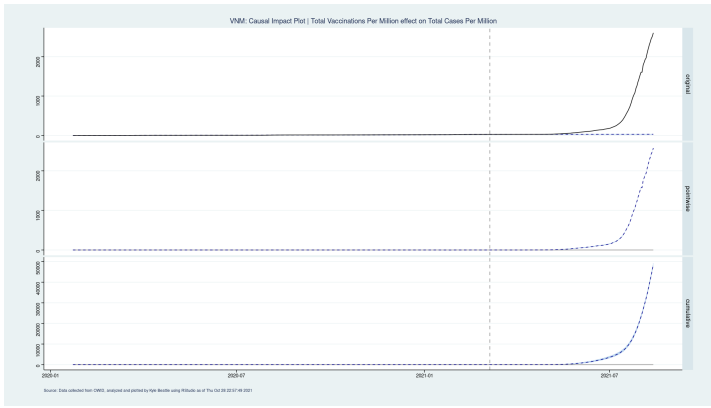
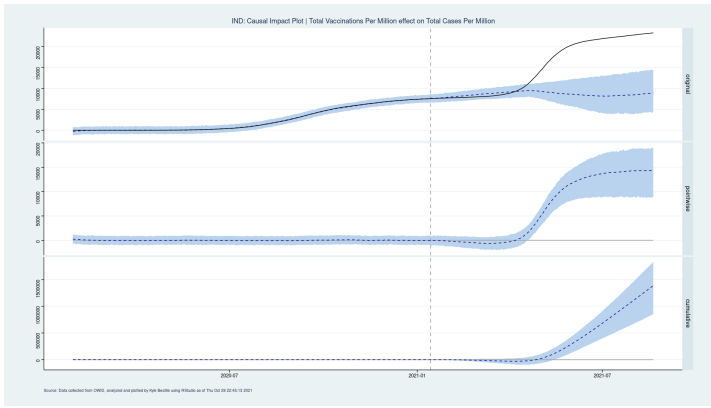Seychelles: +1978% Vaccine Causal Impact on Total Cases Per Million
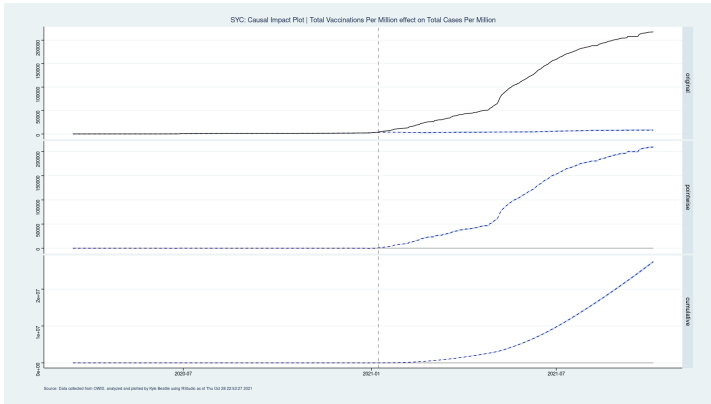
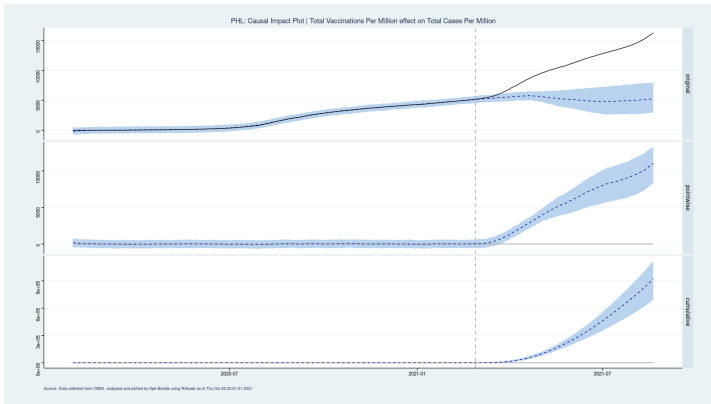

Mongolia: +3391% Vaccine Causal Impact on Total Cases Per Million
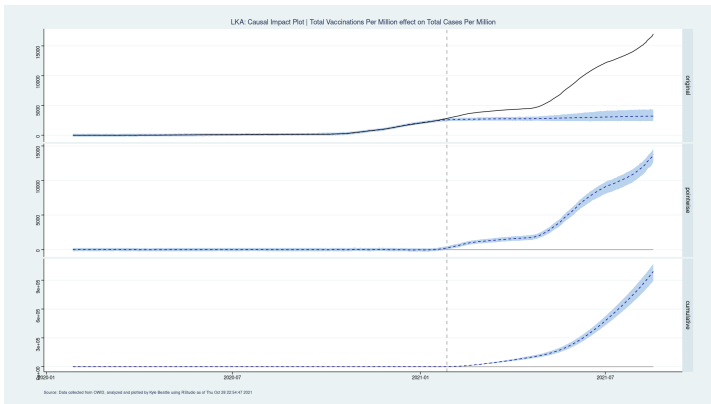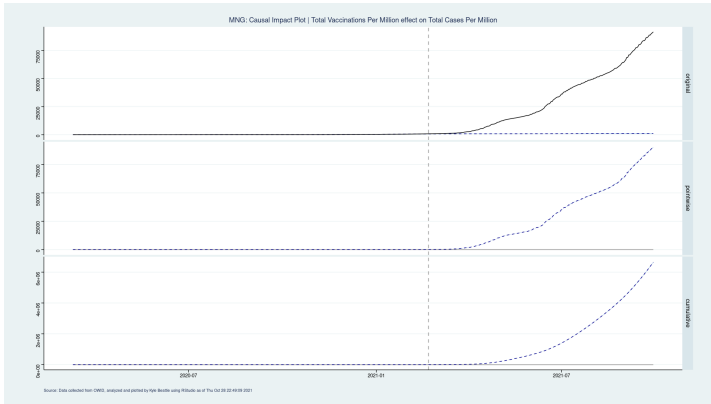

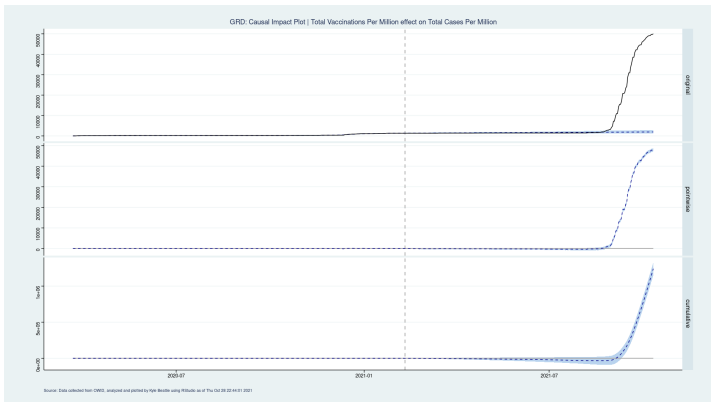Laos: +6955% Vaccine Causal Impact on Total Cases Per Million

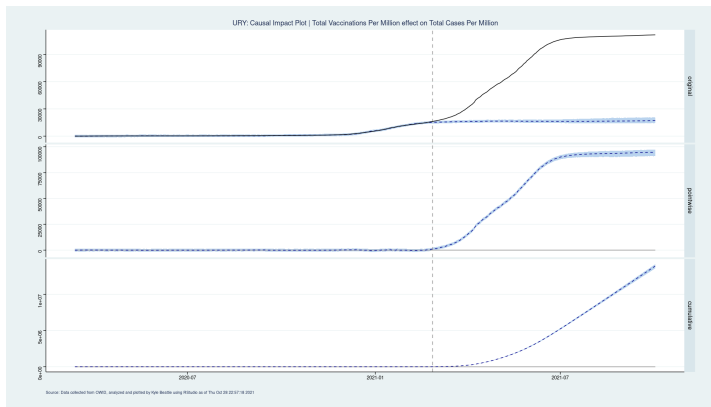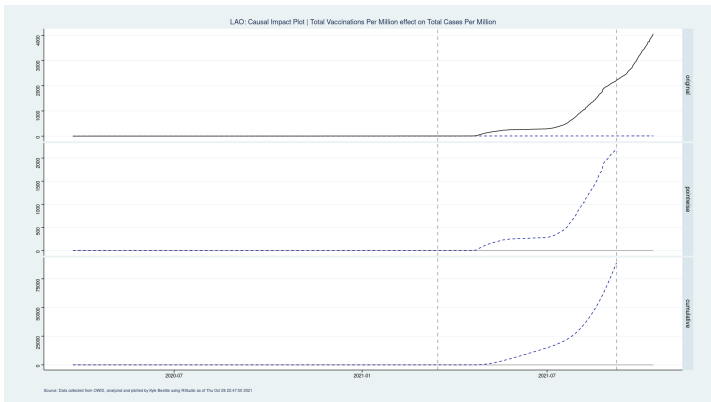#### Fiji: +12240% Vaccine Causal Impact on Total Cases Per Million

Fiji: Causal Impact Plot | Total Vaccinations Per Million effect on Total Cases Per Million

Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Thu Oct 28 22:42:56 2021
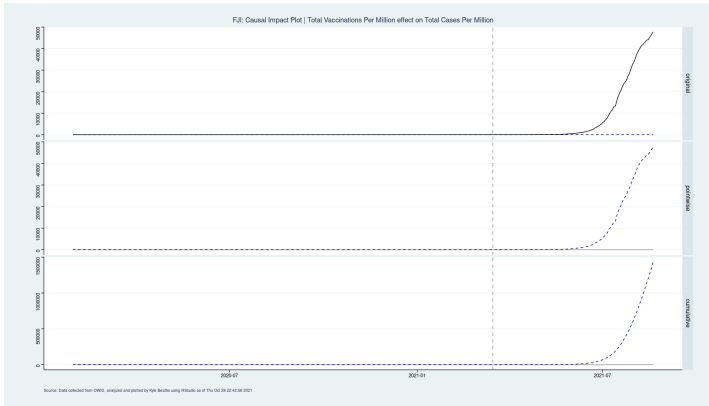
Density Plots

The following density plots and tables present a larger view of the results. The density plots present data for each continent for all countries with results up to +500%.

Density Plot 1: Effect of Vaccines on Total Deaths Per Million grouped by Continent

Density Plot 2: Effect of Vaccines on Total Cases Per Million grouped by Continent

Density Plot: Effect of Vaccines on Total Cases Per Million grouped by Continent

Quantiles are represented by the blue horizontal lines on each curve.
This data shows % Causal Impact of Vaccines up to +500%.

% Causal Impact of Vaccines on Total Cases Per Million
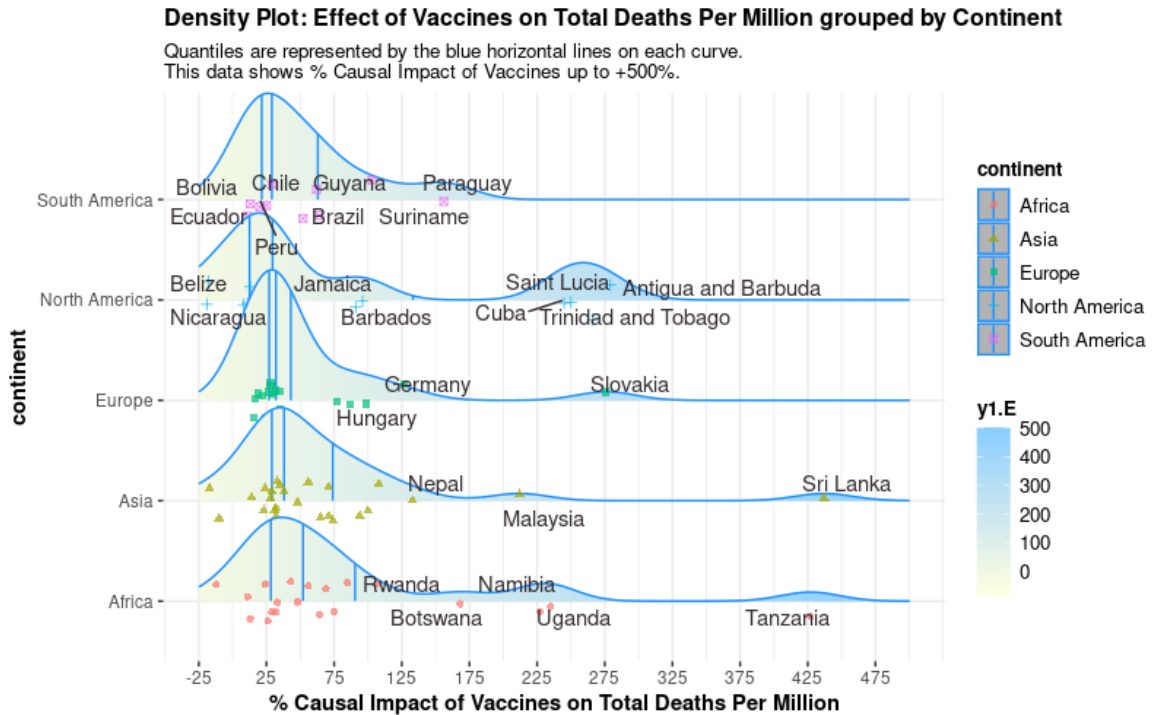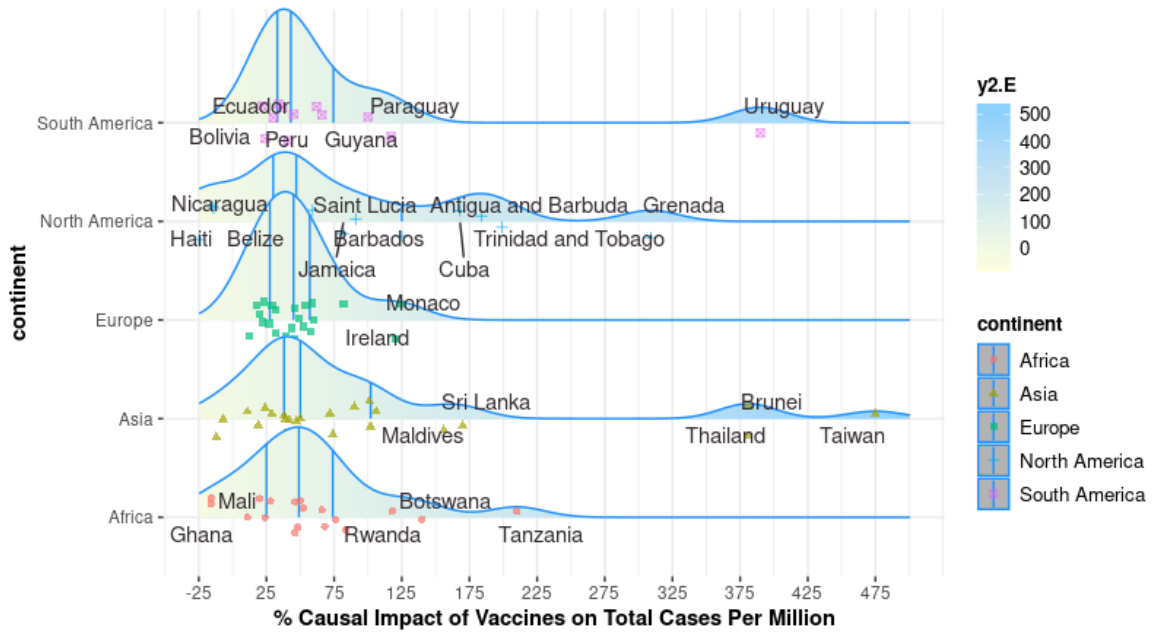
Source: Data collected from OWID, analyzed and plotted by Kyle Beattie using RStudio as of Sat Oct 30 11:24:08 2021

Tables

The two tables here represent all the countries with statistically significant results, they show the percentage change in $y1$ and $y2$ as a direct causal impact of vaccine deployment.

Causal Impact of Vaccine Deployment on Total Deaths per Million associated with COVID-19

X ISO	y1.p	y1.effect	y1.effect_dec
1 AFG	0.001	+32%	0.32
2 AGO	0.001	+29%	0.29
3 ALB	0.001	+35%	0.35
4 ARE	0.001	+71%	0.71
5 ARG	0.001	+23%	0.23
6 ATG	0.001	+279%	2.79
7 AUS	0.007	–22%	–0.22
8 AUT	0.005	+26%	0.26
9 AZE	0.004	+17%	0.17
10 BGD	0.001	+33%	0.33
11 BHS	0.019	+13%	0.13
12 BIH	0.001	+30%	0.30
13 BLR	0.001	+43%	0.43
14 BLZ	0.009	–19%	–0.19
15 BOL	0.027	+13%	0.13
16 BRA	0.001	+52%	0.52
17 BRB	0.001	+96%	0.96
18 BRN	0.001	+65%	0.65
19 BTN	0.033	+35%	0.35
20 BWA	0.001	+168%	1.68
21 CAF	0.046	–12%	–0.12
22 CAN	0.001	+31%	0.31
23 CHL	0.001	+29%	0.29
24 CHN	0.026	–20%	–0.20
25 CIV	0.001	+43%	0.43
26 CMR	0.001	+25%	0.25
27 COL	0.001	+25%	0.25
28 CPV	0.001	+32%	0.32
29 CUB	0.001	+245%	2.45
30 CYP	0.001	+87%	0.87
31 DEU	0.001	+127%	1.27
32 DJI	0.001	+27%	0.27
33 DZA	0.007	–7%	–0.07
34 ECU	0.030	+11%	0.11

X ISO	y1.p	y1.effect	y1.effect_dec
35 EGY	0.001	+20%	0.20
36 ESP	0.001	+16%	0.16
37 ETH	0.001	+19%	0.19
38 FIN	0.001	+35%	0.35
39 FJI	0.001	+2499%	24.99
40 FRA	0.001	+28%	0.28
41 GAB	0.001	+26%	0.26
42 GBR	0.001	+35%	0.35
43 GEO	0.001	+23%	0.23
44 GHA	0.001	+26%	0.26
45 GIN	0.001	+56%	0.56
46 GNB	0.003	+13%	0.13
47 GRD	0.001	+1180%	11.80
48 GTM	0.044	+8%	0.08
49 GUY	0.001	+64%	0.64
50 HKG	0.030	−13%	−0.13
51 HND	0.001	+26%	0.26
52 HRV	0.001	+43%	0.43
53 HTI	0.002	+11%	0.11
54 HUN	0.001	+99%	0.99
55 IDN	0.001	+100%	1.00
56 IND	0.001	+29%	0.29
57 IRL	0.001	+33%	0.33
58 ITA	0.001	+24%	0.24
59 JAM	0.001	+91%	0.91
60 JOR	0.001	+56%	0.56
61 JPN	0.001	+48%	0.48
62 KAZ	0.001	+94%	0.94
63 KEN	0.001	+33%	0.33
64 KWT	0.001	+24%	0.24
65 LBN	0.001	+74%	0.74
66 LBR	0.001	+75%	0.75
67 LCA	0.001	+250%	2.50
68 LKA	0.001	+437%	4.37
69 MAR	0.047	−12%	−0.12
70 MCO	0.001	+741%	7.41
71 MDA	0.001	+17%	0.17
72 MDG	0.021	+11%	0.11
73 MDV	0.001	+108%	1.08
74 MEX	0.001	+28%	0.28
75 MKD	0.001	+28%	0.28
76 MLI	0.018	+13%	0.13
77 MLT	0.005	+19%	0.19
78 MMR	0.001	+71%	0.71
79 MNE	0.001	+30%	0.30
80 MNG	0.001	+19015%	190.15

X ISO	y1.p	y1.effect	y1.effect_dec
81 MOZ	0.001	+64%	0.64
82 MUS	0.001	+69%	0.69
83 MWI	0.003	+22%	0.22
84 MYS	0.001	+212%	2.12
85 NAM	0.001	+227%	2.27
86 NER	0.029	−12%	−0.12
87 NIC	0.005	−18%	−0.18
88 NPL	0.001	+133%	1.33
89 NZL	0.002	−19%	−0.19
90 OMN	0.001	+32%	0.32
91 PAK	0.001	+28%	0.28
92 PER	0.001	+20%	0.20
93 PHL	0.001	+38%	0.38
94 PNG	0.001	+263%	2.63
95 PRY	0.001	+156%	1.56
96 PSE	0.004	+14%	0.14
97 ROU	0.001	+34%	0.34
98 RUS	0.001	+77%	0.77
99 RWA	0.001	+107%	1.07
100 SAU	0.018	−10%	−0.10
101 SEN	0.001	+43%	0.43
102 SGP	0.025	−17%	−0.17
103 SOM	0.005	+24%	0.24
104 SRB	0.001	+32%	0.32
105 SUR	0.001	+103%	1.03
106 SVK	0.001	+276%	2.76
107 SVN	0.001	+27%	0.27
108 SWE	0.001	+22%	0.22
109 SYC	0.001	+10680%	106.80
110 SYR	0.001	+31%	0.31
111 TGO	0.001	+24%	0.24
112 THA	0.001	+887%	8.87
113 TLS	0.002	+2356%	23.56
114 TTO	0.001	+266%	2.66
115 TUN	0.001	+56%	0.56
116 TUR	0.001	+32%	0.32
117 TWN	0.001	+2767%	27.67
118 TZA	0.001	+427%	4.27
119 UGA	0.001	+235%	2.35
120 UKR	0.001	+43%	0.43
121 URY	0.001	+507%	5.07
122 USA	0.001	+31%	0.31
123 VCT	0.013	+26%	0.26
124 VEN	0.001	+62%	0.62
125 VNM	0.001	+707%	7.07
126 VUT	0.025	−39%	−0.39

X ISO	y1.p	y1.effect	y1.effect_dec
127 ZMB	0.001	+85%	0.85
128 ZWE	0.001	+48%	0.48

Causal Impact of Vaccine Deployment on Total Cases per Million associated with COVID-19

X ISO	y2.p	y2.effect	y2.effect_dec
1 AFG	0.006	–12%	–0.12
2 AGO	0.001	+52%	0.52
3 ALB	0.001	+54%	0.54
4 ARE	0.001	+72%	0.72
5 ARG	0.001	+66%	0.66
6 ATG	0.001	+184%	1.84
7 BGD	0.014	+11%	0.11
8 BHS	0.001	+47%	0.47
9 BIH	0.047	+18%	0.18
10 BLZ	0.042	–14%	–0.14
11 BOL	0.001	+22%	0.22
12 BRA	0.001	+37%	0.37
13 BRB	0.001	+91%	0.91
14 BRN	0.001	+381%	3.81
15 BTN	0.001	+90%	0.90
16 BWA	0.001	+140%	1.40
17 CAF	0.001	–38%	–0.38
18 CAN	0.001	+59%	0.59
19 CHL	0.001	+45%	0.45
20 COG	0.018	–11%	–0.11
21 COL	0.001	+34%	0.34
22 CPV	0.001	+46%	0.46
23 CUB	0.001	+168%	1.68
24 CYP	0.001	+82%	0.82
25 DEU	0.001	+52%	0.52
26 DMA	0.001	+407%	4.07
27 ECU	0.001	+24%	0.24
28 ESP	0.001	+44%	0.44
29 FIN	0.001	+59%	0.59
30 FJI	0.001	+12240%	122.40
31 FRA	0.001	+49%	0.49
32 GBR	0.001	+46%	0.46
33 GEO	0.003	+24%	0.24
34 GHA	0.002	–16%	–0.16
35 GNQ	0.001	–23%	–0.23
36 GRD	0.001	+309%	3.09
37 GTM	0.001	+30%	0.30
38 GUY	0.001	+100%	1.00
39 HND	0.001	+19%	0.19

X ISO	y2.p	y2.effect	y2.effect_dec
40 HTI	0.001	–15%	–0.15
41 HUN	0.005	+27%	0.27
42 IDN	0.001	+106%	1.06
43 IND	0.001	+74%	0.74
44 IRL	0.001	+120%	1.20
45 IRN	0.001	+64%	0.64
46 IRQ	0.001	+51%	0.51
47 ITA	0.001	+39%	0.39
48 JAM	0.001	+82%	0.82
49 JOR	0.001	+53%	0.53
50 JPN	0.001	+45%	0.45
51 KAZ	0.001	+38%	0.38
52 KEN	0.001	+28%	0.28
53 KHM	0.001	+5808%	58.08
54 KNA	0.001	+1051%	10.51
55 KOR	0.001	+49%	0.49
56 KWT	0.001	+50%	0.50
57 LAO	0.001	+6955%	69.55
58 LBN	0.001	+47%	0.47
59 LBR	0.001	+20%	0.20
60 LBY	0.001	+31%	0.31
61 LCA	0.001	+125%	1.25
62 LKA	0.001	+170%	1.70
63 MCO	0.001	+124%	1.24
64 MDA	0.007	+12%	0.12
65 MDG	0.024	–16%	–0.16
66 MDV	0.001	+156%	1.56
67 MEX	0.001	+41%	0.41
68 MKD	0.003	+23%	0.23
69 MLI	0.027	+11%	0.11
70 MLT	0.001	+46%	0.46
71 MMR	0.037	+39%	0.39
72 MNE	0.001	+20%	0.20
73 MNG	0.001	+3391%	33.91
74 MOZ	0.001	+46%	0.46
75 MRT	0.001	–22%	–0.22
76 MUS	0.001	+536%	5.36
77 MYS	0.001	+102%	1.02
78 NAM	0.001	+84%	0.84
79 NGA	0.024	–15%	–0.15
80 NIC	0.001	–25%	–0.25
81 NLD	0.001	+31%	0.31
82 NPL	0.001	+39%	0.39
83 OMN	0.027	+42%	0.42
84 PAK	0.029	-7%	–0.07
85 PER	0.001	+30%	0.30

X ISO	y2.p	y2.effect	y2.effect_dec
86 PHL	0.001	+101%	1.01
87 PNG	0.001	+191%	1.91
88 PRY	0.001	+117%	1.17
89 PSE	0.002	+19%	0.19
90 RUS	0.001	+22%	0.22
91 RWA	0.001	+118%	1.18
92 SAU	0.003	–28%	–0.28
93 SDN	0.001	–29%	–0.29
94 SEN	0.001	+24%	0.24
95 SGP	0.001	–46%	–0.46
96 SLE	0.017	–10%	–0.10
97 SMR	0.001	+25%	0.25
98 SRB	0.001	+32%	0.32
99 SUR	0.001	+62%	0.62
100 SVK	0.001	+29%	0.29
101 SVN	0.001	+58%	0.58
102 SWE	0.001	+60%	0.60
103 SWZ	0.001	+36%	0.36
104 SYC	0.001	+1978%	19.78
105 SYR	0.001	+29%	0.29
106 TGO	0.001	+76%	0.76
107 THA	0.001	+381%	3.81
108 TLS	0.001	+839%	8.39
109 TTO	0.001	+199%	1.99
110 TUN	0.001	+68%	0.68
111 TWN	0.001	+475%	4.75
112 TZA	0.001	+210%	2.10
113 UGA	0.001	+48%	0.48
114 UKR	0.001	+32%	0.32
115 URY	0.001	+390%	3.90
116 USA	0.001	+38%	0.38
117 VCT	0.001	+35%	0.35
118 VEN	0.001	+41%	0.41
119 VNM	0.001	+1099%	10.99
120 ZMB	0.001	+66%	0.66
121 ZWE	0.001	+50%	0.50

Categorical Variables - Dot plots of variables $y1.E$, $y2.E$, $c1$, $c2$, and $c3$

To test and display categorical variables, dot plots were created using the R package *ggstatsplot* (⊕Patil 2021). These plots demonstrate the distribution of vaccine effect across *continent*, *country* and *vaccine brands in use in each country*. As

this data is non–parametric, these charts also present the Wilcoxon signed rank value (V) to demonstrate whether the differences between groups is different from the null of 0. This demonstrates whether different continents, countries, or vaccine brands show different levels of $y1.E$ or $y2.E$ or whether the effect of vaccine administration was uniform.

Numerical Variables - Correlation Matrix and Scatterplot of $y1.E$, $y2.E$, $n1$, $n2$, and $n3$

As this data is non–parametric, all significant correlation calculations are presented with Spearman's ρ (*rho*). This correlation matrix shows how all numerical variables relate to one another. The stronger the correlation, the bluer the box appears, if there is an X over the number that means it has a p-value > 0.05 and is not statistically significant. The only statistically significant result of importance for this study is the correlation between *total vaccinations per hundred people* and *effect of vaccine intervention on total cases per million* (ρ: 0.3384496, p: 9.7^{-4}).

Correlation Matrix

R Graphics Output   1 / 1   —   99%   +

Spearman Scatterplot of y2.E ~ n2

The scatter plot shown here provides the following detailed information for the only statistically significant correlation among independent numerical variables:

- Correlation coefficient (r) = The strength of the relationship.

- p-value = The significance of the relationship.

- Histogram with kernel density estimation and rug plot.

- Scatter plot with fitted line.

Discussion

Increase in Death and Cases as a Causal Impact of Vaccine Administration

Countries with few COVID-19 deaths in the year 2020 appear to have fared the worst of all countries after vaccine administration (e.g Thailand, Vietnam, Mongolia, Taiwan, Seychelles, Cambodia, etc.). The causal impact results from vaccine administration seen in these countries of hundreds or thousands of percentage increases in total deaths and cases per million are also the causal impact results we can be most statistically confident in due to the direct increase of COVID-19 associated deaths and cases after vaccine administration, where prior to vaccine administration there were few or none. Notably, the results we can be least statistically confident about are many of the results suggesting a negative causal impact from vaccine administration (e.g. Saudi Arabia, China, Nigeria, Belize, etc.).

Some might try to argue that these results indicate a rise in cases and deaths associated with COVID-19 of those who have not taken COVID-19 experimental gene therapy injections, or that perhaps these deaths are as a result of some new more contagious variant such as Delta. As to the first point, while this data is still inconclusive, especially on a worldwide scale as was this study's focus, there is beginning to emerge a pattern of a similar amount of cases and deaths of COVID-19 in relation to the population that is vaccinated as is evidenced by public records from Public Health England and the Israeli Ministry of Health. In addition, if that

counterargument were true, we would expect to see countries with higher vaccination rates also have lower (or negative) impacts from vaccine administration on rates of cases and deaths associated with COVID-19. Instead, we see the opposite, a low positive correlation (ρ: 0.34, $p < 0.001$) between *total vaccinations per hundred* and *the impact of vaccine administration on cases associated with COVID-19*. These results concur with the fact that the vaccines only offer a low absolute risk reduction (ARR) (0.8–1.9%) (⊕Olliaro, Torreele, and Vaillant 2021) in the first place and have been shown to wane over time to an even lower ARR (⊕Levin et al. 2021; ⊕Chemaitelly et al. 2021; ⊕Wang et al. 2021).

To the latter point, the calculations in the *CausalImpact* package and the code presented above accounted for the differing dates when the vaccine administration started in each country and it is not likely that the Delta variant arrived in each of these countries precisely at the time each vaccine administration also started. Rather, it is more likely that the vaccine administration causes a bottleneck effect in each region and helps to create even more deadly variants as ⊕Ausschuss et al. (2021), ⊕Bossche (2021), and ⊕Ricke and Malone (2020) all warned, which may translate into a rise in cases and deaths associated with COVID-19 as a result of the causal impact of vaccine administration.

Farr's Law, Gompertz Function

As noted by other authors (⊕Pacheco-Barrios et al. 2020), many countries that reported high cases and deaths associated with COVID-19 during 2020 and early 2021 showed a standard example

of Farr's Law and/or the Gompertz Function when viewed cumulatively per capita as an inverse Gompertz function (⊕Haynes and Kulkarni 2021) or when predicted as a "best straight line" (⊕Levitt, Scaiewicz, and Zonta 2020). In other words, these countries appeared to have largely achieved natural immunity by late spring of 2021 in the Northern Hemisphere, which is why many of their trend lines go flat for a time. Unfortunately, once the vaccine administration started for the general population, or shortly thereafter, those trend lines began to increase again in many countries, and unnaturally so in the middle or towards the end of summer in the Northern hemisphere or in countries where previously there had been very few if any cases or deaths. Normally, seasonal die off from pneumonia, influenza, or COVID-19 is in the winter, so this spike that appears in many countries after vaccine administration at this time of the year or in countries with no previous outbreaks is highly suspect as not being a natural trend, but rather vaccine-induced.

These results are consistent with the waning effect of COVID-19 vaccine protection mentioned above, the amount of "breakthrough" cases we are currently witnessing (⊕Musser et al. 2021), and the overwhelming and historically unprecedented quantity of reports of vaccine adverse events in the Vaccine Adverse Events Reporting System (VAERS), over 16,000 reported deaths as of writing this report (⊕Health and Services 2021), suggesting a highly untested vaccine. At the same time, the robust, durable, and long-lasting natural immunity that occurs with infection from SARS-CoV or SARS-CoV-2 (⊕Majdoubi et al.

2021) combined with extremely low absolute risk reduction (0.8-1.9%) (⊕Olliaro, Torreele, and Vaillant 2021) from available vaccines, make the risks of these vaccines likely outweigh the benefits for most if not all of the population.

At the very least, these results suggest that COVID-19 vaccine administration as a public policy over 80% of the time does not have a statistically significant causal impact of lowering total deaths or cases per million, but rather a statistically significant impact in increasing total deaths or cases per million associated with COVID-19 over and above what would have been expected if no vaccines were ever administered.

Obviously, the results here will be shocking to many who have perhaps been paying more attention to official government/media narrative rather than on the ground data and evidence. However, for those that have been paying attention to the pleas, warnings, and publicly voiced frustrations of many of the brightest and bravest minds in the scientific, medical and investigative community (e.g. Dr. Luc Montagnier (⊕Ausschuss et al. 2021), Dr. Michael Yeadon (⊕Yeadon 2020; ⊕Borger et al. 2021), Dr. Robert Malone (⊕Ricke and Malone 2020), Dr. Wolfgang Woodard (⊕Wodarg and Scarlattilaan 2020), Dr. Geert Vanden Bossche (⊕Bossche 2021), Dr. Peter McCullough (⊕Bruno et al. 2021), and over 860,000 others (⊕Kulldorff, Gupta, and Bhattacharya 2020; ⊕Kampf and Kulldorff 2021)) or have been studying the raw data themselves, these results likely come as no surprise.

Vaccine Mandates

If they ever were ethical, which this author disagrees with strongly based on scientific, medical, and research ethics that regard the patient rights to informed consent and non-prejudcial refusal of treatment or experimentation as iterated and agreed to under the Nuremberg Code (⊕Code 1998), Helsinki Accords (⊕Association and others 2009), and the Human Rights Declaration on Bioethics (⊕UNESCO 2019) as inalienable, essential, and non-negotiable, vaccine mandates under these conditions and results are beyond unethical at this point, they are clearly discriminatory and likely criminal, a determination courts and lawyers will ultimately decide.

The results of this study taken together demonstrate a product that directly causes more COVID-19 associated cases and deaths than otherwise would have existed with zero vaccines. Consequently, these experimental gene therapy injections known as COVID-19 vaccines cannot be mandated by any public policy that intends to continue following the regulations of the Nuremberg Code (⊕Code 1998), the Helsinki Accords (⊕Association and others 2009), and the Human Rights Declaration on Bioethics (⊕UNESCO 2019).

Limitations

There exist limitations with this study including: (1) potentially inaccurate publicly recorded data (2) the causal impact measures only the vaccine effect on COVID-19 related cases and death and not all-cause mortality (3) data for COVID-19

related cases and death largely do a poor job differentiating between vaccinated and non-vaccinated individuals (4) the predictor sets can be debated over (5) a lack of pre-intervention data for some countries prevents analysis of more countries.

As to the first limitation, whether the public data being provided by certain countries is 100% accurate is a matter for another study. Here it is important to use the public data available because this is the data that the world is using to determine COVID-19 government policies. This is the data that is quoted daily by government officials, journalists, and scientists alike. The results presented here are based on that data in its unadulterated form.

Regarding the second limitation, future research will utilize the methods formulated in this study to analyze all-cause mortality or specific-cause mortality data, making this R script highly versatile for future research.

Third, the publicly available data for COVID-19 related cases and death does a poor job at differentiating between vaccinated and non-vaccinated individuals making it an important category to study within this context, however one that should be performed on a smaller-scale case study if valid data should be made available. This study, while recognizing the urgent importance of the outcomes for this variable, could not perform large-scale data analysis with data that is so sparse and ill-defined worldwide. Unfortunately, much of the data surrounding this variable is highly politicized at this moment

in time and is therefore difficult to obtain in large enough quantities in an unbiased fashion.

Fourth, the use of predictor data sets is a complicated business, however; the decisions made here to include data sets from four African countries (Burkina Faso, Chad, Democratic Republic of the Congo, and South Sudan) whose populations and governments have largely abstained from mass vaccination provides, in this author's estimation, a more natural synthetic control with the least endogeneity interference possible. Additionally, by using total deaths and cases per million over time the trend line is much more consistent than looking at the dramatic ups and downs of new deaths/cases per million thus allowing for a more accurate and robust predictive capability for the counterfactual.

Finally, as some countries were not reporting data last year, or stopped reporting data at a certain point, there is limited pre-intervention data for these countries from 2020 with which to use for the Causal Impact Analysis. An analysis of those countries will have to be done separately.

Future Research

This study provides various suggestions for different forms of future research on this topic.

As mentioned above, a future study utilizing these methods, but looking at all-cause mortality rates rather than total deaths/cases per million will provide better pre-intervention data (as it goes back decades in some cases), which will be useful for prediction of trend lines. It will also provide a better understanding of the overall

causal impact of COVID-19 vaccine administration on the general health of the population rather than just the causal impact on COVID-19 associated deaths and cases.

An effort to obtain non-politicized raw data surrounding deaths and cases associated with COVID-19 in vaccinated and non-vaccinated individuals after vaccine administration began in each country will be vital for further accurate analysis of this important variable. This will allow us to distinguish between the effects of both the vaccines and the variants, but it must be done in a fashion that categorizes individuals properly. In this author's estimation, the proper categorization should be as follows: *Group 1: vaccinated* = anyone who has had any COVID-19 gene therapy injection at any moment following injection; *Group 2: non-vaccinated* = anyone who has never had any COVID-19 gene therapy injection.

Given the recent announcement by the Japanese Ministry of Health about their reasoning for rejecting a batch of 2.6 million Moderna vaccines because it contained a "metallic foreign substance that reacts to magnets" (⊕Urasaki and Nomura 2021), the possibility of different vaccines or batches having different effects due to contamination must be considered in any future analyses.

Additionally, the more we learn about natural immunity suggests that many populations may have had previous antibodies to SARS and SARS-like viruses, which may affect the vaccine impact.

It should be recalled this study only represents a snapshot in time, as such, future data points as they become available can be reviewed utilizing the calculation methodology explained in this paper to see if there are any changes in the results over time.

Finally, due to the announcement of the use of ivermectin by the Tokyo Medical Association and the state of Uttar Pradesh in India and the dramatic reduction in cases and deaths after administration (⊕Seth 2021; ⊕Hannah Ritchie and Roser 2020), and based on other promising research in repurposing ivermectin for use against COVID-19 (⊕Santin et al. 2021) a similar study to that presented here could be done to understand the causal impact of mass ivermectin administration as has been practiced now in Japan, India, Peru, and El Salvador, aside from the many tropical countries that administer ivermectin bi-yearly as a malaria prophylactic. Should enough countries adopt this policy and should the data become available in the future, this would be a worthwhile causal impact analysis to pursue.

Data Availability

All results, data, plots, and R code are included in this study for the easy replication by others. To access all files please refer to this Google Drive link: https://drive.google.com/drive/folders/1xOphw78-BhsMly09lpxCVPZ-WGieMSqO?usp=sharing

A PDF version of this study can be accessed here: https://drive.google.com/file/d/1DLlRa9rUqvW9pG1vNEsWMEydWwsmSMbe/view usp=sharing

All *CausalImpact* figures as compiled into multiple PDF pages can be downloaded here: https://drive.google.com/file/d/1lX3NVqY-sbxVLM81lgK5f6I5ny9c6zGL/view?usp=sharing

This author welcomes suggestions on how to improve any aspect of the methodology utilized in this study, please refer questions, suggestions, or criticism to Kyle A. Beattie kbeattie@ualberta.ca.

References

R (Version 4.0.3; ⊕R Core Team 2020) and the R-packages *AICcmodavg* (Version 2.3.1; ⊕Mazerolle 2020), *anytime* (Version 0.3.9; ⊕Eddelbuettel 2020), *broom* (Version 0.7.9; ⊕Robinson, Hayes, and Couch 2020), *car* (Version 3.0.10; ⊕Fox and Weisberg 2019; ⊕Fox, Weisberg, and Price 2020), *carData* (Version 3.0.4; ⊕Fox, Weisberg, and Price 2020), *cluster* (Version 2.1.0; ⊕Maechler et al. 2019), *corpus* (Version 0.10.2; ⊕Perry 2020), *correlation* (Version 0.6.1; ⊕Makowski et al. 2020), *corrplot2017* (⊕Wei and Simko 2017), *data.table* (Version 1.14.0; ⊕Dowle and Srinivasan 2020), *dplyr* (Version 1.0.7; ⊕Wickham et al. 2020), *finalfit* (Version 1.0.3; ⊕Harrison, Drake, and Ots 2020), *flextable* (Version 0.6.6; ⊕Gohel 2020), *forcats* (Version 0.5.1; ⊕Wickham 2020a), *Formula* (Version 1.2.4; ⊕Zeileis and Croissant 2010), *fpc* (Version 2.2.9; ⊕Hennig 2020), *GGally* (Version 2.1.2; ⊕Schloerke et al. 2020), *ggcorrplot* (Version 0.1.3; ⊕Kassambara 2019), *ggplot2* (Version 3.3.5; ⊕Wickham 2016), *ggpubr* (Version 0.4.0; ⊕Kassambara 2020), *ggraph* (Version 2.0.5; ⊕Pedersen 2020), *ggrepel* (Version 0.9.1; ⊕Slowikowski 2020), *ggstatsplot* (Version 0.8.0; ⊕Patil 2021), *ggthemes* (Version 4.2.4; ⊕Arnold 2019), *ggvis* (Version 0.4.7; ⊕Chang and Wickham 2019), *ggwordcloud* (Version 0.5.0; ⊕Le Pennec and Slowikowski 2019), *Hmisc* (Version 4.5.0; ⊕Harrell Jr, Charles Dupont, and others. 2020), *igraph* (Version 1.2.6; ⊕Csardi and Nepusz 2006), *kableExtra* (Version 1.3.4; ⊕Zhu 2020), *KernSmooth* (Version 2.23.20; ⊕Wand 2020), *knitr* (Version 1.36; ⊕Xie 2015), *lattice*

(Version 0.20.44; ⊕Sarkar 2008), *lsr* (Version 0.5; ⊕Navarro 2015), *NLP* (Version 0.2.1; ⊕Hornik 2018), *nortest* (Version 1.0.4; ⊕Gross and Ligges 2015), *papaja* (Version 0.1.0.9997; ⊕Aust and Barth 2020), *pdftools* (Version 3.0.1; ⊕Ooms 2020), *PerformanceAnalytics* (Version 2.0.4; ⊕Peterson and Carl 2020), *plm* (Version 2.4.1; ⊕Croissant and Millo 2008; ⊕Millo 2017), *plyr* (Version 1.8.6; ⊕Wickham et al. 2020; ⊕Wickham 2011), *psych* (Version 2.1.6; ⊕Revelle 2020; ⊕Makowski 2018), *psycho* (Version 0.6.1; ⊕Makowski 2018), *purrr* (Version 0.3.4; ⊕Henry and Wickham 2020), *quanteda* (Version 3.0.0; ⊕Benoit et al. 2018), *RColorBrewer* (Version 1.1.2; ⊕Neuwirth 2014), *readr* (Version 2.0.1; ⊕Wickham, Hester, and Francois 2018), *readtext* (Version 0.80; ⊕Benoit and Obeng 2020), *rvest* (Version 1.0.1; ⊕Wickham 2020b), *see* (Version 0.6.4; ⊕Lüdecke et al. 2020), *SnowballC* (Version 0.7.0; ⊕Bouchet-Valat 2020), *stringr* (Version 1.4.0; ⊕Wickham 2019), *survival* (Version 3.2.13; ⊕Terry M. Therneau and Patricia M. Grambsch 2000), *tesseract* (Version 4.1.1; ⊕Ooms 2019), *text2vec* (Version 0.6; ⊕Selivanov, Bickel, and Wang 2020), *tibble* (Version 3.1.5; ⊕Müller and Wickham 2020), *tidyr* (Version 1.1.4; ⊕Wickham 2020c), *tidytext* (Version 0.3.1; ⊕Silge and Robinson 2016), *tidyverse* (Version 1.3.1; ⊕Wickham et al. 2019), *tinylabels* (Version 0.2.1; ⊕Barth 2021), *tm* (Version 0.7.8; ⊕Feinerer, Hornik, and Meyer 2008), *topicmodels* (Version 0.2.12; ⊕Grün and Hornik 2011), *wordcloud* (Version 2.6; ⊕Le Pennec and Slowikowski 2019; ⊕Fellows 2018), *xml2* (Version 1.3.2; ⊕Wickham, Hester, and Ooms 2020), *xtable* (Version 1.8.4; ⊕Gohel 2020;

⊕Dahl et al. 2019), *xts* (Version 0.12.1; ⊕Ryan and Ulrich 2020), and *zoo* (Version 1.8.9; ⊕Zeileis and Grothendieck 2005)

Acknowledgements

The conclusions presented here are my sincere attempt to arrive at some truth of the matter at hand. Ultimately, God knows best where the truth lies.