International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
Maths 2016; 1(6): 15-20

www.mathsjournal.com
Received: 08-09-2016
Accepted: 12-10-2016

Aniket Modi
Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS Deemed-to-be-University, V. L. Mehta Road, Vile Parle, West, Mumbai, Maharashtra, India

Kartikay Laddha
Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS Deemed-to-be-University, V. L. Mehta Road, Vile Parle, West, Mumbai, Maharashtra, India

Ujwal Shah
Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS Deemed-to-be-University, V. L. Mehta Road, Vile Parle, West, Mumbai, Maharashtra, India

Vidhi Kapoor
Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS Deemed-to-be-University, V. L. Mehta Road, Vile Parle, West, Mumbai, Maharashtra, India

Corresponding Author:
Aniket Modi
Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS Deemed-to-be-University, V. L. Mehta Road, Vile Parle, West, Mumbai, Maharashtra, India

A statistical analysis on the impact of BCG vaccination policy of different economies on coronavirus

Aniket Modi, Kartikay Laddha, Ujwal Shah and Vidhi Kapoor

Abstract
The novel coronavirus SARS-Cov-2 (Cov-2) results in COVID-19 infection which affects the respiratory tract of the human body, so we investigated whether the BCG vaccine offers some protection against CoV-2 infection. Herein we examine whether the national programs that use BCG vaccination to reduce tuberculosis infections might be responsible for different incident and mortality rates which is observed in Covid-19 among various countries.
There is some mechanistic evidence which shows that BCG vaccination can have some protective effects against the Covid-19 virus. Through our observation, we see that countries which still follow BCG vaccination policy have a significantly lower mortality rate than those countries who have discontinued this policy or never had one.

Keywords: Coronavirus, vaccination, BCG, morbidity, mortality

1. Introduction

It is the live, attenuated mycobacterium bacillus Calmette Guerin strain of M.bovis -attenuated strain of tuberculosis bacillus. Bacillus Calmette–Guérin (BCG) is a freeze-dried vaccine and is used to prevent the disseminated form of tuberculosis such as TBM and miliary TB. The protection offered by BCG vaccine against infections affecting the respiratory system is not specific and hence it is a subject of active investigation. Tuberculosis (TB) is a disease caused by Mycobacterium tuberculosis bacteria and since Bacillus Calmette-Geurin (BCG) vaccine has been highly successful in preventing it, many countries have implemented national BCG Vaccination programs for preventing this disease. This disease primarily impacts the lungs and may even lead to death. When the BCG vaccination policy was not implemented in many countries, TB-related death was very common. BCG vaccine provides immunity against TB and protection against other infectious diseases also. The BCG vaccine induces some immune cells that provide immunity against anti-mycobacterial, bacterial, fungal, and viral diseases too. Through hypothesis testing, researchers have proved that BCG offers protection against RNA and DNA viral infections and even against diseases like asthma. Hence, a possibility arises that BCG vaccine might offer protection against cov-2 virus. Cov-2 results in the disease called COVID-19 which affects the respiratory system of humans.

BCG vaccination policy has been implemented in countries that have high incidences of TB. It is given to children in early infancy. Some countries amended this policy when their TB incidences dropped. Few countries that had very few TB incidences never had the BCG vaccination policy. These groups have been categorized as discontinued/not having the BCG vaccination policy. BCG vaccination provides protection for approximately 30 to 40 years of post-vaccination. Researchers have proved that live, attenuated vaccines instigate non-specific benefits related to mortality. Therefore, it was inquired whether BCG vaccination has some kind of relation with regards to covid-19 cases and mortality rate of each country.

2. Materials and Methods

2.1 Data Collection and Description

Data for this study has been accessed and collated from three sources (as on 16th June 2020) which are mentioned below.

- **The World Bank:** The World Bank data classifies countries according to their GNI (Gross National Income) per capita in 2018 in 4 different categories.
- **World meter:** Data from world meter accounts for the information about the number of cases and deaths caused due to the spread of Coronavirus across the globe.
- **The World Atlas of BCG:** The data from the world atlas of BCG provides a list of countries describing the programs of BCG vaccination policies in their respective states.

The data was available for 168 countries from all the three sources and hence formed the basis of the analysis. Hence the number of training examples in the dataset are 168. The national programs for BCG vaccination policy currently exist in 142 countries and 26 countries have no current program of national BCG vaccination policy in their state. Due to unbalanced data, 26 countries were randomly selected from the list of nations where BCG vaccination policy still exists.
The assembled dataset for 168 countries recorded 19 features regarding the information on BCG vaccination policy and COVID19 scenario. These features have been mentioned below.

1. Country Name	2. Est. Date	3. End Date
4. Serious Critical	5. BCG Coverage	6. Life Exp
7. Total Cases	8. New Cases	9. Total Death
10. New Death	11. Total Recovered	12. Active Cases
13. Tot Cases / 1M pop	14. Deaths/1M pop	15. C/X
16. Total Tests	17. Total Tests / 1M pop	
18. Population	19. Economy Type	

Where;
C = Countries with current active BCG vaccination policy.
X = Countries with No BCG vaccination policy or had the policy sometime in past.

The data obtained from World Bank divided the countries in 4 different data groups based on the ascending order of their economy (GNI per capita). The four groups have been mentioned below:

a) Lower b) Lower Middle c) Upper Middle d) Upper

For proper sampling of data in accordance with non-vaccination countries and T-test was performed and due to unequal variance and non-normality of data, Kruskal Wallis Test has been also used to determine concrete results.

2.2 Data Distribution

The aim of this study is to examine the relationship between the 'Death/M' (Death per million people), the BCG vaccination policy and the role of economy of a nation. This study focuses on to check whether BCG vaccination policy has an impact on the Death/M caused due to the outbreak of coronavirus and does this differ across various economies around the globe.

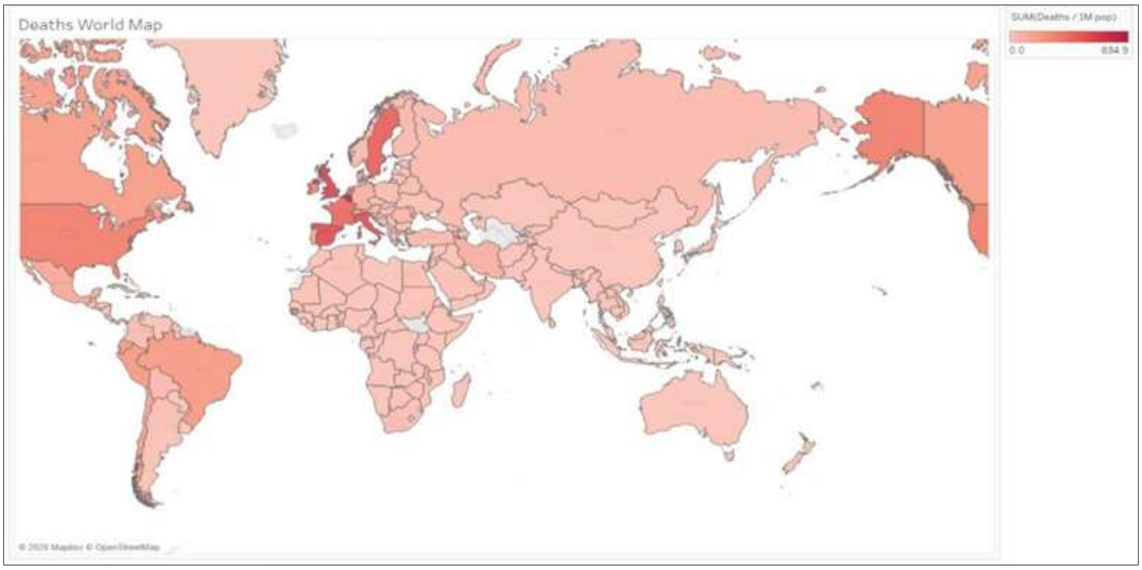

Fig 1: World Map of Deaths/M due to Coronavirus

Therefore, descriptive statistics measures such as Kurtosis and Skewness were applied to the data of 'Death/M' to understand the nature of data distribution (whether it is normally distributed or not), following is a brief about the insights these measures provide regarding the distribution of the data.

a) Skewness

Skewness is a descriptive statistic technique which measures the extent to which the disrtibution of the observed dataset under study is different/varies from a normally distributed data. It quantifies by how much the data leans leftwards or rightwards while representing its distribution when compared to a normally distributed dataset. Skewness is used along with kurtosis to better judge the likelihood of events falling in the tails of a probability distribution.

$$Sk_1 = \frac{\bar{X} - Mo}{s}$$

$$Sk_2 = \frac{3\bar{X} - Md}{s}$$

where:
Sk_1 = Pearson's first coefficient of skewness and Sk_2 the second
s = the standard deviation for the sample
$\bar{X}$ = is the mean value
Mo = the modal (mode) value
Md = is the median value

Box 1: Skewness formula
The value of Skewness for Deaths/M is 3.514306

b) Kurtosis
Kurtosis is another descriptive statistic technique, which helps in measuring the thickness of the tail of a distribution. It depicts the difference between the shape of tails of a normally distributed data and the observing dataset under study. Kurtosis focuses on the tail shape.

$$\mathrm{Kurt}[X] = \mathrm{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mathrm{E}[(X-\mu)^4]}{(\mathrm{E}[(X-\mu)^2])^2} = \frac{\mu_4}{\sigma^4},$$

where:
X = Observations
E = Expected Value
μ^4 = Fourth central moment
μ = Mean
σ = standard deviation

Box 2: Kurtosis Formula
The observed value of kurtosis for Deaths/M data is 15.93457

The following table helps in accounting the values for the measures of descriptive statistics such as kurtosis and skewness for 'Deaths/M' data:

Table 1: Results of tests for Skewness and Kurtosis

Tests	Skewness	Kurtosis
Hypothesis Ho: Null Hypothesis Ha: Alternate Hypothesis	Ho: Not Skewed Ha: Skewed (Positively/Negatively)	Ho: Data Distribution is Mesorkurtic Ha: Data Distribution is not Mesorkurtic, i.e., either Platykurtic or leptokurtic)
All Sample's Mean Values:	3.514306[a]	15.93457[a1]
Results/Conclusion	As mean skewness is greater than 0 and hence, data is positively skewed.	As the mean (coefficient of kurtosis) is greater than 3, i.e., then the data distribution is leptokurtic.

The value of skewness does not lie between (-1 to +1), which indicates substantially skewed (positively skewed) distribution of 'Deaths/M' data. This means that the majority of data values are less than mean of the distribution and are clustered towards the left tail of the distribution. The coefficient of Kurtosis for 'Death/M' data is greater than 3, which indicates that the data is leptokurtic in distribution. Both of these measures indicate that the distribution of data is highly peaked. The values of skewness and kurtosis for 'Deaths/M' data do not lie in the general domain of. (-1 to +1) and lesser than 3 respectively for skewness and kurtosis. The distributions which accounts for greater values of skewness and kurtosis, compared to the above-mentioned guidelines are considered as non-normal distributions. Hence it was concluded that the data has a non-normal distribution, which can also be seen from Figure (2).

[a] Values are calculated for the entire data set and not with respect to different samples.

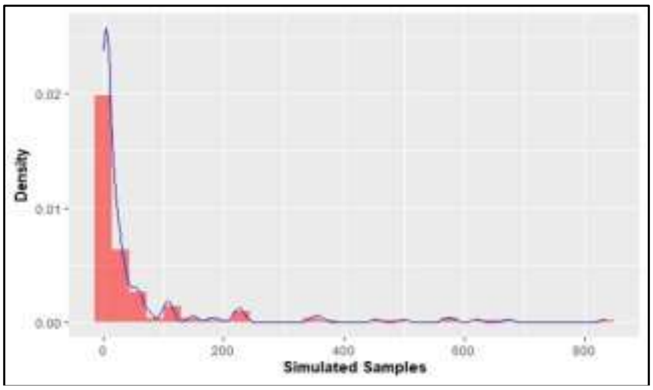

Fig 2: Non-Normal Distribution of Data

Figure (3) also proves that the data is not normally distributed as points are not centred towards the QQ-Line.

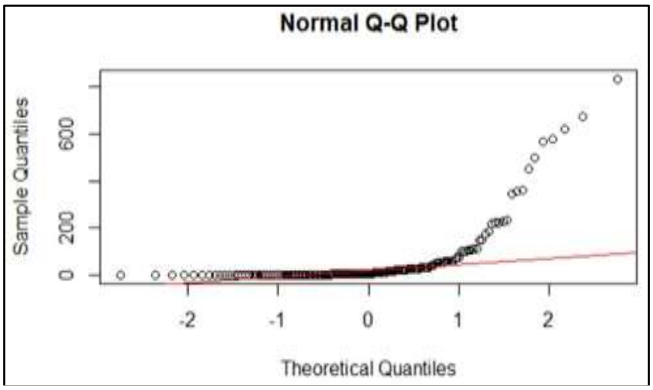

Fig 3: Normal Q-Q plot for Deaths/M

2.3 Sampling
Instead of taking the entire population altogether for the analysis, it was divided into 100 samples to cover the whole population. The other advantage sampling provides is the turnaround time for code. The accuracy of data remains high, since the scope of sampling is high. The Probability Sampling technique used for the same is the two-Stage Cluster Sampling. Cluster Sampling is a probability sampling technique wherein the researcher divides the entire population into multiple clusters. Some of these clusters are selected randomly for sampling. It is usually being employed by various researchers as it helps in Cost reduction and therefore increasing the efficiency.
Here the entire population is divided into 2 different cluster of Vaccination Policy (C & X) which are mentioned below:

- There are 142 countries in "C" Cluster
- There are 26 countries in "X" Cluster

The nature of cluster elements in a cluster sampling technique should be as heterogeneous as possible. This implies that the population should contain distinct subpopulations of various different types. Each cluster in this technique should be mutually exclusive & exhaustive collectively. It should be impossible for each cluster to occur together. C & X are two distinctive subpopulations of Vaccination Policy type. Countries in "C" & "X" clusters do not overlap with each other. So, the conditions require for a cluster sampling are fulfilled in this research.
To select the clusters which can be included in the study, a 'random sampling' technique has been used on the relevant clusters. In two-stage cluster sampling, a random sampling technique is applied to the elements also from each of the

selected clusters. For second-stage cluster sampling, a random selection of 26 countries from "C" cluster is taken into consideration. From the "X" cluster, all the countries in the cluster (Total: 26) are selected for sampling. That means in all the 100 sample, every time all "X" countries are taken in sampling. So, a total of 52 (26(C) +26(X)) are selected for sampling. This process gets repeated for all the 100 samples created.

2.4 Fligner - Killeen Test

Due to the non-normal distribution of data, as already observed by the measures of Skewness & kurtosis, a non-normal data distribution test has been performed on data to get better results for understanding the variance. To check that the variance across the Samples is the same or not, Fligner-Killeen Test of Homogeneity of Variances has been performed.

The Fligner Killeen test checks the homogeneity of group variances based on ranks. It is a non-parametric test which is useful on non-normally distributed data which contains no outliers. The median-centring version of this test is being followed.

$$FK = \frac{\sum_{j=1}^{k} n_j(\bar{a}_j - \bar{a})^2}{s^2}$$

where:

k = number of groups
n_j = size of j^{th} group
a_j bar = mean of the normalization values for j^{th} group
a bar = mean of all the normalization values
s^2 = variance of all normalization values

Box 3: Fligner-Kileen Test Formula

Fligner's Test on 'Deaths/M' data has been performed 3 times with three different independent variables such as C/X, Economy & the Interaction effect of these 2 independent variables. The results of the 3 tests which are mentioned below have been depicted. By the table below.

- Fligner's test (Deaths/M ~ C/X)
- Fligner's test (Deaths/M ~ Economy)
- Fligner's test (Deaths/M ~ ((C/X) *Economy))

Table 2: Results for Fligner-Kileen Test

Tests	Fligner (C/X) - ChiSq	Fligner (Econ) - ChiSq	Fligner (C/X vs Econ) - ChiSq	Fligner (C/X) - p	Fligner (Econ) - p	Fligner (C/X vs Econ) - p
HYPOTHESIS Ho: Null Hypothesis Ha: Alternate Hypothesis	H_0: σ^2 (1) =σ^2 (2) .... =σ^2(n) H_a: At least 1 σ^2(n) is not equal					
All Sample's Mean Values:	26.4684779	21.5665284	29.3397509	1.18E-06	0.000198634	4.10E-05
Results / Conclusion	Since all the mean P-Values are <0.05 there is no homogeneity of variances on 'C/X', 'Economy' & Interaction of both the effect for all of the 100 samples					

The p-value observed for the 3 tests mentioned above is lesser than 0.05, the same result is also obtained for all the 100 samples where p-value is obtained <0.05. Therefore, it can be now confirmed that variance across the sample is not equal & as the data is non-normal (as proved earlier from result of kurtosis & Skewness test) a parametric test cannot be performed on these samples & hence only option is an application of a non-parametric tests on these sample.

2.5 Kruskal - Wallis Test

It is important to see if there exists a significant difference between the median Deaths/M of the C & X groups to prove that BCG really has an impact on death rate between C and X categories countries. As non-Parametric tests being the only option, Kruskal-Wallis-Test has been performed. The Kruskal-Wallis test is a nonparametric (distribution free) test and is used when the assumptions of one-way ANOVA are not met. Therefore, the Kruskal-Wallis test can be used for both continuous and ordinal-level dependent variables. The Kruskal–Wallis test by ranks, Kruskal–Wallis H test or one-way ANOVA on ranks is a non-parametric method for testing whether samples originate from the same distribution. It is used for comparing two or more independent samples of equal or different sample sizes. A significant Kruskal–Wallis test indicates that at least one sample stochastically dominates one other sample. The hypothesis for this test is formulated in the. following manner.

Null hypothesis (H_o): Null hypothesis assumes that the samples (groups) are from identical populations.

Alternative hypothesis (H_a): Alternative hypothesis assumes that at least one of the samples (groups) comes from a different population than the others.

The P-value of Kruskal-Wallis test for all Sample's Mean is less than the significance level of 0.05, hence it is evident that there are significant differences between 'Deaths/M' with respect to C/X groups.

Kruskal Wallis Test is again performed on (Deaths /M) v/s (Economy) in order to analyse if the economy of a country is also a contributing factor for Deaths due to the coronavirus. Again, all Sample's Mean P-value of Kruskal-Wallis Test for economy is less than the significance level of 0.05, hence it can be concluded without any doubt that there are significant differences between Deaths /M with respect to different economies groups.

Table 3: Results for Kruskal-Wallis Test

Tests	KW (C/X) - ChiSq	KW (C/X) - p	KW(Econ) - ChiSq	KW (Econ) - p
Hypothesis Ho: Null Hypothesis Ha: Alternate Hypothesis	Ho: The distribution of Deaths/M is the same for each Vaccination Policy category. Ha: The distribution of Deaths/M is not the same for at least 1 Vaccination Policy category.		Ho: The distribution of Deaths/M is the same for each Economy category. Ha: The distribution of Deaths/M is not the same for at least 1 Economy category.	
All Sample's Mean Values:	20.0186521	2.68E-05	20.24694	0.000549169
Results / Conclusion	Since the mean p<0.05 for the samples, hence then we can conclude that there are significant differences of Deaths/M		Since the mean p<0.05 for all the samples, hence the we can conclude that there are significant differences of Deaths/M	

	between the C/X groups.	between the different economies groups.

Hence, 2 important conclusions can be made by the above analysis, which are:

Conclusion 1: BCG has an important role to play in the deaths of people in countries with respect to Covid-19

Conclusion 2: Country's economy also has an important role to play in the deaths of people in countries

2.6 Dunn's Test

For post Hoc analysis on Kruskal – Wallis test, Dunn's test on Kruskal Test is performed to give the following analysis. Pairwise-Wilcox test can't be performed on this data as this test does ranking in repeatable data values. Dunn Test is used for multiple comparisons of rank sums based on the z-statistics of the standard normal distribution. The test returns the lower triangle of the matrix that contains the p-values of the pairwise comparisons as shown in the. table below.

Table 4: Average values of Dunn's Test

Hypothesis:	Ho: The data comes from the same distribution. Ha: The data comes from the different distribution. (Therefore, any one of the distributions is more likely to draw a larger value than the other).
Result of Dunn Test:	There is no significant difference for Deaths/M between Lower, Lower Middle & Upper Middle Economies countries (As P-value >0.05) Whereas there is a significant difference for Deaths/M between Upper v/s Lower & Upper v/s Lower Middle Economies countries. (As P-value <0.05) There is no significant difference for Deaths/M between Upper & Upper Middle Economies Countries (As P-value>0.05)

The function posthoc. kruskal. dunn. test calculates the corresponding level of significance for the estimated statistics z. There is no significant difference for Deaths/M between Lower, Lower Middle & Upper Middle Economies countries (As P-value >0.05), whereas there is a significant difference for Deaths/M between Upper v/s Lower & Upper v/s Lower Middle Economies countries (As
P-value <0.05), also there is no significant difference for Deaths/M between Upper & Upper Middle Economies Countries (As P-value>0.05). The results of the test are depicted below.

Table 5: Results of Dunn's Test

Dunn's Test Mean	Lower	Lower Middle	Upper
Lower Middle	0.77924354	NA	NA
Upper	0.01661387	0.0340032	NA
Upper Middle	0.47319032	0.589139	0.1683211

3. Results and Discussion

With Kruskal Test it can be concluded that there is a significant difference in Deaths/M of countries with & without a BCG vaccination policy. So, BCG vaccination may play a significant role in enhancing immunity towards Covid-19 disease resulting in significantly less deaths for countries who are having Current BCG vaccination policy

With Post-Hoc Analysis on Kruskal Test using Dunn test, it can be concluded that Upper Economy countries have more impact on Deaths/M rate than that lower economy countries, Lower Middle & Upper Middle Economy countries don't have a significant impact on Deaths/M & their Deaths/M. Whereas Death/M of Upper Economy countries are significantly different as compared to lower & lower Middle economies countries.

Figure (4 & 5) shows the high difference of Deaths/M between countries with C & X vaccination policy using boxplot. Figure (6) shows the differences of Deaths/M between Economies of countries with C & X vaccination policy.

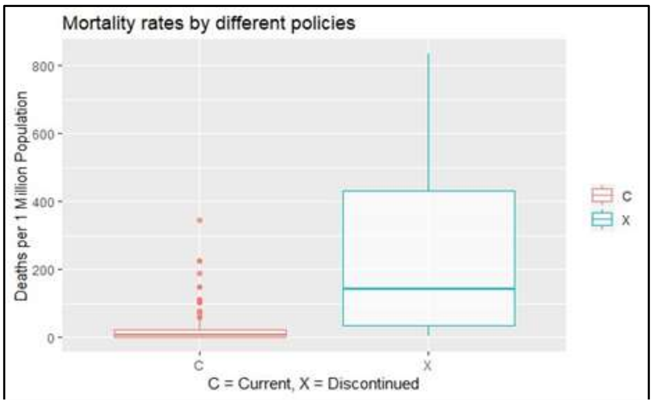

Fig 4: Box plot on Deaths/M for countries with different vaccination policies

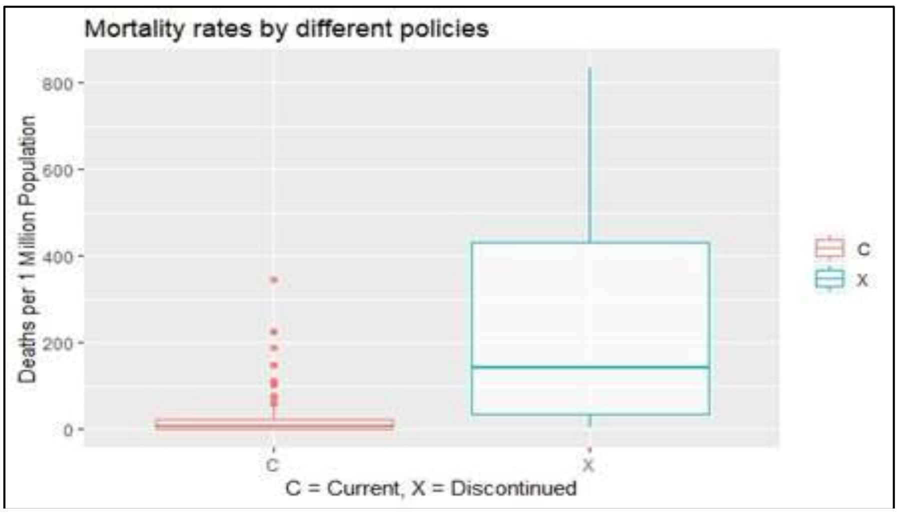

Fig 5: Difference in Deaths/M for different countries which differ in BCG vaccination policies

Fig 6: Differences in Deaths/M between economies of countries with different vaccination policy

4. Conclusion

Although the above approaches may not have a conclusive evidence that BCG Vaccination policy prevents the spread of the coronavirus or reduces the morbidity rate, but the statistical investigation above is a proof that it surely helps an affected person in recovering from the coronavirus and hence reduces the mortality rate, therefore it has been observed that there is a drop in the mortality rate for nations that have an active BCG Vaccination Policy.

5. References

1. Miller A, Reandelar MJ, Fasciglione K, Roumenova V, Li Y, Otazu GH. Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: an epidemiological study. Med Rxiv. 2020.
2. Curtis N, Sparrow A, Ghebreyesus TA, Netea MG. Considering BCG vaccination to reduce the impact of COVID-19. The Lancet 2020;395(10236):1545-1546.
3. Hensel J, McGrail DJ, McAndrews KM, Dowlatshahi D, LeBleu VS, Kalluri R. Exercising caution in correlating COVID-19 incidence and mortality rates with BCG vaccination policies due to variable rates of SARS CoV-2 testing. Medrxiv 2020.
4. Hegarty PK, Kamat AM, Zafirakis H, Dinardo A. BCG vaccination may be protective against Covid-19. Preprint. 2020.
5. Escobar LE, Molina-Cruz A, Barillas-Mury C. BCG vaccine protection from severe coronavirus disease 2019 (COVID-19). Proceedings of the National Academy of Sciences 2020;117(30):17720-17726.
6. Berg MK, Yu Q, Salvador CE, Melani I, Kitayama S. Mandated Bacillus Calmette-Guérin (BCG) vaccination predicts flattened curves for the spread of COVID-19. Medrxiv 2020.
7. Klinger D, Blass I, Rappoport N, Linial M. Significantly improved COVID-19 outcomes in countries with higher BCG vaccination coverage: A multivariable analysis. Med Rxiv 2020.
8. Galway L, Bell N, SAE A, Hagopian A, Burnham G, Flaxman A *et al*. A two-stage cluster sampling method using gridded population data, a GIS, and Google Earth TM imagery in a population-based mortality survey in Iraq. International Journal of Health Geographics 2012;11(1):12.
9. Joanes DN, Gill CA. Comparing measures of sample skewness and kurtosis. Journal of the Royal Statistical Society: Series D (The Statistician) 1998;47(1):183-189.
10. Sheskin DJ. The Kruskal–Wallis One-Way Analysis of Variance by Ranks. Handbook of Parametric and Nonparametric Statistical Procedures 2020, 1001-1026.
11. Dinno A. Nonparametric Pairwise Multiple Comparisons in Independent Groups using Dunn's Test. The Stata Journal: Promoting communications on statistics and Stata 2015;15(1):292-300.
12. AP Statistics Curriculum Non Param VarIndep-Socr. Ucla.edu 2020.
13. Pohlert T. (N. D.). The Pairwise Multiple Comparison of Mean Ranks Package (PMCMR).