

Statistical Analysis of COVID-19 by 6 Sigma

N. R. Bhatia,
Mechanical Department
V.J.T.I. Mumbai, India

S.R Mulani
Mechanical department,
V.J.T.I., Mumbai, India

S. D. Khandagale
Mechanical department,
V.J.T.I. Mumbai, India

Prof. H. P Khairnar,
Mechanical department
VJTI., Mumbai, India

Abstract— The Covid Crisis looming over the world has overtaken the number of deaths caused in the world wars & has caused greater economic shrinkage than the Great Depression of 1928. Needless to say this one of the biggest healthcare & economic crisis that has ever hit humanity. In this paper we try to study issues like virus transmission, mortality rate, vaccination rates etc & provide possible solutions on the basis of statistical analysis using Six Sigma.

With the help of the analysis we statistically point out various areas of concern & provide required solutions. We observed that the corona virus mutates at a very faster rate due to which with time the vaccines developed may be deemed unfit to deal with the new variants. One way to tackle this problem is a swift & immediate mass vaccination drive which not only provides herd immunity to a country but also opens up the possibility of reopening the economy. For such a mass vaccination drive we have suggested an app which can be used to assign people vaccination time based on their current address to speed up the mass immunization drive.

Keywords— Herd immunity, Six Sigma, Mass immunization

I. INTRODUCTION

The objective is to reduce the Mortality Rate to the minimum and achieve maximum containment of the virus & conduct a swift vaccination drive. Also to measure the virus outbreak and its impact by using Pareto principles, Run Charts, various probability distributions (Eg Normal, Poisson, Chi-square distribution, ANOVA, etc.) and determine the factors that are Critical to Quality (COQ) i.e in reducing the mortality as well as virus containment along with steps for vaccination drive. These steps give us a well-defined process to choose from on the basis of the lowest mortality rate. Using the Process Capability Study, Ishikawa/Fishbone Diagrams, Failure Mode & Effects Analysis (FMEA) and Hypothesis Testing to determine the important causes of deaths, special and common causes of variation, Deaths per Million (DPO). We choose the method with the minimum death rate & optimal vaccination percentage determined by designing of the solution design matrix by analysing the key criteria and providing weightage to each criterion. This step also involves use of Project Planning & Management tools to determine various possible solutions and the selection of the most feasible solution. Management parameters including risk, cost, scope etc. are controlled in this phase & Measure the gains made by the project through Statistical Process Control and Control charts. This Phase is to ensure that the project

achieves the goals set in the Define phase.

We observed that about 80 percent of the cases & deaths were concentrated in a limited no of countries & regions which goes a long way in explaining the transmission of the virus. The hotspots are one of the major reasons for the difficulty in mitigating the virus along with its high spread ratio. With the help of six sigma analysis

we determined the relation between rise or fall in cases with respect to temperature, population density, percentage population vaccinated & active cases.

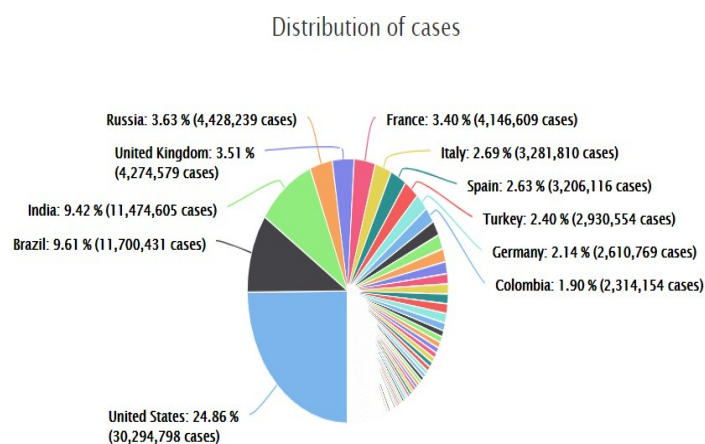


Fig 1. Distribution of Cases

II. Six Sigma Analysis

1. Define:

The problem statement is “To reduce the mortality Rate & suggest measures for the virus containment along with practical solutions for conducting a mass vaccination drive”.

Pareto Charts:

A Pareto chart is a bar graph. The lengths of the bars represent frequency or cost (time or money), and are arranged with longest bars on the left and the shortest to the right. In this way the chart visually depicts which situations are more significant. This cause analysis tool is one of the basic quality tools used in the Six Sigma Analysis. We use this tool for analysis of multiple parameters for selected countries.

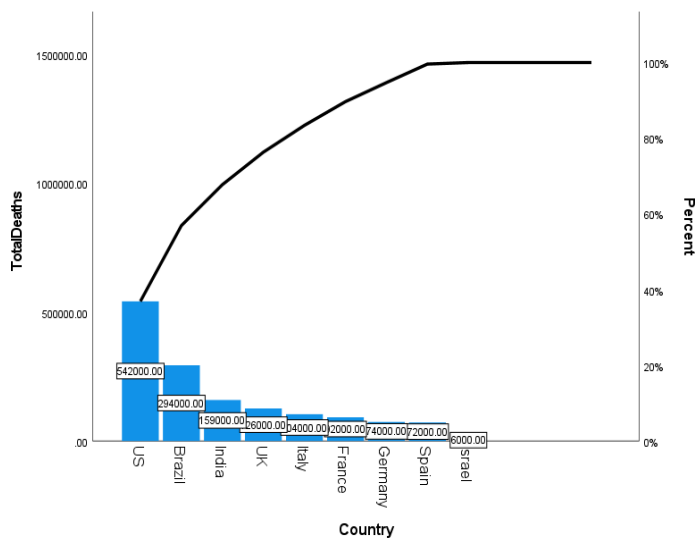


Fig 2a. Total Deaths vs Countries

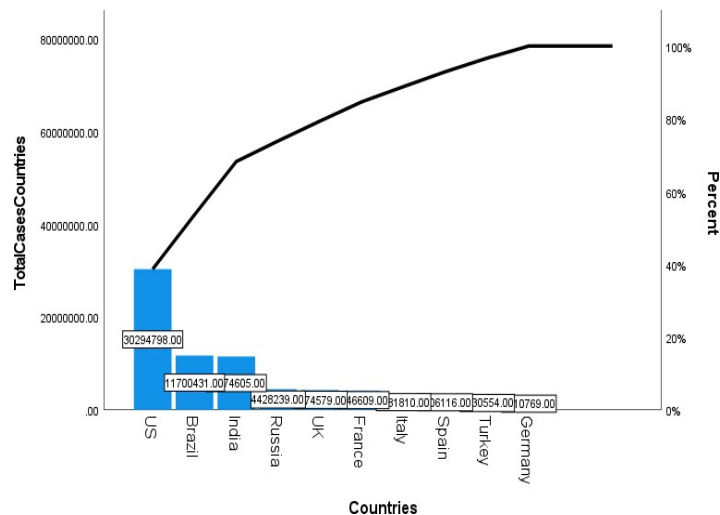


Fig 2b. Active Cases vs Countries

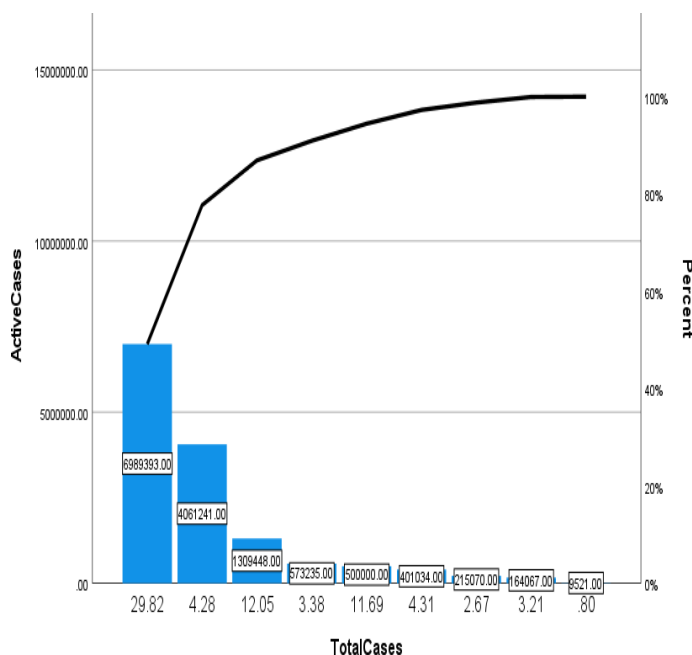


Fig 2c. Active vs Total Cases

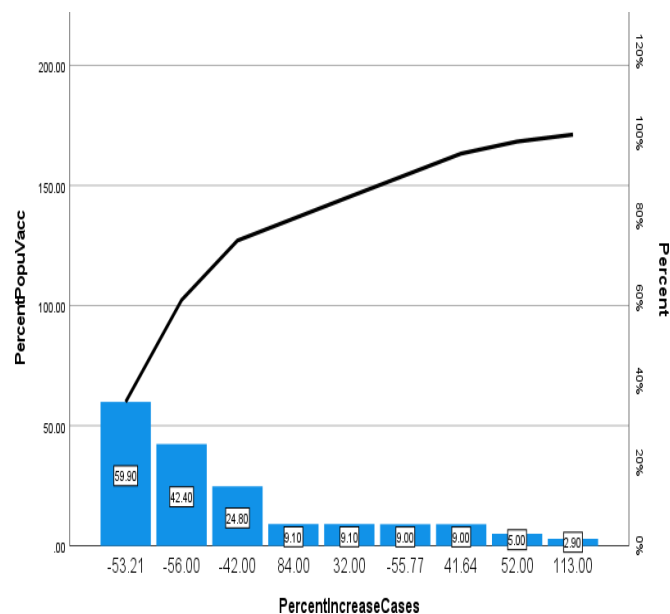


Fig 2d. % Vaccinated vs % Increase in Cases

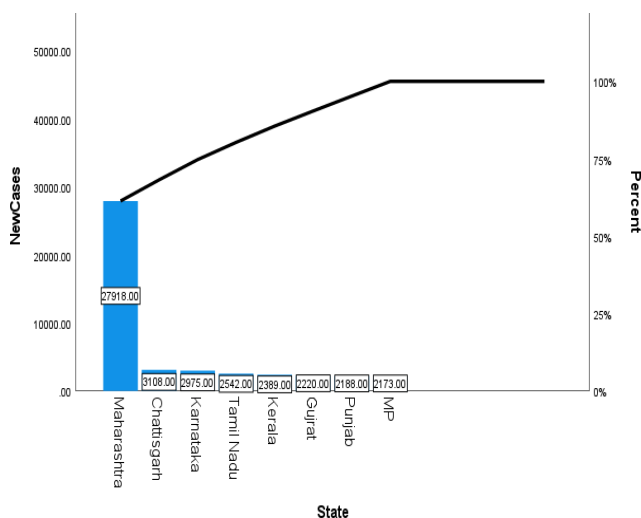


Fig 2e. New Cases in States

From the above graphs we learn that majority of the cases are concentrated in only 5-6 countries or regions. Also that most of the deaths have also happened in the countries with highest of cases & most number of active cases. One of the most pertinent statistic is the one that shows that the countries with higher percent of vaccinated people has a drastic decrease in daily cases. Even in Indian State more than 85% of the cases are concentrated in 8 out of the 29 states. This further goes on to show that the transmission of the cases is mainly in certain concentrated regions & hence greater attention & resources are to be devoted to these regions. The concentration of cases in such regions may be by virtue of factors like high population density, lower testing rate initially, primitive medical infrastructure, lack of restrictions, climatic conditions, proportion of active cases etc. which form some of the factors for our further analysis.

2. Measure

Run Charts

daily cases on 15 jan - Run Chart

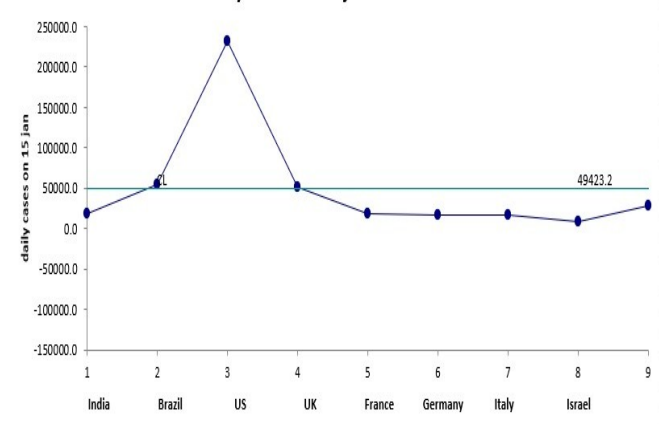


Fig 3a. Run Chart: Daily Cases

People tested in a second - Run Chart

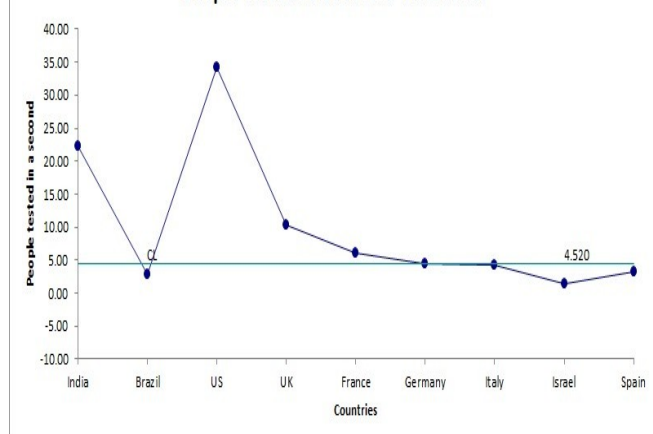


Fig 3b. Run Charts: Tests per second

People vacc in a second - Run Chart

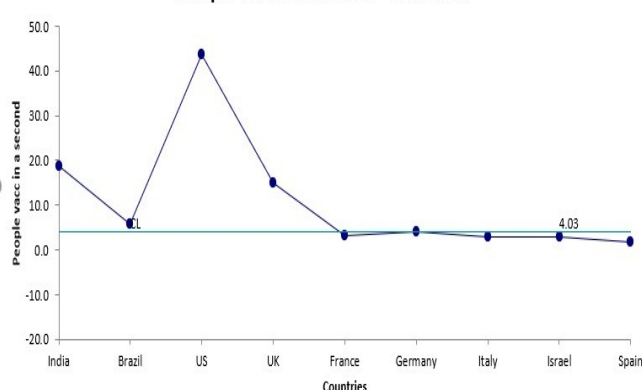


Fig 3c. Run Chart: Vaccinations per second

Severity Rating (5- Very severe, 1- Good)

1. India- 4 2. Brazil- 4 3. US- 5 4. Uk- 5 5. France- 4 6. Germany- 3 7. Italy- 5 8. Israel- 2 9. Spain- 3 10. Africa- 1

3. Analyse

Ishikawa Diagram: An Ishikawa diagram is a diagram that shows the causes of an event and is often used in manufacturing and product development to outline the different steps in a process, demonstrate where quality control issues might arise and determine which resources are required at specific times.

Fig 4a. Ishikawa Diagram: For rise in cases during first wave

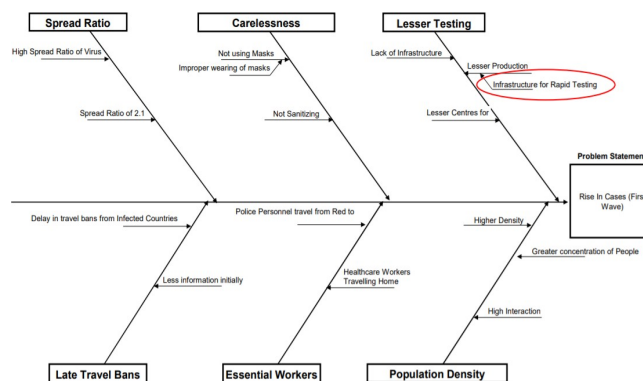
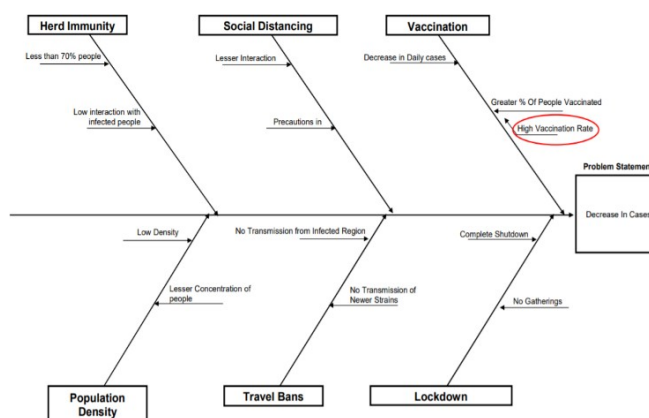


Fig 4b. Ishikawa Diagram: For dip in cases in US and UK



Risk Priority Number (RPN) allocated:

Infected People Coming in from foreign countries-448, Infected covid warriors going home in green zones-336, Improper wearing of masks, touching the nose, eyes or mouth without sanitization of hands-315, Hand shakes, objects exchanged & not sanitized-336, Sports arenas, political rallies, elections, protests etc.-810, Lack of Doses-504, Partial number of vaccination centres being functional-336, Lack of vaccination infrastructure-320, Extreme waiting periods not feasible for elderly & adults with comorbidities-210, Side effects after vaccination-42, Doses going to waste due to low turnout-392

Null hypothesis

Chi-square Test for Independence Death Rate vs Population Density

Death Rate * Population Density Crosstabulation					
Death Rate	High	Population Density		Total	
		high	low		
High	Count	11	3	14	14.0
	Expected Count	9.1	4.9		
Low	Count	2	4	6	6.0
	Expected Count	3.9	2.1		
Total	Count	13	7	20	20.0
	Expected Count	13.0	7.0		

Outcome: There is association between death rate & Population density.

Chi-Square test of Independence between Covid Cases & Temperature Condition

Case Rate * temp Crosstabulation					
Case Rate	high	Count	temp		Total
			cold	hot	
		Count	7	1	8
		Expected Count	6.2	1.8	8.0
	low	Count	0	1	1
		Expected Count	.8	.2	1.0
Total	Count		7	2	9
	Expected Count		7.0	2.0	9.0

Outcome: There is association between case rate & temperature condition.

ANOVA for % Change in Daily Cases & Vaccination Rate:

Countries	% Increase in daily cases	Countries	% Increase in daily cases
US	-76.19	India	27.52
UK	-88.63	Brazil	23.21
Israel	-72.61	France	33.6
Spain	-83.11	Germany	-39.4
Russia	-53.15	Italy	37.1
Total	-379.695	Total	82.01
Mean	-75.939	Mean	16.402

Outcome: ANOVA for this cases proves that the mean of increase in percentage cases in regions with high vaccination rates is not equal to regions in low vaccination rates. Hence, we conclude that rate of vaccination has an impact on the daily cases & by observed data we can conclude that daily cases decrease as vaccination rate increases.

Co-relation between total cases and active cases

Correlations			
Total Cases	Pearson Correlation	Total Cases	Active Cases
		1	.799**
			.010
		9	9
Active Cases	Pearson Correlation	.799**	1
			.010
		9	9

** . Correlation is significant at the 0.01 level (2-tailed).

Outcome: Strong correlation between total cases & active cases

Correlation between % Population Vaccinated vs % Increase in Cases.

Correlations			
% Population Vaccinated	Pearson Correlation	% Population Vaccinated	% Increase in Cases
		1	-.703*
			.035
		9	9
% Increase in Cases	Pearson Correlation	-.703*	1
			.035
		9	9

*. Correlation is significant at the 0.05 level (2-tailed).

Outcome: Strong negative correlation exists.

4. Improve

Process capability study

Serial no.	Country	% Population Vaccinated
1	Israel	61.29
2	UK	47
3	Chile	38.1
4	US	33.5
5	Baharain	32.45
6	Hungary	27.84
7	Uruguay	22.97
8	Germany	14.57
9	Turkey	12.6
10	Brazil	9.12
11	India	5.98
12	Russia	5.58
13	Canada	17.18
14	France	14.89
15	Spain	14.52
16	Italy	14.02
17	Greece	12.99
18	Turkey	12.6
19	Switzerland	12.11
20	Japan	0.82
21	South Africa	0.47
22	Denmark	14.19
23	Mexico	6.68
24	Finland	17.93
25	Serbia	22.82

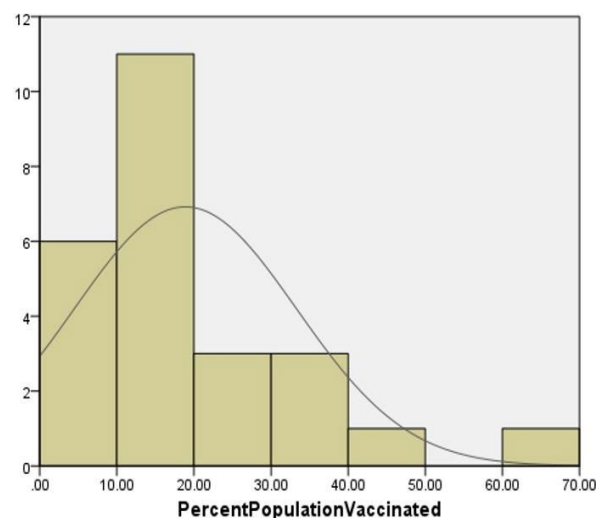


Fig 5. Process Capability Chart for % population vaccinated

We Observe that for the sample of 25 countries that most countries have vaccinated between 0-20% of their population. However, to be able to achieve herd immunity about 70% of the population should be vaccinated. Here the acceptable bell curve peak should be between 30 – 70 % wherein most of the countries should lie. Here 20 out of 25 countries lie below this range ie 800000 out of a million and hence the sigma level is 1. Now we need to propose a solution for rapid vaccination drive so as to improve the sigma level to 6 to reduce the daily cases, deaths and to be able to reopen the economy.

5. Control

Control Charts

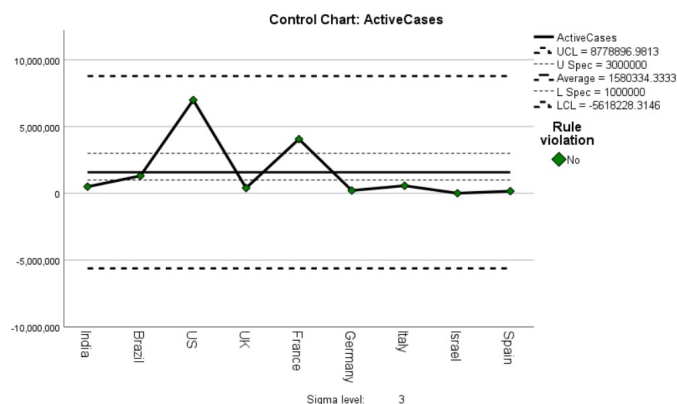


Fig 6a. Control Chart: Active cases in Countries

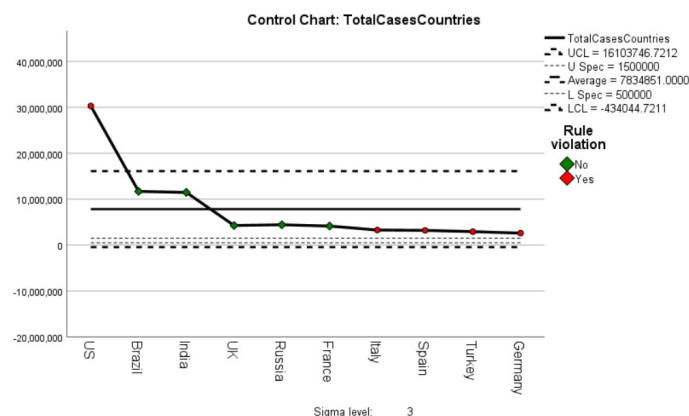


Fig 6b. Control Chart: Total cases in Countries

5S Technique

1. For restraints during the lockdown, the following key points are absolutely pragmatic: sort, zones, improve, form standard wards, sustain, emergency.
2. For vaccination drive, the following key points are pragmatic: sort, arrange, chaos-free, standardize, sustain.

III. Results:

Some generalized results which could be observed from the data per se could be stated as:

1. Majority of the cases are concentrated in limited number of countries (5-6) & regions which goes on to emphasize the need to focus on those countries to mitigate the virus.
2. The countries with higher vaccination rates see a decrease in daily cases.
3. The countries with greater number of active cases witness greater cases & higher death rates.
4. The virus spreads faster in the regions that have colder climatic conditions.
5. Virus spreads faster in countries with a population density
6. Death rate is higher in countries with high population density.
7. Death rate is higher in countries with greater

- percentage of people above the age group of 50.
8. If rapid vaccination is not carried out we would continue to experience wave after wave of the virus inspite of living in constraints.
9. There has been a constant surge in the daily cases since the onset of the virus across the globe, the generalized period of which could be observed and understood as the end of January or the beginning of February of the previous year. This surge however, cannot be deemed to be of a constant nature, since disruptions in a regular rise in the graphs of the plotted data sets have been observed.
10. Furthermore, when the European nations are peculiarly concerned, these nations owing to their relatively lower population, the deaths observed can be deduced to be of a greater proportion of their existing population, especially if, such sort of a comparison is primarily made with a nation like India.
11. The varied demographic circumstances all across the globe have shown a considerable impact on the spread of the virus, where those nations having a greater older population have been largely affected due to the vulnerable population.
12. The pollutants in the atmosphere also do have an impact on the spread of the virus, where the regions having a greater atmospheric pollution have shown a greater spread, for an instance, the Lombardy region of Italy.

IV. Discussions

The research conducted has been consistent with the findings and the reports registered with authentic data sources like the World Health Organization i.e. WHO, the official data collected from the Government websites of the concerned nations. There has been a constant surge in the daily cases since the onset of the virus across the globe. For the lower population nations, the death rate as a proportion of the total population is larger as compared to the higher population nations. The exponential rise in the infections can be related to lifted lockdowns, increased public exposure, increase in community transmission as also various Covid variants. The time period of massive attacks by the virus i.e. the waves seem to show a non-uniform global trend. European nations peaked in March of last year, to fall to negligible values in June. Whereas, the US has shown a constant upward trend of cases. Brazil has an all-together different trend of a surge in cases as compared to the other nations. India had seen a peak in cases quite later the earlier year as compared to the other nations, and presently is under the attack of the second wave, where the cases have already surpassed the previous numbers. Maharashtra has the highest number of cases in India.

Vaccination across the globe:

The lower population nations have inoculated some significant proportion of their population by at least one shot of the vaccine, however, these proportions

seem significant owing to the lower population in these nations, for an instance the European nations and Israel. When we consider the proper statistical data, the United Kingdoms have been able to inoculate a massive 47.9% of the population. The United States have administered vaccine to 114,436,039 people which constitute 34.9% of the entire population. People in Israel fully vaccinated are 54.3%. When we consider a nation like India, wherein the proportion of the vaccinated people, even by at least one shot of the vaccine is comparatively lower, we need to consider the massive population of India. The United States has done a commendable job in inoculating people. The vaccines used across the globe with two dosages are Pfizer, AstraZeneca (Covishield), Moderna, BioNtech, whereas Johnson & Johnson has a single shot, each vaccine with respective efficacies. These are the most widely used vaccines all over the globe. However, when we peculiarly consider the region of South America, many nations have shown reliance on the Chinese vaccines, for an instance, Peru, Colombia, Argentina. The Indian states have been administering Covishield and Covaxin to their citizens, the vaccination being the highest in Maharashtra, Gujarat, Uttar Pradesh, Rajasthan, West Bengal.

V. Conclusions:

The supply of vaccines all across the globe is extremely important for, more the reachability greater shall be the resistance to the virus. If such a co-operative global approach were not to be followed, the emerging mutants in the massively affected nations might raise the chances of more and more waves of different variants all across the globe. The vaccination drives of the nations across the globe is consistent with the demographic division observed, with the older population being considered for inoculation first.

1. When we consider vaccine efficacy, the mRNA based vaccines of Pfizer and Moderna have shown an efficacy of 95% and 94.1% respectively in an absence of an earlier Covid infection. Whereas, the carrier vaccines of AstraZeneca have shown an efficacy of 85% in citizens over 65 years of age. Johnson & Johnson has an efficacy of 86% against severe disease in the US.
2. Vaccine diplomacy has been pragmatic since the initial days of development of vaccines, with nations like India taking a lead in this. Whereas, the US and Canada have displayed examples of vaccine nationalism, with the US placing an embargo on the raw materials and Canada with surplus vaccines order.
3. These vaccines have not been able to arrest the onset of Covid, however, have shown a reduction in severity of infection when administered.
4. Mass vaccination for the globe is required to

counter the devastation ushered by the virus.

VI. Suggestions:

1. Formulations of zones of virus with the boundaries completely sealed except for medical emergencies.
2. Ramp up the vaccination rate in the country to conduct & finish mass vaccination within 12 months.
3. Finish Vaccination of 70% of the populations as quickly as possible to attain herd immunity which will also be helpful in reopening the economy.
4. Use the app suggested in the improve phase to conduct rapid mass vaccination by assigning a vaccination time to each individual.
5. Focus most of the resources in areas that are hotspots for the virus as those are limited & contribute majority of cases.
6. A quite sensible approach is to develop a vaccination system and use it for pilot run on a regional basis, further expanding its paradigm as per the success. This sounds to be very consistent with time, efforts and economics.

VII. Website developed

A website has been developed for representing the daily cases and the vaccination data. This website shall provide the user with some interactive, graphical information of the vaccination status in the major countries, which have been under study for the Covid situation. Also, the statistical data of dose 1, dose 2 and total vaccination of the various Indian states is displayed. All of the data is in accordance with the official government sites. The first interface is the Home Page where the user shall interact with four links respectively for country matrix, Indian states matrix, WHO site and mygov.in site. Furthermore, clicking on any of the highlighted areas of the world map shall direct the user to the interactive graphs of vaccination and cumulative cases and deaths in the nations. Clicking on any of the highlighted areas in the Indian map shall direct to the vaccination status of those respective states. Thus, this consolidated data shall be beneficial for the user for any aid in research.



Fig 7a. Website Home Page PAGE NO: 6

Key message

It is imperative to understand the parameters which more or less impact the spread of the virus. This paper aims at studying these influencing parameters from a statistical point of view and drawing appropriate conclusions from the same, so as to understand and decipher some solutions which might aid the global society in curbing the spread and the existence of this virus, as a whole.

Acknowledgement

All the authors are thankful to the faculties and staff of AIIMS, Patna, Bihar for giving support and guidance in the study.

Disclosure:

All the authors declare no conflicts of interest for this work.

References

1. Jernigan DB; CDC COVID-19 Response Team. Update: public health response to the coronavirus disease 2019 outbreak—United States, February 24, 2020. MMWR Morb Mortal Wkly Rep 2020;69:216–9. 10.15585/mmwr.mm6908e1 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
2. Qualls N, Levitt A, Kanade N, et al.; CDC Community Mitigation Guidelines Work Group. Community mitigation guidelines to prevent pandemic influenza—United States, 2017. MMWR Recomm Rep 2017;66(No. RR-1). 10.15585/mmwr.rr6601a1 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
3. Moriarty LF, Plucinski MM, Marston BJ, et al.; CDC Cruise Ship Response Team; California Department of Public Health COVID-19 Team; Solano County COVID-19 Team. Public health responses to COVID-19 outbreaks on cruise ships—worldwide, February–March 2020. MMWR Morb Mortal Wkly Rep 2020;69:347–52. 10.15585/mmwr.mm6912e3 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
4. Patel A, Jernigan DB; 2019-nCoV CDC Response Team. Initial public health response and interim clinical guidance for the 2019 novel coronavirus outbreak—United States, December 31, 2019–February 4, 2020. MMWR Morb Mortal Wkly Rep 2020;69:140–6. 10.15585/mmwr.mm6905e1 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
5. Bialek S, Bowen V, Chow N, et al.; CDC COVID-19 Response Team. COVID-19 Response Team. Geographic differences in covid-19 cases, deaths, and incidence—United States, February 12–April 7, 2020. MMWR Morb Mortal Wkly Rep 2020;69:465–71. 10.15585/mmwr.mm6915e4 [PMC free article] [PubMed] [CrossRef] [Google Scholar]



Fig 7b. World map with interactive areas



Fig 7c. Data for Italian Republic

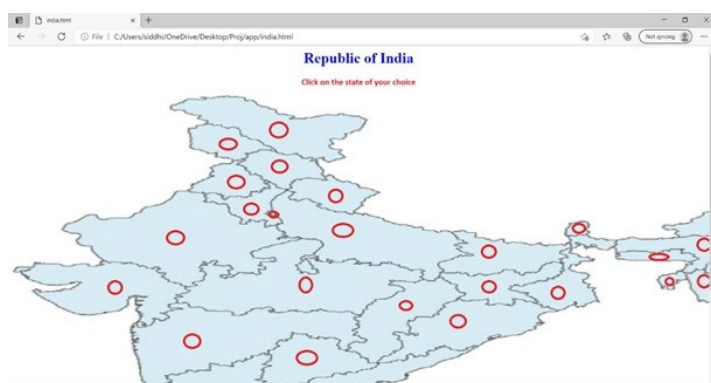


Fig 7d. Indian Map with interactive areas



Fig 7e. Data for Maharashtra

6. 6. McMichael TM, Currie DW, Clark S, et al. Epidemiology of Covid-19 in a long-term care facility in King County, Washington. *N Engl J Med* 2020;NEJMoa2005412. 10.1056/NEJMoa2005412 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
7. 7. Heinzerling A, Stuckey MJ, Scheuer T, et al. Transmission of COVID-19 to health care personnel during exposures to a hospitalized patient—Solano County, California, February 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:472–6. 10.15585/mmwr.mm6915e5 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
8. 8. Pung R, Chiew CJ, Young BE, et al.; Singapore 2019 Novel Coronavirus Outbreak Research Team. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *Lancet* 2020;395:1039–46. 10.1016/S0140-6736(20)30528-6 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
9. 9. Rocklöv J, Sjödin H. High population densities catalyze the spread of COVID-19. *J Travel Med* 2020; Epub March 29, 2020. 10.1093/jtm/taaa038 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
10. 10. Zwald ML, Lin W, Sondermeyer Cooksey GL, et al. Rapid sentinel surveillance for COVID-19—Santa Clara County, California, March 2020. *MMWR Morb Mortal Wkly Rep* 2020;69:419–21. 10.15585/mmwr.mm6914e3 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
11. 11. Gandhi M, Yokoe DS, Havlir DV. Asymptomatic transmission, the Achilles' Heel of current strategies to control Covid-19. 2020. *N Engl J Med* 2020; Epub April 24, 2020. 10.1056/NEJMe2009758 [PMC free article] [PubMed] [CrossRef] [Google Scholar]
12. Caldwell, J. and Ram, Y.M., "Study of a deterministic model in the theory of contagious disease," *Int. J. Math.Educ. Sci. Techol.* 1995.
13. Bailey, N.T.J., *The Mathematical Theory of Epidemics*, C.Griffin, London, 1957.