



ISSS616 – Applied Statistical Analysis with R
Project Proposal

*Analyzing the potential COVID -19 vaccinators across the
globe for Arcturus Therapeutics*

(Group 4)

Gopinath Architha

Li Xiang

Raunak Kapur

Raveena Chakrapani

Samyukta Arun Hariharan

ABSTRACT

The COVID-19 pandemic has led to a dramatic loss of human life worldwide and has impacted almost every corner of life, causing global economies to stall, changing the way we work and stretching healthcare systems to the limit. Governments around the world have been forced to implement harsh restrictions on human activity to curb the spread of the virus. COVID-19 vaccination is now offering a way to transit out of this phase of the pandemic. Fortunately, the beginning of 2021 saw numerous vaccines given emergency approval and begin their roll out in countries across the world. Collaborating with Duke-NUS Medical School in Singapore, US firm Arcturus Therapeutics is in process of developing the vaccine ARCT-154 and ARCT-165¹ which will target the newer COVID variant. This vaccine is currently between Phase 2 and Phase 3 and is yet to be authorized for Emergency use by the WHO and other countries.

OVERALL CONCEPT

With vaccines of high efficacy such as Pfizer and Moderna (95%)² and other vaccines such as Sinopharm which can be maintained with fewer storage costs, it is going to be a challenge for Arcturus Therapeutics to enter the market. They would need to understand the current requirements of countries, particularly analyzing the current number of vaccinated and unvaccinated people. Insights can be derived on how they can devote more time on the R&D, manufacturing process, marketing strategies so that their vaccines can cater to maximum people across the globe. This includes clinical trials, targeting people with co-morbidities, children under 12 years, and other such scenarios.

The purpose of this project is to analyze the potential COVID-19 vaccinators across the globe for Arcturus Therapeutics vaccine. This analysis will be conducted from two points of view – The population that would require these doses (unvaccinated and booster-less population) and the properties of existing vaccines that deprive certain sections from getting vaccinated.

This study performs Exploratory Data Analysis on existing vaccines and its vaccination rate by countries, composition, and storage capabilities and on the population, considering parameters such as age groups, co-morbidities, booster status, financial status of countries, and other reasons for lack of vaccination.

Finally, the two are correlated to provide Arcturus Therapeutics the right platform to target their vaccinations, using statistical methods and inferential analysis techniques.

¹ <https://www.biospace.com/article/releases/arcturus-therapeutics-updates-data-from-arct-154-and-arct-165-booster-clinical-trial-demonstrating-robust-neutralizing-antibody-responses-to-sars-cov-2-and-several-variants/>

² <https://www.moh.gov.sg/covid-19/vaccination/faqs---safety-and-efficacy-of-the-covid-19-vaccine>

DATA SOURCES

This project uses datasets from the following sources:

1. Data.gov.sg ([Source Link 1](#))
2. IMF- WHO Covid-19 Vaccine Tracker ([Source Link 2](#))
3. Our World in Data ([Source Link 3](#))
4. ClinicalTrials.gov ([Source Link 4](#))
5. Worldbank.org ([Source Link 4](#))

The following table describes the data (not limited to) which is relevant to our project:

DATASET	VARIABLE(S)	DESCRIPTION/USAGE	DATA SOURCE	TARGET GROUP FOR ANALYSIS
vaccinations-by-age-group	location, date, age_group and people_vaccinated_per_hundred	People vaccinated (by hundred) per age group by date and location	Our World Data	Population
vaccinations-by-manufacturer	location, date, vaccine and total_vaccinations	Total vaccinations by vaccine manufacturer by date and location	Our World Data	Vaccines
Worldbank tracker	% total population vaccinated with at least one dose	To determine the percentage of population per country who received atleast one dose	Worldbank.org	Population
Worldbank tracker	% total population fully vaccinated	To determine the fully vaccinated percentage of population per country	Worldbank.org	Population
Worldbank tracker	Doses administered as a percentage of doses received	To check the vaccine usage/wastage per country	Worldbank.org	Population
Worldbank tracker	Vaccination rate trend of the past month (4-wk avg.)	To check the vaccination rate trend, whether it is stable, upward or downward	Worldbank.org	Population
Worldbank tracker	Availability of bank financing for vaccine deployment (US\$ millions)	To find out whether the countries can afford the vaccine	Worldbank.org	Population

bilateral_deal	country, deal/report date, vaccine, number of doses	To view the bilateral vaccine supply chain to country over time	IMF- WHO Covid-19 Vaccine Tracker	Vaccines
donation_delivered	country, vaccine, month, number of doses delivered through direct donations	To gauge the ratio between direct delivery of vaccines and donations by other countries	IMF- WHO Covid-19 Vaccine Tracker	Vaccines
supply_tracker	ISO3, population, Secured and/or Expected Vaccine (millions of courses)	To track the current expectation of confirmed vaccine delivery with regards to the population	IMF- WHO Covid-19 Vaccine Tracker	Vaccines

METHODOLOGY

DESCRIPTIVE STATISTICS

Before we commence with our observations, we plan to perform the descriptive statistical analysis to better understand the data. We will study the population data and based on the skewness of the normal distribution curve of numerical data we will chalk out our observations. We plan to validate our samples created with the help of Central Limit Theorem and calculate cumulative and probability density functions to get a better understanding of the data.

For visualization, we plan to use the following graphs:

- **Box Plot:** We can compare categorical and numerical data and learn how the central tendencies vary for each category/group
- **Scatter Plot:** This helps us understand the correlation between 2 numerical data points such as comparing the current COVID-19 cases vs Number of people vaccinated in a particular country
- **Geospatial Chart:** This is a great visualization tool for the readers to understand the data by looking at the world map such as number of people required to be vaccinated in each country to achieve a 70% fully vaccinated benchmark, doses administered as a percentage of doses received in each country and so on.
- **Histogram:** This can be useful in understanding the distribution/frequency of the categorical data for us to make appropriate decisions.
- **Curve:** Mainly the fit distribution of our analysis objectives, which can give us an overview of the data we are looking into and give the clues of further statistics analysis.

INFERENCE STATISTICS

Below include (but do not limit to) some inferential statistics proposed:

- Confirmatory analysis techniques such as **hypothesis testing and parametric/non-parametric tests** based on normality of data to gauge the reception of vaccines by certain age groups
- **Regression analysis** to infer the correlation between vaccine situation and COVID-19 situation for countries with downward and stable vaccination trends
- **Clustering analysis** to create homogeneous groups of population that Arcturus Therapeutics can target
- **Decision Tree** to plan the best time (in month-year format) for Arcturus Therapeutics to tap on the target countries
- **Confidence Interval Estimations** to assess the conclusions assurance.

R LIBRARIES

This project uses the list of (not limited to) following libraries

- dplyr – Data Manipulation
- tidyverse, tidyr, lubridate – Data Cleaning
- shiny – Web Application
- plotly – Interactive Graphs
- knitr – Report creation
- leaflet, rgdal, tmap, sp, htmlwidgets, corrplot, ggplot2- Data Visualization

CONCLUSION

From our analysis, we can recommend the following to the relevant department heads at Arcturus Therapeutics:

- Potential countries and the best time to target those who need vaccines
- Tweaks in vaccine composition to reach out to the larger group provided the in-progress phase testing timelines are not hampered
- Fill in the voids which were created by other competitors and create appropriate marketing strategies to reach maximum numbers