Covid-19 Rates in Countries with Different Air Travel Restrictions

Carly Mahoney

4/11/2021

Introduction

COVID-19 has managed to reach every corner of the world, however it is not effecting regions equally. Some countries have nearly eradicated the virus while others are seeing major increases in cases. Possible reasons for the incidence differences are population and region specific regulations. This study looks at three neighboring countries each with unique air travel restrictions, from January 01, 2021 to March 01, 2021, to determine whether this type of restriction has an effect on COVID-19 incidence.

On April 12, 2021 Venezuela, Colombia, and Brazil all had different air travel restrictions that had been in place since at least the beginning of the year. Venezuela was completely closed to air travel, meaning only citizens and persons under special circumstances could enter the country. Colombia was partially open. Under the country's restrictions, a person could enter the country by air only if proof of a negative COVID-19 test was provided. Brazil was open, which meant at the time anyone could enter the country by air.

Three potential results could arise after adjusting for population. One, air travel restrictions have no effect on incident cases in a country. Two, a country has more restrictions because it has a relatively large number to cases. I hypothesize the third option that tighter restrictions will result in lower COVID-19 incidence.

Data Collection and Organization

The data for this analysis were collected by John Hopkins University of Medicine every day starting in January of 2020. The raw data set contains the cumulative case counts for states and provinces of various countries by date. It also provides the latitude and longitude for each province or state.

The data set was cleaned and organized to include only data for Brazil, Colombia, and Venezuela as whole countries instead of dividing them into provinces. It also was cleaned to include only the specified date range for this analysis. After cleaning and organization, the data set was manipulated to create two new variables, incident_cases and Population. The specific variables used in analysis were Date, country_region, incident_cases, cumulative_cases, and Population. Summaries for some of these variables can be found in Table 1. Date is a date, country_region is the country in which the cases occurred, incident_cases is the number of cases that occurred for a specific country and date, cumulative_cases is the number of cases that occurred up to the specified date, and Population is the population of the specified country.

Table 1: Numeric Summaries

country_region	Average Number of Cases Per Day	SD of Daily Cases	Cumulative Cases
Brazil	49149.6724	19376.5808	2850681
Colombia	10289.8276	5359.9176	596810
Venezuela	435.0345	103.5601	25232

Statistical Methods

In order to determine how COVID-19 situations in countries with different air travel restrictions compare and whether tighter restrictions mean fewer cases for a country, I used incidence rate ratios. An incidence rate as defined by the CDC is "how quickly disease occurs in a population" and is calculated using the equation:

$$IR = \frac{Number\ of\ new\ cases\ of\ disease\ during\ specified\ period}{Time\ each\ person\ was\ observed,\ totaled\ for\ all\ persons}.$$

The sum of the incident_cases variable for a particular country from January 01, 2021 to March 01, 2021 is the numerator. The denominator of the equation is in person per some unit of time. In this analysis I used person-years which used the Population variable. Table 2 has the total person-years for each country and the values used in the calculation made by the equation:

$$Total\ Person - Years = \frac{Population\ of\ the\ country}{365} * Number\ of\ days\ observed.$$

It is important to note using person-time requires the assumption that during the time period of observation, the probability of a person contracting COVID-19 was constant. After incidence rates are calculated, incidence rate ratios, which are simply ratios made up of two incidence rates, can be used to compare how incidence differs between regions. I calculated incidence rates and incidence rate ratios using the epi.2by2 function in R's epiR package. The function was adjusted to give incidence rates per 1000 persons per year. I used the function three separate times in order to compare each country with the other two countries.

Table 2: Person-Years Table

country_region	Total Incident Cases	ndays	Total Person Years
Brazil	2850681	58	33776565
Colombia	596810	58	8085500
Venezuela	25232	58	4518588

Results

Since the countries of Brazil, Colombia, and Venezuela have vastly different populations, it was necessary to look at the case counts after adjusting for population. Figure 1 shows the incidence of COVID-19 in Brazil, Colombia, and Venezuela per 100,000 people. The plot tells us during the time period analyzed, Brazil and Colombia both had a relatively high incidence of COVID-19 with respect to population compared to Venezuela which had a constant low incidence throughout the time period. The noticeable difference between Brazil and Colombia is how incidence changed throughout the time period. Brazil had a pattern which stayed constant while Colombia saw a steady reduction in incidence starting around January 20th.

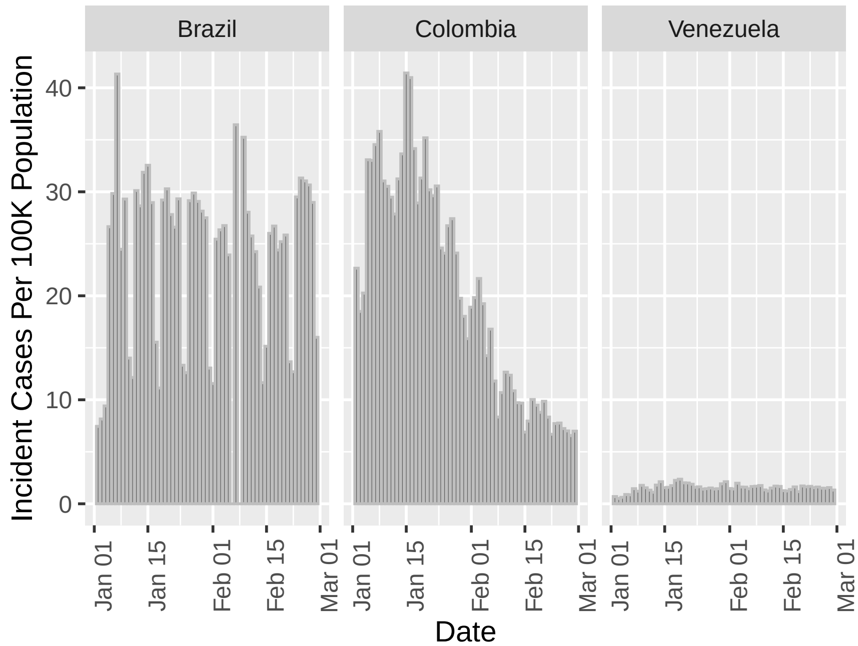

Figure 1: Incidence with Respect to Population

The calculated incidence rates for each country are provided in Table 3. Brazil's incidence rate was 84.40 persons per 1000 person-years, Colombia's was 73.81 persons per 1000 person-years, and Venezuela's was 5.58 persons per 1000 person-years.

Country	Estimate
Brazil	84.40
Colombia	73.81
Venezuela	5.58

Table 3: Incidence Rates

The epi.2by2 function also calculated estimates and confidence intervals for the incidence rate ratios between the three countries shown in Table 4. An example of how an estimated value is read is, the risk of contracting COVID-19 in Brazil is 15.114 times that of contracting the virus in Venezuela. All of the confidence intervals are entirely above the value one meaning all the pairwise comparisons are significant.

	Estimate	Lower	Upper
Brazil vs Venezuela	15.114	14.929	15.302
Colombia vs Venezuela	13.218	13.056	13.383
Brazil vs Colombia	1.1434	1.1405	1.1463

Table 4: Incidence Rate Ratio Confidence Intervals and Estimates

From both the incidence rates and incidence rate ratios, we can see that the incidence of COVID-19 in Brazil was greater than that of Colombia which was greater than that of Venezuela during the time period from January 01, 2021 to March 01, 2021. The difference between Brazil and Colombia is small compared to how both these countries differ from Venezuela.

Discussion

The results of the analysis suggest that air travel restrictions may have an effect on COVID-19 incidence rates. My original hypothesis was countries with open air travel statuses would have larger COVID-19 incidence rates. This hypothesis was correct when considering the analysis results. Brazil was open to air travel and had the highest incidence rate and Venezuela was closed to air travel and had a significantly lower incidence rate.

Even though the results are consistent with my hypothesis, there are still more questions to be answered. Just because we see this pattern of fewer air travel restrictions equaling larger incidence rates for these three countries, we cannot conclude this is the case for all countries in similar situations. Brazil, Colombia, and Venezuela are not the only neighboring countries with three different air travel restriction statuses. Further analysis should be done on other sets of three countries to see if this pattern exists elsewhere or if these results are just a coincidence.

Another question we could ask is, what other factors influence COVID-19 incidence? It is important to note that the pattern seen in this analysis may not be solely explained by air travel restrictions. It could be the case countries with tighter air travel restriction have tighter restrictions overall and therefore have lower numbers of COVID-19 cases. Understanding all the factors that contribute to COVID-19 incidence will need more research and analysis.

Appendices

Appendix I: Reading and Maniplulating the Data

```
jhu_url <- paste("https://raw.githubusercontent.com/CSSEGISandData/",
  "COVID-19/master/csse_covid_19_data/",
  "csse_covid_19_time_series/",
  "time_series_covid19_confirmed_global.csv",sep="")

multi_confirmed_long_jhu <- read_csv(jhu_url) %>%
  dplyr::rename(province = "Province/State", country_region = "Country/Region") %>%
  filter(country_region %in% c("Venezuela","Colombia","Brazil"))  %>%
  group_by(country_region) %>%
  pivot_longer(-c(province, country_region, Lat, Long), names_to = "Date",
               values_to = "cumulative_cases") %>%
  mutate(Date = lubridate::mdy(Date)) %>%
  mutate(incident_cases = c(0, diff(cumulative_cases)))

# Adding the populations

popdf <- data.frame(country_region=c("Venezuela","Colombia","Brazil"),
                    Population=c(28435940,50882891,212559417))

dfcompare <- dplyr::left_join(multi_confirmed_long_jhu,popdf, by="country_region") %>%
  mutate(incper100k=100000*(incident_cases/Population))
```

Appendix II: Numeric Summaries

```
statstable1 <- dfcompare %>% filter(incident_cases > -1) %>%
  filter(Date < lubridate::mdy("03/01/2021") & Date > lubridate::mdy("01/01/2021")) %>%
  group_by(country_region) %>%
  dplyr::summarise(`Average Number of Cases Per Day`= mean(incident_cases),
                   `SD of Daily Cases`= sd(incident_cases), `Cumulative Cases`=
                    sum(incident_cases))
statstable2 <- statstable %>%
  mutate(`Total Person Years`=ndays*c(212559417,50882891,28435940)/365)
```

Appendix III: Graphs

```
multi_confirmed_long_jhu %>% filter(incident_cases > -1) %>%
  filter(Date < lubridate::mdy("03/01/2021") & Date > lubridate::mdy("01/01/2021")) %>%
  ggplot(aes(x=Date,y=incident_cases)) +
  geom_bar(colour="grey", stat="identity") +
  facet_grid(~ country_region) +
  xlab("Date") + ylab("Incident Cases") +
  theme(axis.text.x = element_text(angle = 90))

dfcompare %>% filter(incident_cases > -1) %>%
  filter(Date < lubridate::mdy("03/01/2021") & Date > lubridate::mdy("01/01/2021")) %>%
  ggplot(aes(x=Date,y=incper100k)) +
```

```
  geom_bar(colour="grey", stat="identity") +
  facet_grid(~ country_region) +
  xlab("Date") + ylab("Incident Cases Per 100K Population") +
  theme(axis.text.x = element_text(angle = 90))
```

Appendix IV: Analysis

```
#Determining total person years
 dfcompare1 <- dfcompare %>% filter(incident_cases > -1) %>%
  filter((Date < lubridate::mdy("03/01/2021")) & (Date > lubridate::mdy("01/01/2021")))

statstable <- dfcompare1 %>%
  group_by(country_region) %>%
  dplyr::summarise(`Total Incident Cases` = sum(incident_cases),ndays=n())


BC <- as.table(matrix(c(2850681, 33776565, 596810, 8085500), nrow = 2, byrow = TRUE))
BCval <- epiR::epi.2by2(dat = BC, method = "cohort.time", conf.level = 0.95, units = 1000,
outcome = "as.columns")
print(BCval)

BV <- as.table(matrix(c(2850681, 33776565, 25232, 4518588), nrow = 2, byrow = TRUE))
BVval <- epiR::epi.2by2(dat = BV, method = "cohort.time", conf.level = 0.95, units = 1000,
outcome = "as.columns")
print(BVval)

CV <- as.table(matrix(c(596810, 8085500, 25232, 4518588), nrow = 2, byrow = TRUE))
CVval <- epiR::epi.2by2(dat = CV, method = "cohort.time", conf.level = 0.95, units = 1000,
outcome = "as.columns")
print(CVval)
```

References

CDC. Principles of Epidemiology | Lesson 3 - Section 2 https://www.cdc.gov/csels/dsepd/ss1978/lesson3/section2.html (accessed Apr 30, 2021).