Abstract

The scope of this report was to investigate Coronavirus status in seven major developed countries, i.e. G7, based on published data by World Health Organization (WHO). The G7 consists of Canada, France, Germany, Italy, Japan, the United Kingdom and the United States which represents the most advanced economies in the world. The purpose was to compare the current status of the pandemic among G7 countries individually and collectively with the rest of the world through visualization of datasets in Python Matplotlib.Pyplot library. Basic epidemiological parameters such as case and death per population in 100,00 and case fatality rate were used to determine the severity of the outbreak. Additionally, the impact of quarantine as one the controlling measurements to reduce the severity of pandemic was studied in each G7 countries through mathematical modeling in Python, Numpy and Sklearn modules

In order to analyze the effectiveness of quarantine, the datasets were modeled with before and after models of the quarantine. This report is part II of a series studies dedicated to examining the Covid-19 outbreak globally and in individual countries.

Contents

Introduction

The aim of this study was to analyze the Covid-19 outbreak in the most affluent and developed countries in the world, i.e. G7 countries. The focus concentrated on presenting reliable mathematical models for total number of infected and death toll in G7 countries. The major source of this study was reports published by World Health Organization (WHO).

The efficiency of quarantine in decreasing the intensity of pandemic was investigated in each country through modeling variables like total number of cases and deaths. It should be noted that various factors including countries' health policies and pre-existing risk factors among population can affect the pandemic situation. On the other hand, the public accessibility to detail information of these factors is limited. As a result, the models presented here can be considered as simplified versions of sophisticated models. Nonetheless, the presented primitive models confirm the effectiveness of imposing controlling measurement policies like quarantine.

Data

The WHO's daily situation reports were the main source for obtaining statistical information regarding Covid-19 outbreak. Additionally, countries' basic information including their population was obtained from Wikipedia.

Since this is an active outbreak, WHO, governments, and other health authorities modified their report formats and datasets regularly. The availability of datasets to the public often changed as the situation evolved. For instance, access to complete archive and detailed information regarding the early days of the outbreak was limited within certain governmental websites.

Moreover, investigating countries' health infrastructure, tourism sectors and border policies can be a significant key to decipher the mystery of unexpected rapid rate of hospitalization and death due to Covid-19 infection in certain communities and nations; nonetheless, the public accessibility to these data were restricted.

Methodology

The procurement of the visualization and examination from published Coronavirus related data by the WHO was performed using several Python libraries including Pandas, Numpy, Matplotlib.PyPlot, Sklearn, Tabula and PyPDF2. The procedure of this study can be summarized as follows:

1- The data from WHO PDF documents were extracted via tabula and PyPDF2 modules and stored as Pandas dataframes. Further information and details can be found in: https://github.com/HodaMYoung/Covid19/WHO1_38.py
2- The extracted dataframes were preprocessed before analytical study. Dealing with missing data, data conversion, eliminating unnecessary characters and information as the main part of the preprocessing methodology to keep the accuracy and consistency of dataframes was performed entirely in Pandas and Numpy libraries. It should be noted that formats and datasets of WHO reports have evolved during pandemic. Therefore, preprocessing played a significant role in maintaining the uniformity of the dataframe type and style.
3- The countries' basic information was obtained via scraping Wikipedia websites with BeautifulSoup module and stored as Panda dataframe. More information and detail can be found in: https://gits. github.com/HodaMYoung
4- The extracted data from Wikipedia required further revising before being able to merge the two datasets. For instance, the statistics related to pandemics for three small territories in the Caribbean were presented as a single entity in WHO reports. As a result, the datasets from Wikipedia need to be adjusted. After merging the datasets, new variables like number of cases and death per population in 100,000 for each country were determined.
5- The outbreak evolution in each G7 countries and around the world was examined and visualized by using Pandas and Matplotlib.Pyplot modules.
6- The changes in number of cases and deaths were modeled mathematically into stages: after and before quarantine.
7- Polyfit function in Numpy library was used to model the number of cases as a function of time for both stages. While modeling of number of deaths with two variables of time and number of cases for both stages was performed by using Polynomial Features, Pipeline function in Sklearn module.
8- Evaluation of models predicting total number of cases was conducted for polynomial functions with different degrees by calculating the regression coefficients and mean squared error through using Sklearn library.
9- GridSearchCV model was used to optimize the Ridge model in predicting total number of deaths.
   Further details are presented in:
   https://github.com/HodaMYoung/Covid19/G7Countries.py

Results & Discussion

The results that are presented here can be divided into three major categories: G7 vs. global community, G7 Countries, and Modeling.

G7 vs. Global Community

The G7 countries only accounts for 11% of world populations, while 33% of total confirmed cases and 43% of total death tolls globally occurred in these countries, Fig.1-3.

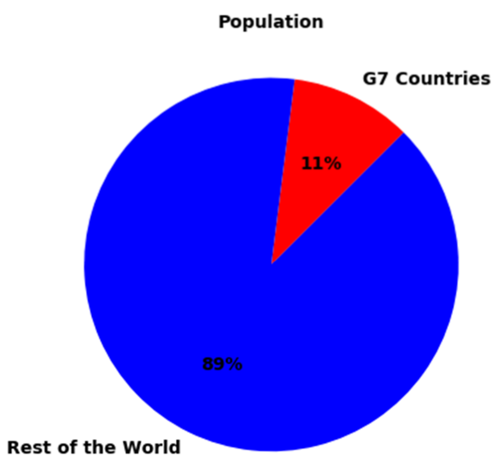


Fig.1. World population vs. the G7 population.

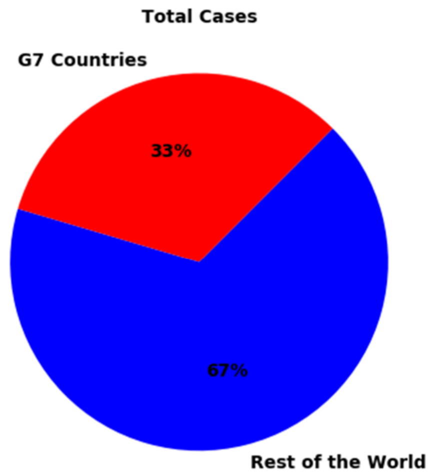


Fig.2. Total Covid-19 total cases in the G7 vs. world on 24th July 2020.

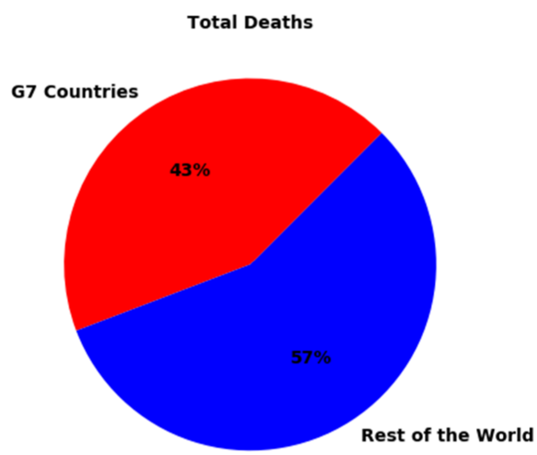


Fig. 3. Total Covid-19 total deaths in the G7 vs. world on 24th July 2020.

On July 24th, 26% of new cases and 12% of total deaths globally were reported in Canada, France, Germany, Italy, Japan, the United Kingdom and the United States, Fig.4-5.

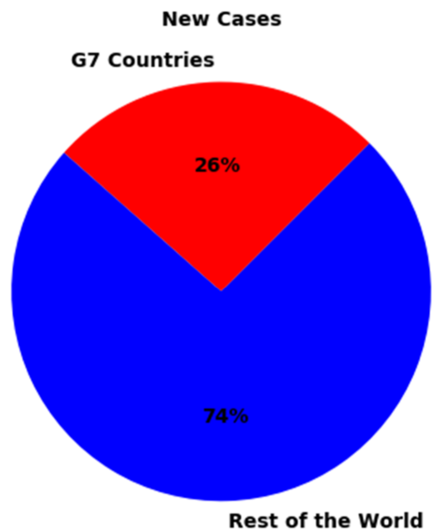


Fig. 4. Total Covid-19 daily cases in the G7 vs. world on Total Covid-19 total deaths in the G7 vs. world on 24th July 2020.

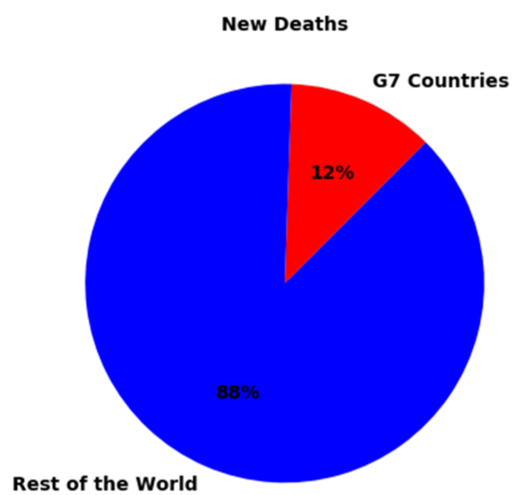


Fig. 5. Total Covid-19 daily deaths in the G7 vs. world on 24th July 2020.

The number of infected per population in G7 countries was 4 times higher than global rate with more than 650 people being affected in 100,00 population, Fig. 6. The G7 countries scored more poorly than the rest of international community in terms of number of deaths per population in 100,000 with approximately 35 deaths per its population compared to global rate (5.45), Fig.7.

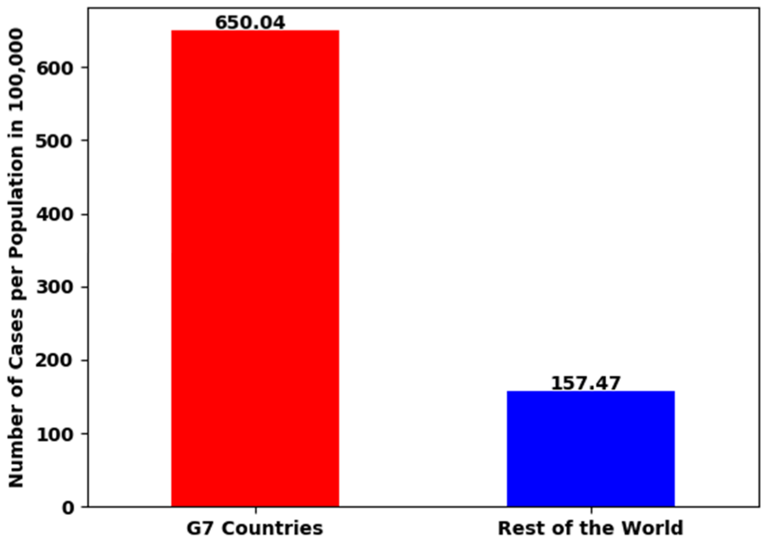


Fig. 6. Number of cases per population in 100,000 on 24th July 2020 for G7 vs. global community.

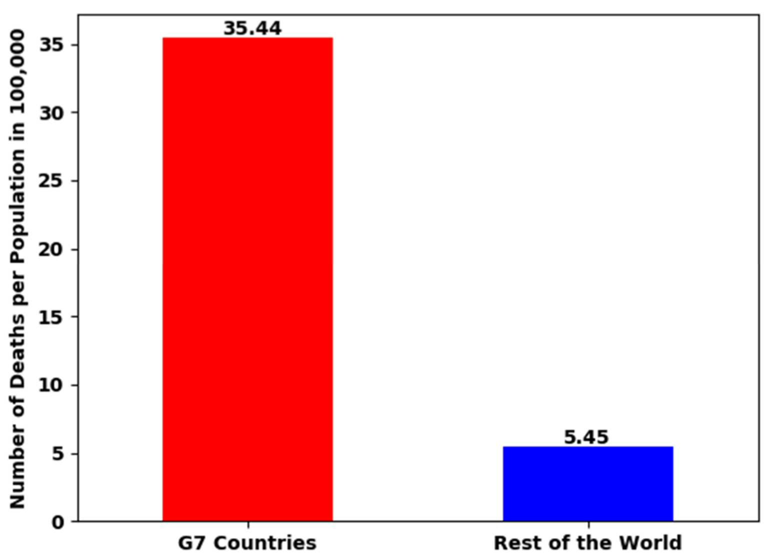


Fig.7. Number of Cases per population in 100,000 on 24th July 2020 for G7 vs. global community.

Finally, the case fatality rate (CFR) in G7 countries was also higher than global value with 5.45% while the world-wide case fatality rate was 3.46%., Fig. 8.

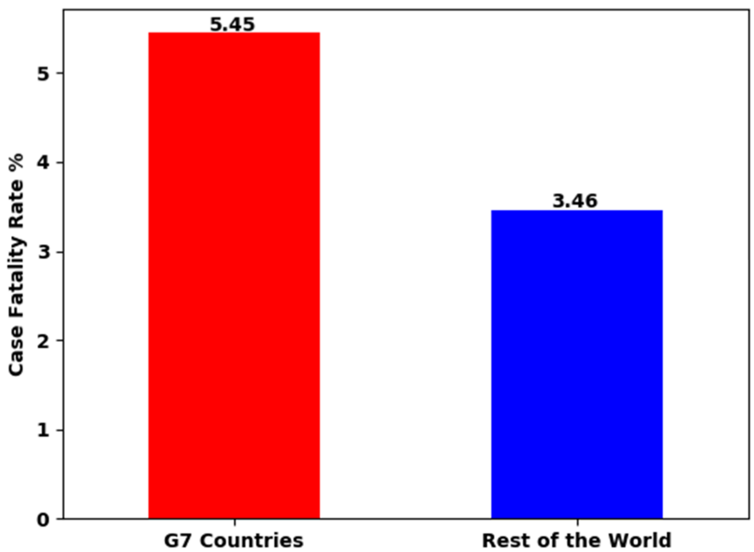


Fig. 8. Case Fatality Rate on 24th July 2020 for G7 vs. global community.

G7 Countries

The focus in this part is concentrated on the evolution and current status of Coronavirus amongst G7 countries themselves. The United States accounts for 43% of the G7's population while Japan (17%), Germany (11%), France (9%), UK (9%), Italy (8%) and Canada (5%) follow the United States, Fig. 9. The United States had the highest number of infected with 79% of the G7's infections took place in the United States, while Germany (6%) and France (5%) ranked second and third, Fig.10. As shown in Fig. 11., 52% of the G7's deaths occurred in the United States while the Japan the second most populated countries in the G7 accounts for only 0.4% of deaths.

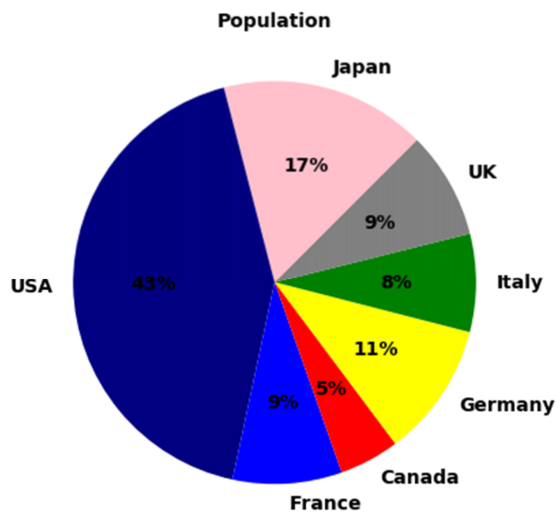


Fig. 9. Distribution of population in the G7 countries.

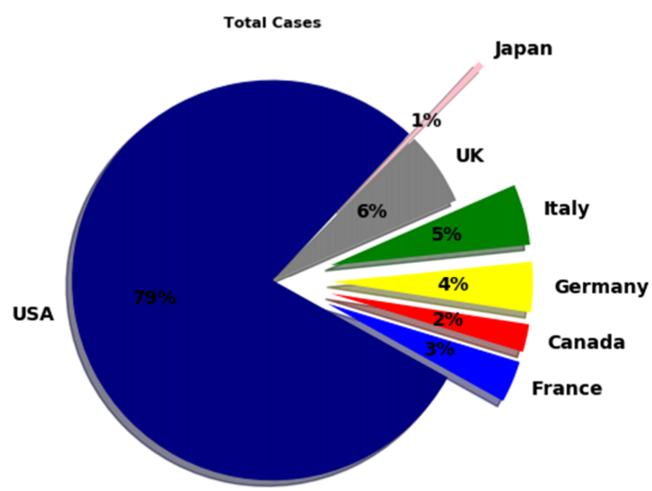


Fig. 10. Number of confirmed cases in G7 countries, on 24th July 2020.

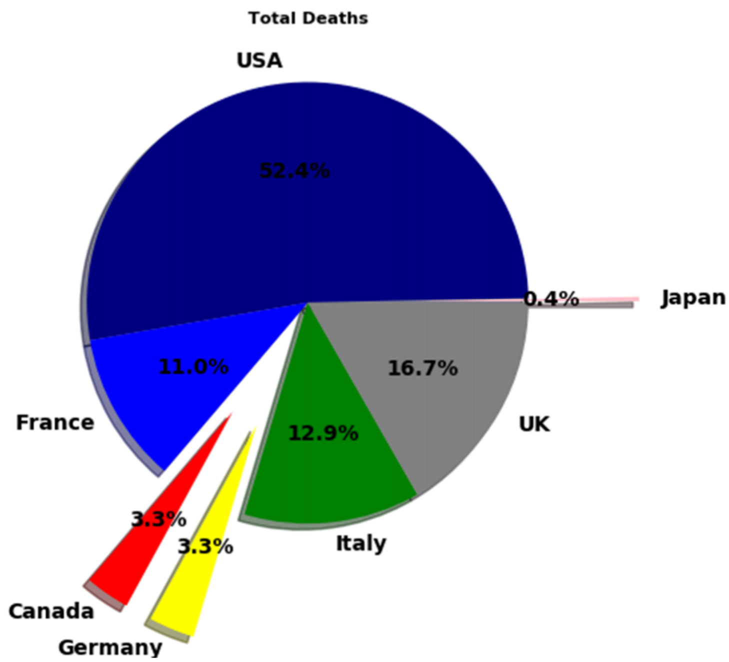


Fig. 11. Number of deaths in G7 countries, on 24th July 2020.

The changes in the number of confirmed cases and death tolls with time for each country in the G7 can be observed in Fig.12. The United States' trends clearly deviated from the rest of the G7 countries in terms of number of cases and death tolls. As a result, Fig.12 cannot demonstrate the performance efficiency of the other members of G7 in flattening the infection and death curves completely. All six remaining countries except for the United Kingdom flattened their death tolls curve successfully, Fig. 13. Additionally, the Fig.13 shows an aberrant trend for the United Kingdom where total number of cases dropped significantly once, this reflects the nature of ongoing outbreak when the authorities modify their datasets and criteria regularly.

As shown in Fig.14, the United States' number of cases per population in 100,000 over the time diverted notably from the rest of G7 countries while more people lost their lives per population in the United Kingdom than any other countries in the G7. All countries apart from the United States and United Kingdom managed to flatten the curve of deaths per population.

Finally, the changes in case fatality rate among all G7 countries can be seen in Fig.15., where France had the highest case fatality rate, and the United Kingdom and Italy ranked second and third respectively.

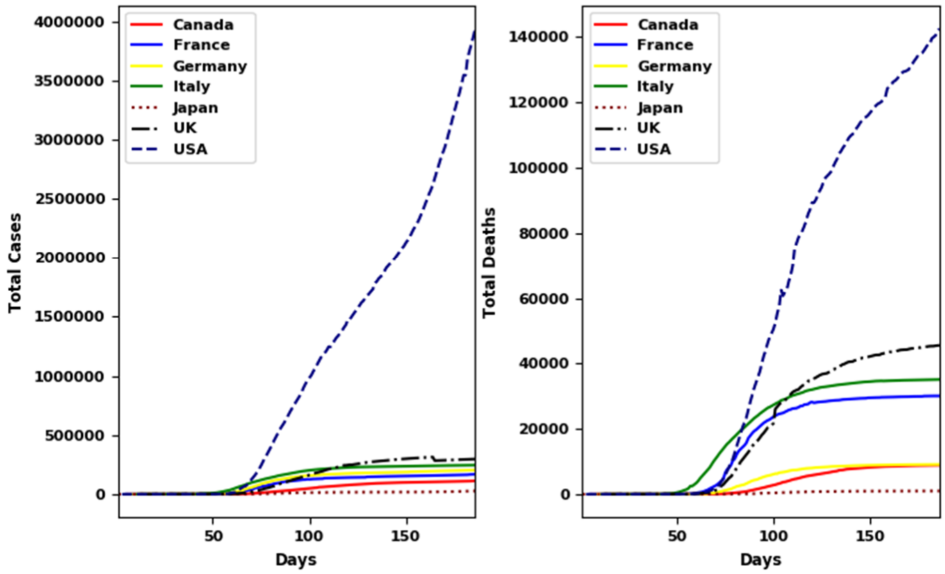


Fig. 12. Changes in the number of cases and deaths with time for G7 countries, on 24th July 2020.

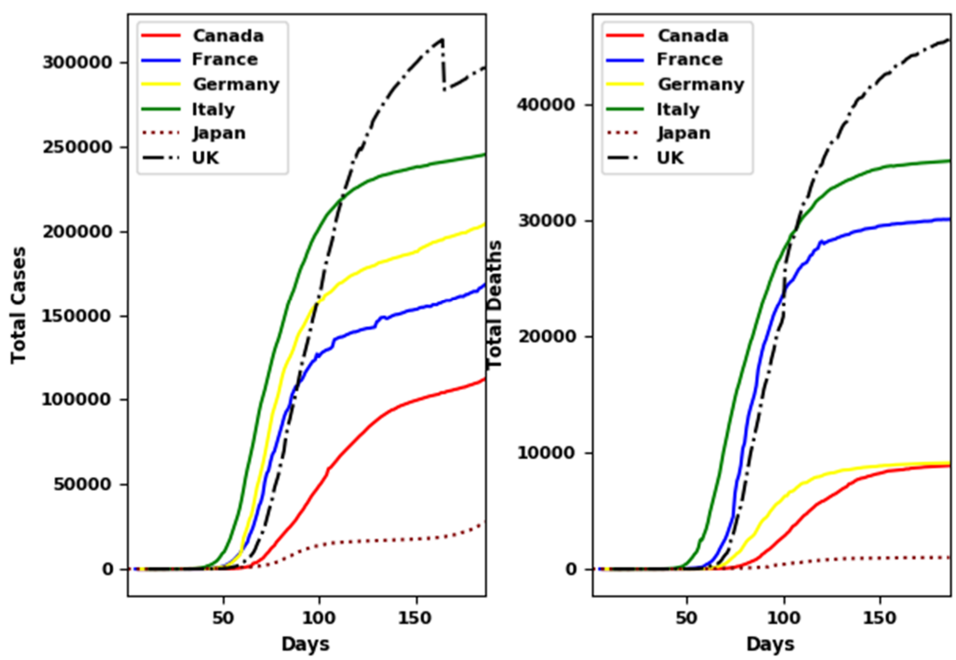


Fig. 13. Changes in the number of cases and deaths with time for G6 countries, on 24th July 2020.

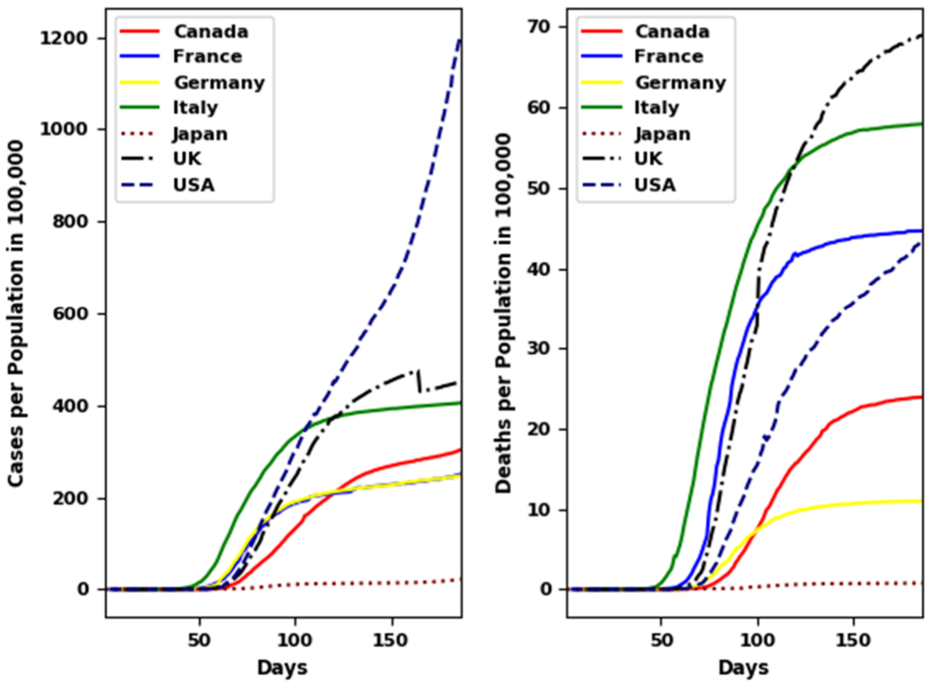


Fig. 14. Changes in the number of cases and deaths per population in 100,000 with time for G7 countries, on 24th July 2020.

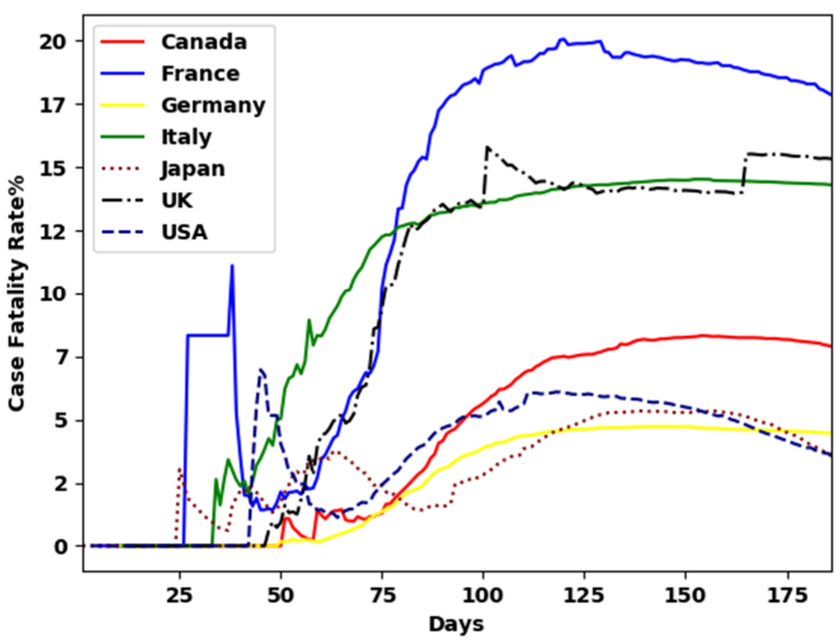


Fig. 15. Changes in case fatality rate with time for G7 countries, on 24th July 2020.

Modeling

This part of the report is dedicated to find mathematical models to represent the number of conformed cases and deaths for each country in the G7. The models present the situation of pandemic before and after imposed quarantine in each country to investigate the effectiveness of the quarantine.

Pre- Quarantine

Many countries including the G7 imposed quarantine to control the pandemic and reduce its intensity. Table 1. summarizes the start and end dates of quarantine in each country. It should be noted that the quarantine period varied in each country and often within its states/territories/provinces with existence of various versions. For instance, Italy had the strictest "Stay at Home Order" with notable fines for violation, while Japan had the mildest quarantine with no fine. The dates in the table for the United States and Canada reflect the quarantine period in the most affected state (New York), and province (Ontario).

Table 1. The quarantine period in G7 countries.

Country	Starting Date	Ending Date
Canada	30-03-2020	19-05-2020
France	17-03-2020	11-05-2020
Germany	23-03-2020	10-05-2020
Italy	09-03-2020	04-05-2020
Japan	16-04-2020	25-05-2020
UK	23-03-2020	01-06-2020
USA	22-03-2020	11-05-2020

The variation of total number of cases with time was represented by a polynomial in this first stage. Optimization process conducted by evaluating the coefficient of regression (R2) and mean squared error (MSE) to find the best degree of polynomial for fitting the data. It should be noted that due to the small size of datasets in this stage, the train and test selection models were not used. The criteria to choose the best polynomial for total cases in Canada was demonstrated in Fig.16. The best fitting occurred when the R2 value was close to 1 and MSE value dropped significantly. As a result, a polynomial with degree 5 was selected to model the total number of cases in pre-quarantine era for Canada. Similar patterns can be observed for other members of G7, Fig.17.

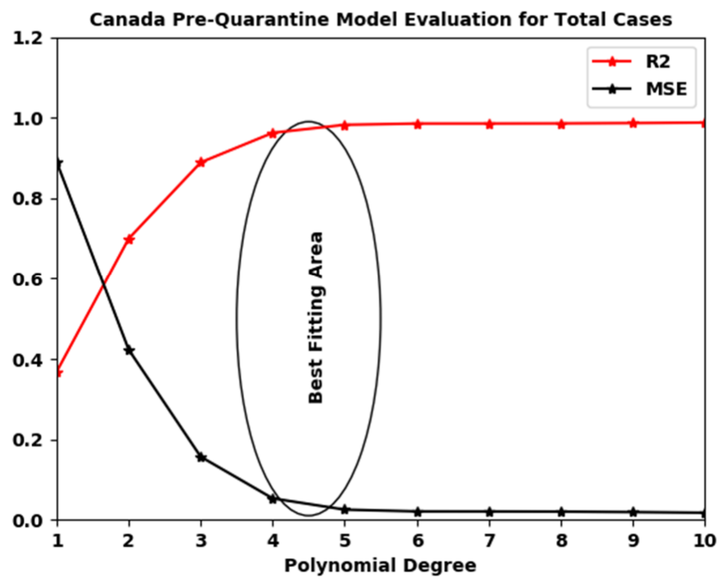


Fig. 16. Canada pre-quarantine model evaluation for total cases.

Polynomial functions with degree 5 were selected to represent the total number of cases with time for all countries except the United States. The details and mathematical formulas of polynomials are summarized in Table 2, where *x* denotes the time.

Table 2. The pre-quarantine mathematical models for number of total confirmed cases in G7 countries.

Country	**Pre-Quarantine Formula for Total Cases**
Canada	$p(x) = 0.000191x^5 - 0.01901x^4 + 1.131x^3 - 30.68x^2 + 370.5x - 1547$
France	$p(x) = 0.000263x^5 - 0.03189x^4 + 1.437x^3 - 29.42x^2 + 266.3x - 820.8$
Germany	$p(x) = 0.001159x^5 - 0.1761x^4 + 10.06x^3 - 256.4x^2 + 3176x - 13530$
Italy	$p(x) = 0.000363x^5 - 0.03488x^4 + 1.281x^3 - 22.47x^2 + 188.5x - 604.4$
Japan	$p(x) = 0.0000247x^5 - 0.00410x^4 + 0.2527x^3 - 6.392x^2 + 62.2x - 144$
UK	$p(x) = 0.000269x^5 - 0.04225x^4 + 2.552x^3 - 73.35x^2 + 994.5x - 5049$
USA	$p(x) = 0.0000366x^6 - 0.006259x^5 + 0.4109x^4 - 12.98x^3 + 211.5x^2 - 1392x + 3141$

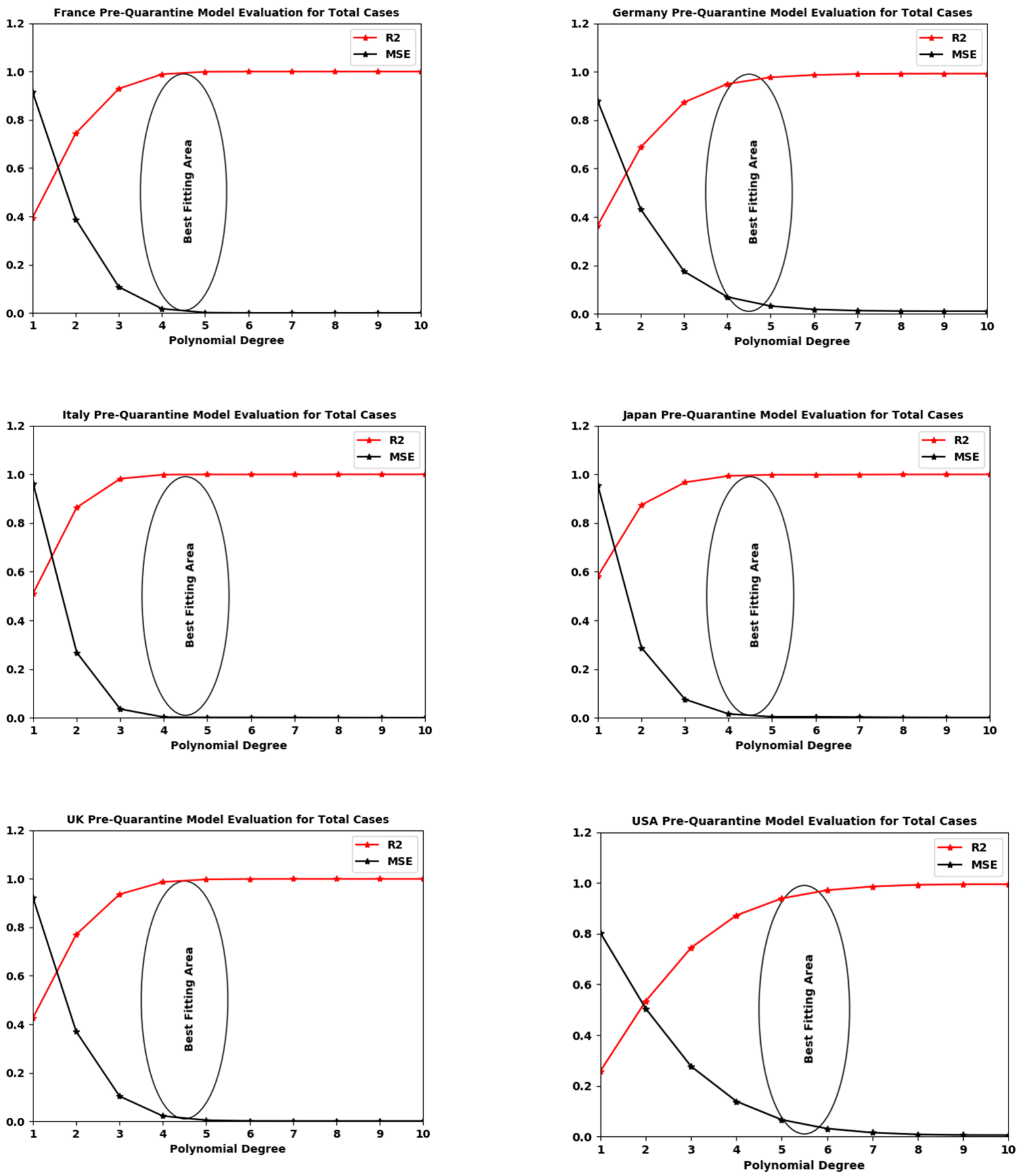


Fig. 17. Pre-quarantine model evaluation for total cases for G7 countries except for Canada.

Fig. 18. depicts the impact on quarantine on number of total cases in Canada, where the hatched area represents the effectiveness of quarantine to curb the number of infected. Similar trends can be observed for remaining G7 countries, Fig.19. It should be noted that these are simplified versions of sophisticated problems with several dependent and independent variables.

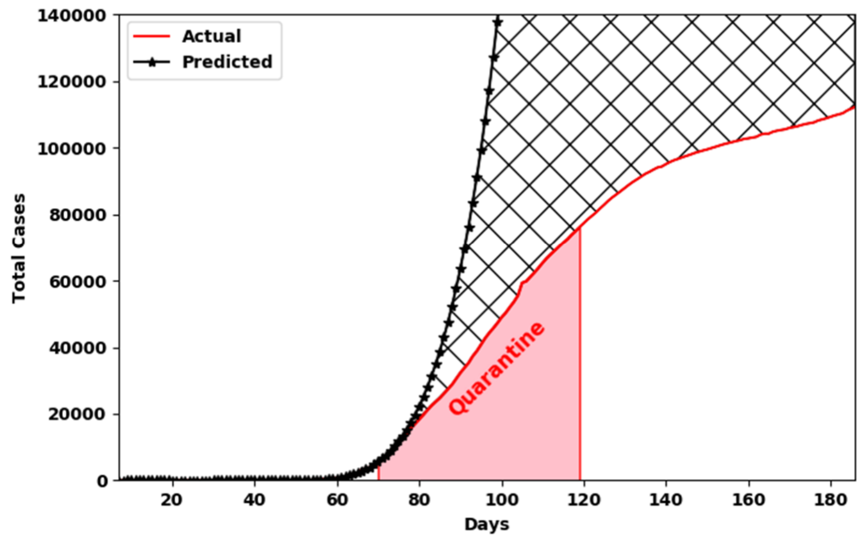
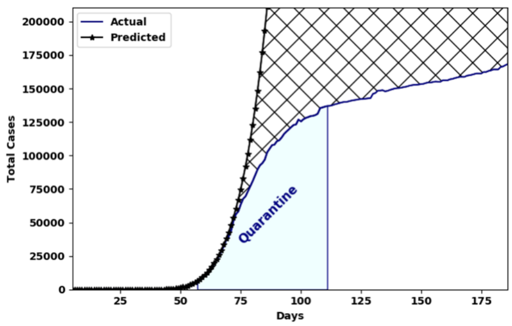


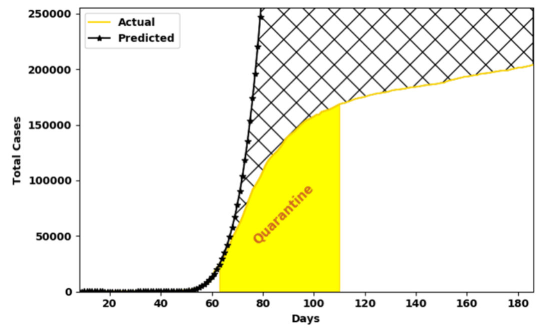
Fig. 18. The impact of quarantine on total number of confirmed cases for Canada.

Similar procedures were applied to obtain pre-quarantine models for total number of deaths. However, the number of total deaths was dependent on two variables: time and number of infected. In this part the Pipeline feature from Sklearn library was used to simplify the steps involving in searching for best fit. Fig. 20. demonstrates the criteria for choosing the best polynomial function to predict the number of deaths in Canada. The change in the values of MSE and R2 for the rest of the G7 countries can be seen in Fig.21.

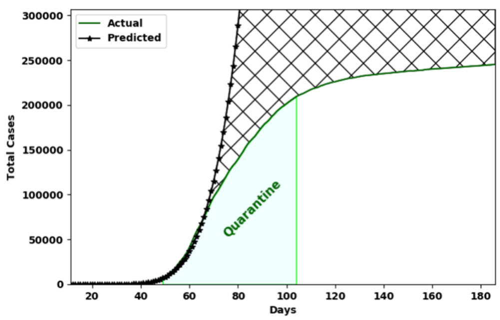
Fig. 22. demonstrates the impact of quarantine on flattening the total number of deaths for Canada, France, Germany and Japan. The hatched area in the graph show the differences between predicted death toll before quarantine and actual death tolls. The predicted death tolls based on pre-quarantine trends in Italy, the United Kingdom and the United States were significantly lower than actual deaths tolls. This indicates the highly non-linear relationship between variables and oversimplification of presented models.



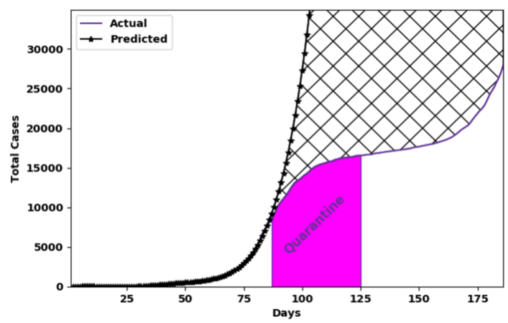
A)



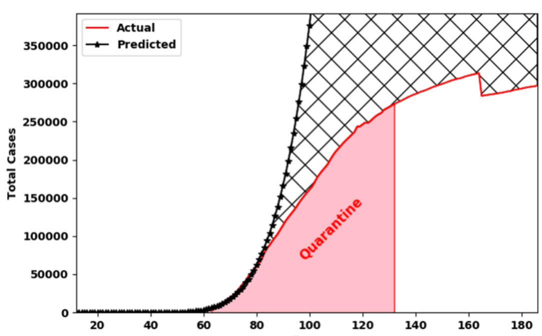
B)



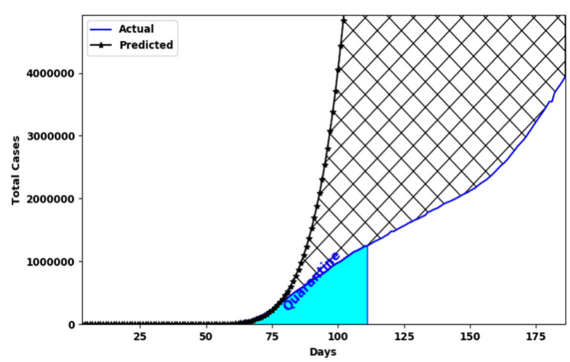
C)



D)



E)



F)

Fig. 19. The impact of quarantine on total number of confirmed cases in G7 countries, A) France, B) Germany, C) Italy, D) Japan, E) United Kingdom and F) United States.

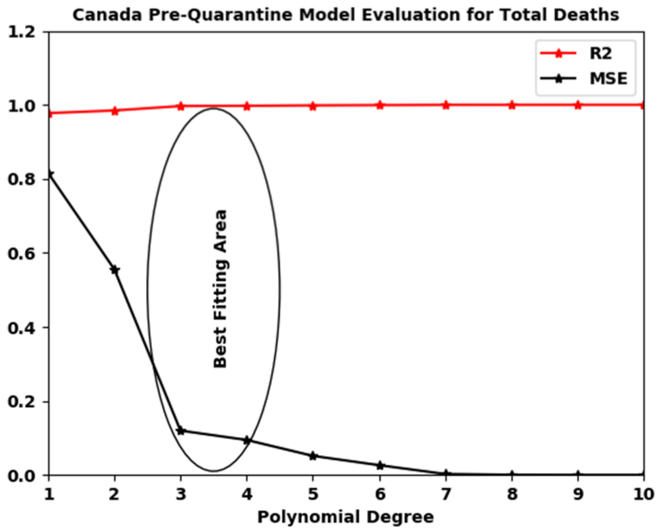


Fig. 20. Canada pre-quarantine model evaluation for total deaths.

The daily changes in number of total confirmed cases and total deaths for Canada can be seen in Fig. 23-24. The hatched area represents the quarantine period, where the highest number of confirmed new cases and new daily deaths took place during the quarantine. The changes in new daily cases and deaths for other members of G7 are presented in Fig.25-26. The highest daily number of confirmed cases in France, Italy and the UK occurred during quarantine period, while Japan and the United Kingdom reported their highest daily cases before quarantine period. The United States continuously recorded higher daily numbers after quarantine. The highest daily number of deaths took place during quarantine period in all G7 countries.

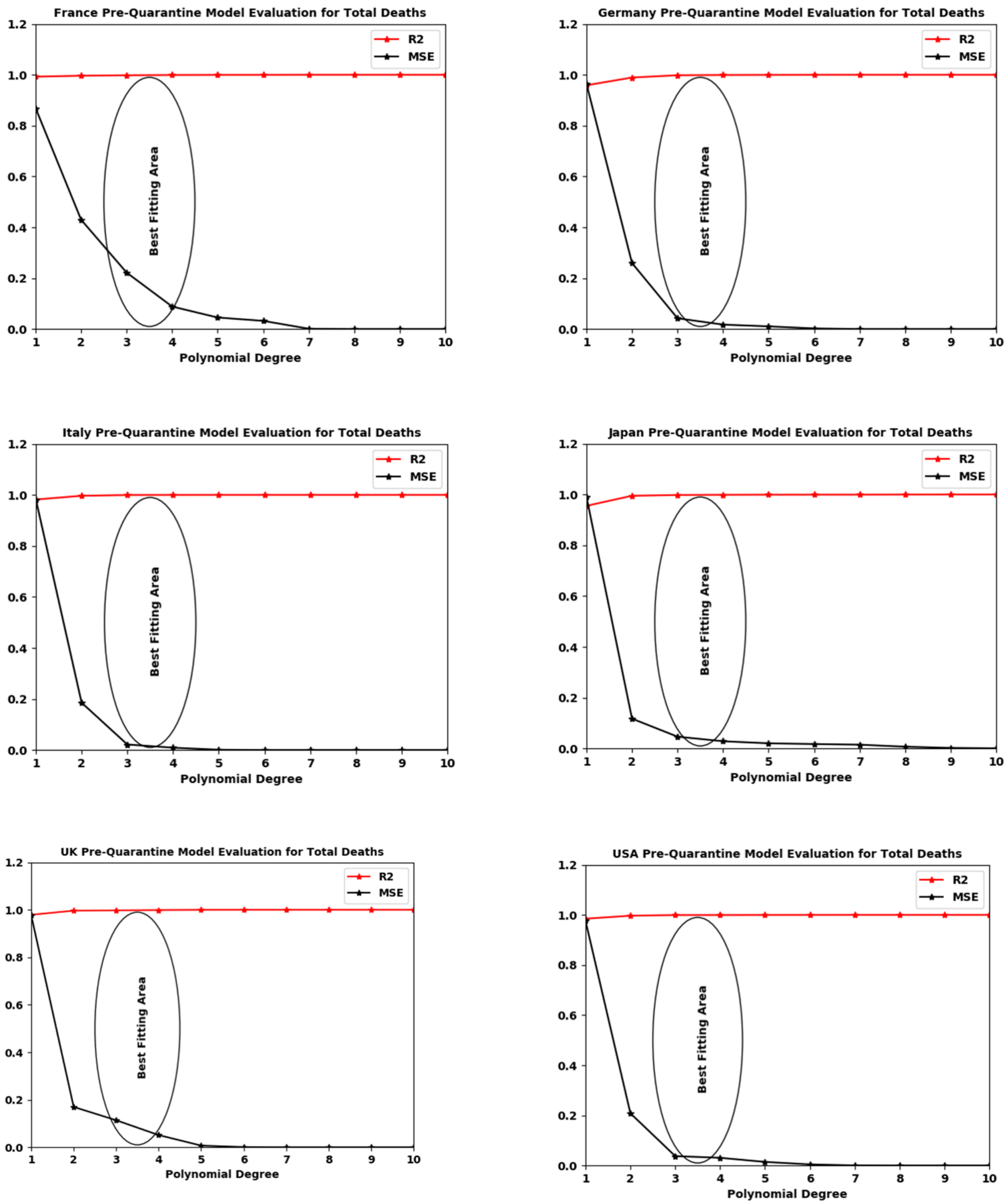
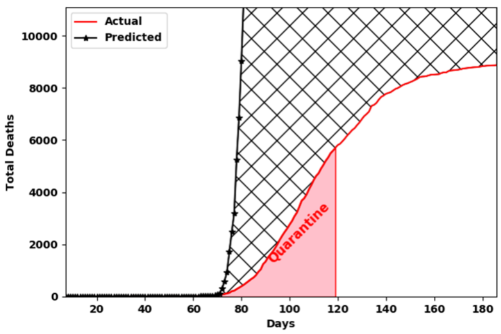
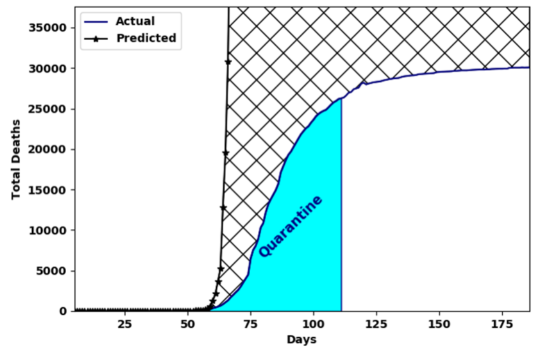


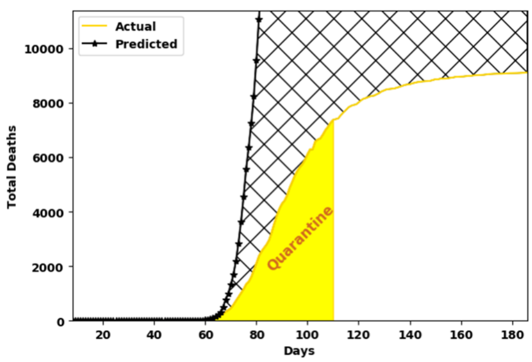
Fig. 21. Pre-quarantine model evaluation for total deaths for G7 countries except for Canada.



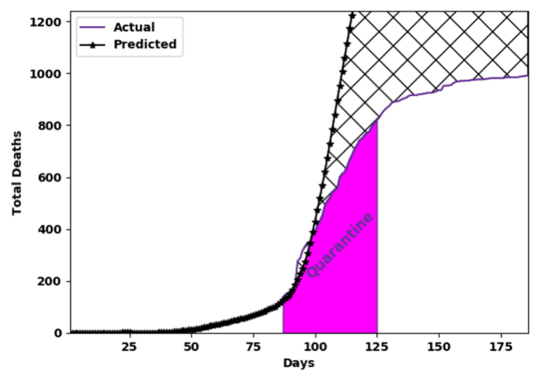
A)



B)



C)



D)

Fig. 22. The impact of quarantine on total number of deaths in A) Canada, B) France, C) Germany, and D) Japan.

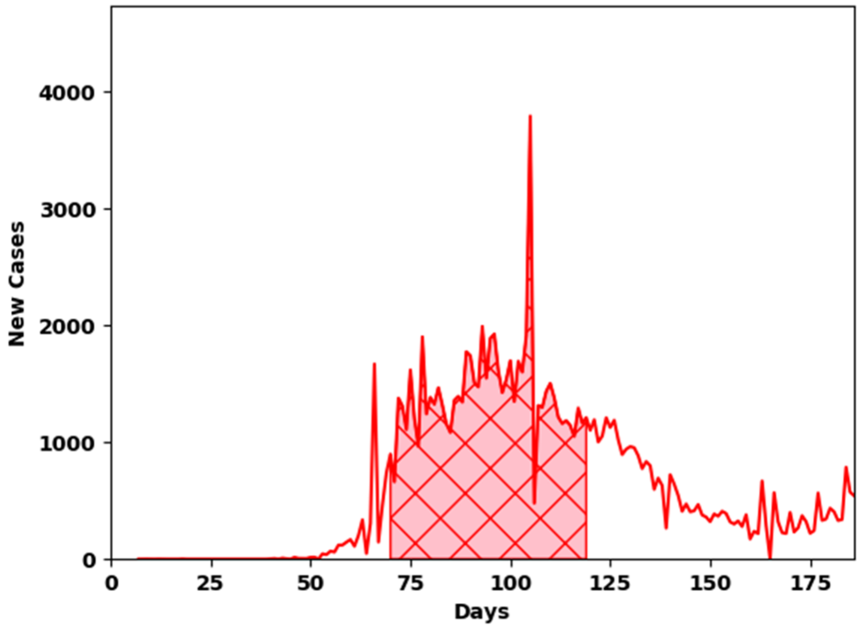


Fig. 23. Canada daily confirmed new cases.

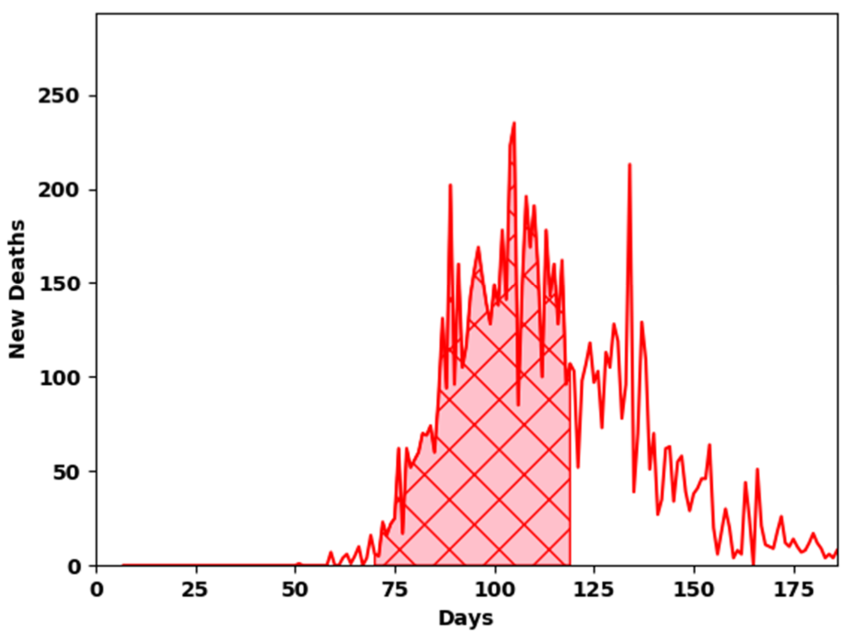
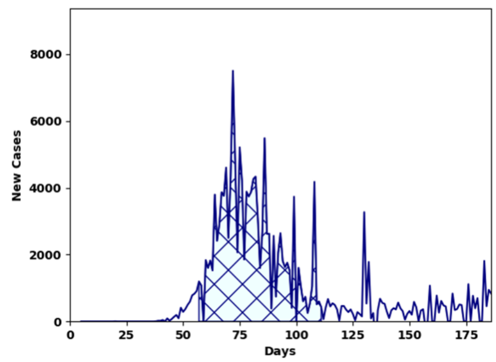
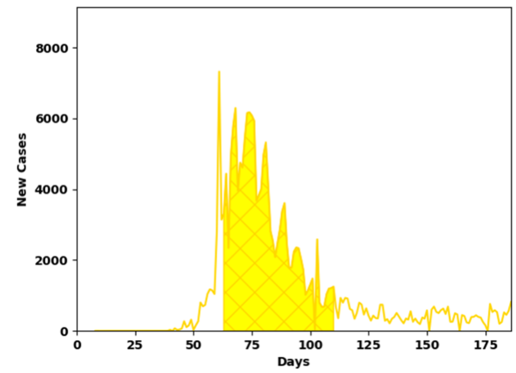


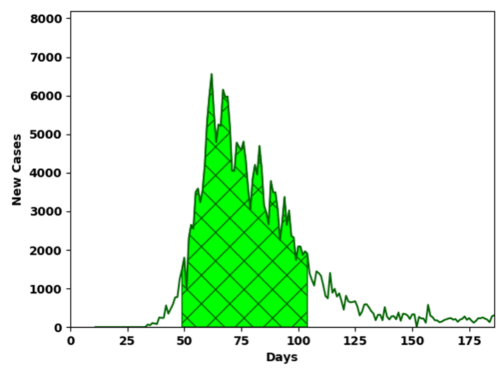
Fig. 24. Canada daily new deaths.



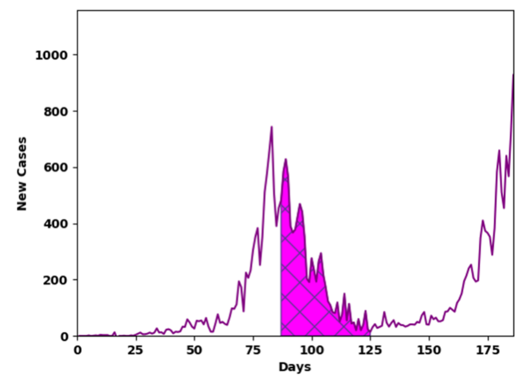
A)



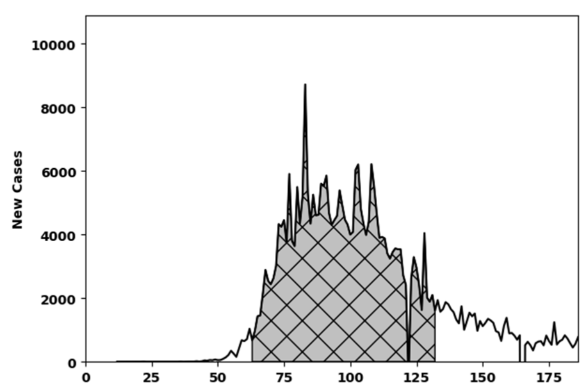
B)



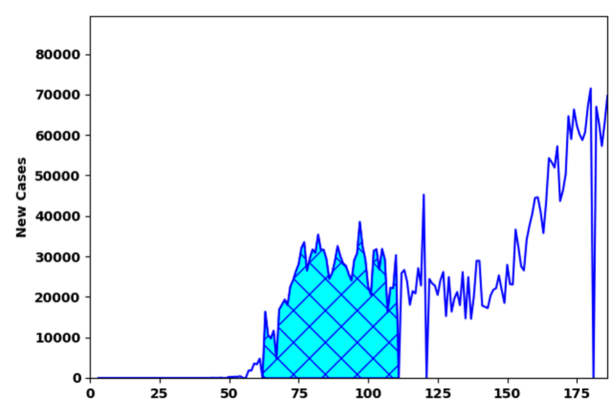
C)



D)



E)



F)

Fig. 25. Daily confirmed cases in G7 countries, A) France, B) Germany, C) Italy, D) Japan, E) United Kingdom and F) United States.

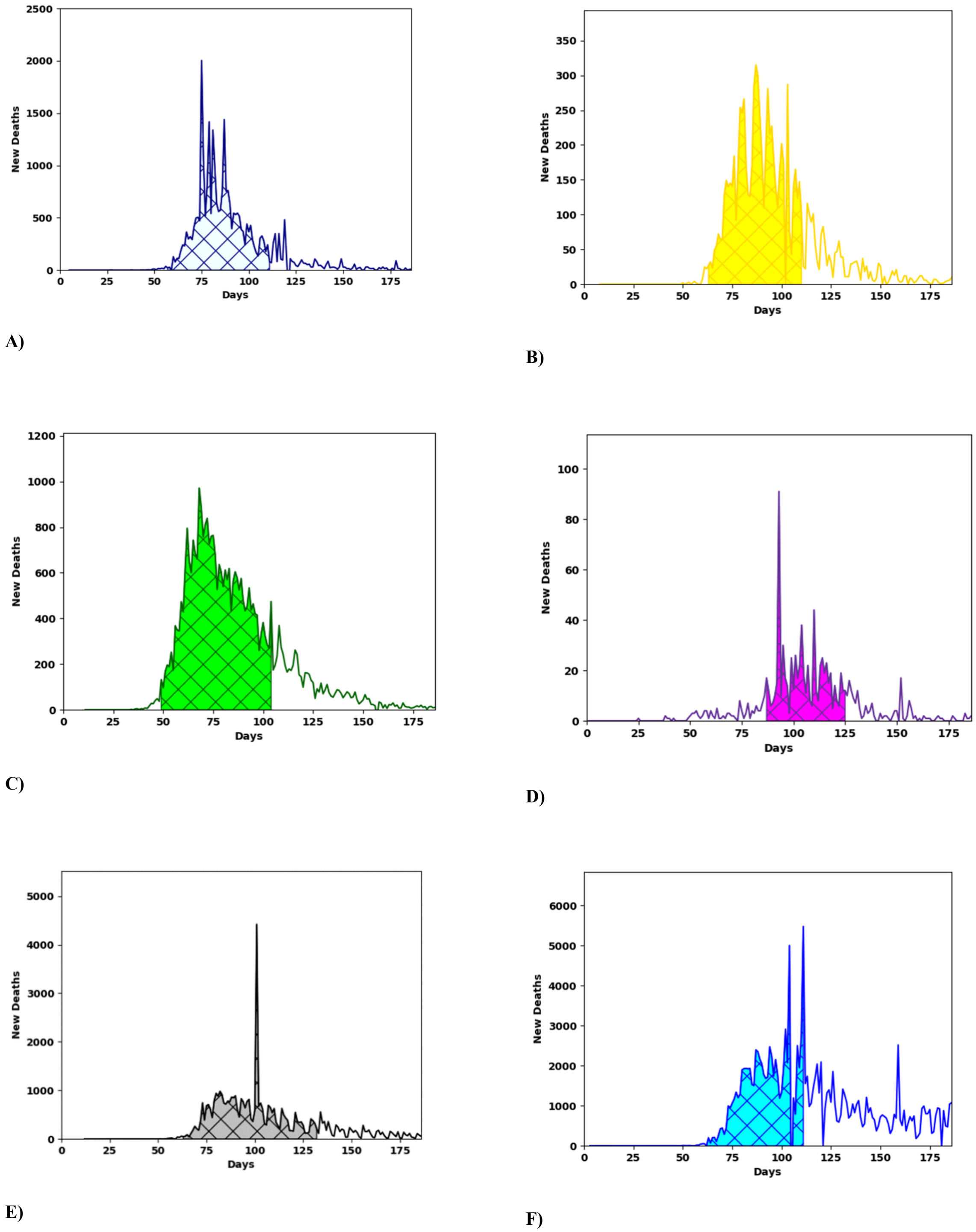


Fig. 26. Daily confirmed cases in G7 countries, A) France, B) Germany, C) Italy, D) Japan, E) United Kingdom and F) United States.

Post- Quarantine

The trends of total confirmed cases and deaths after quarantine imposed in each the G7 countries were presented by polynomial functions. Similarly, the total confirmed cases were defined as functions of time while the total deaths were determined as functions of total confirmed cases and time. In the post-quarantine era train-test selection models were used to increase the accuracy of the models. Ridge function and GridSearchCV functions were also used to find the best fit for the number of total deaths.

For the number of confirmed cases 30% of data was allocated for testing and the rest for training. The changes in R2 and MSE for both test and training datasets with the degree of polynomial for Canada to model total confirmed cases can be observed in Fig. 27. A great agreement between values of R2 and MSE for training and testing was found for all G7 countries, except for Japan and USA with the slight deviation for low degree polynomial, Fig.28.

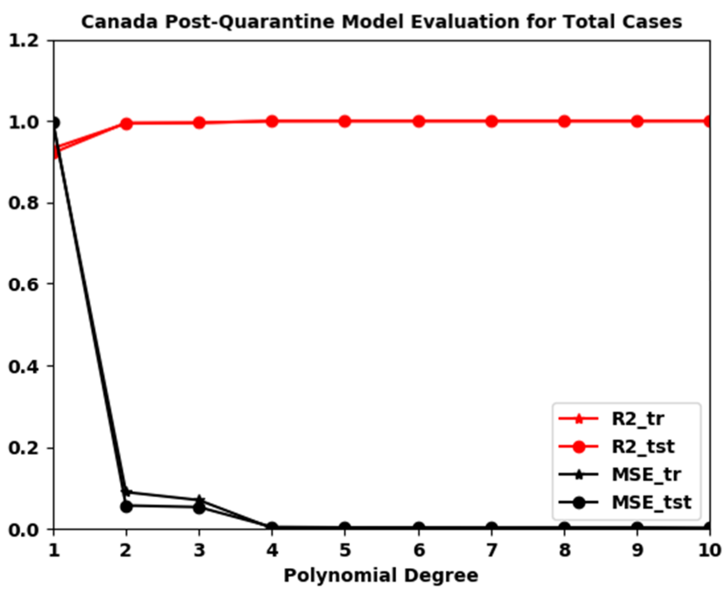


Fig. 27. Canada post-quarantine model evaluation for total cases.

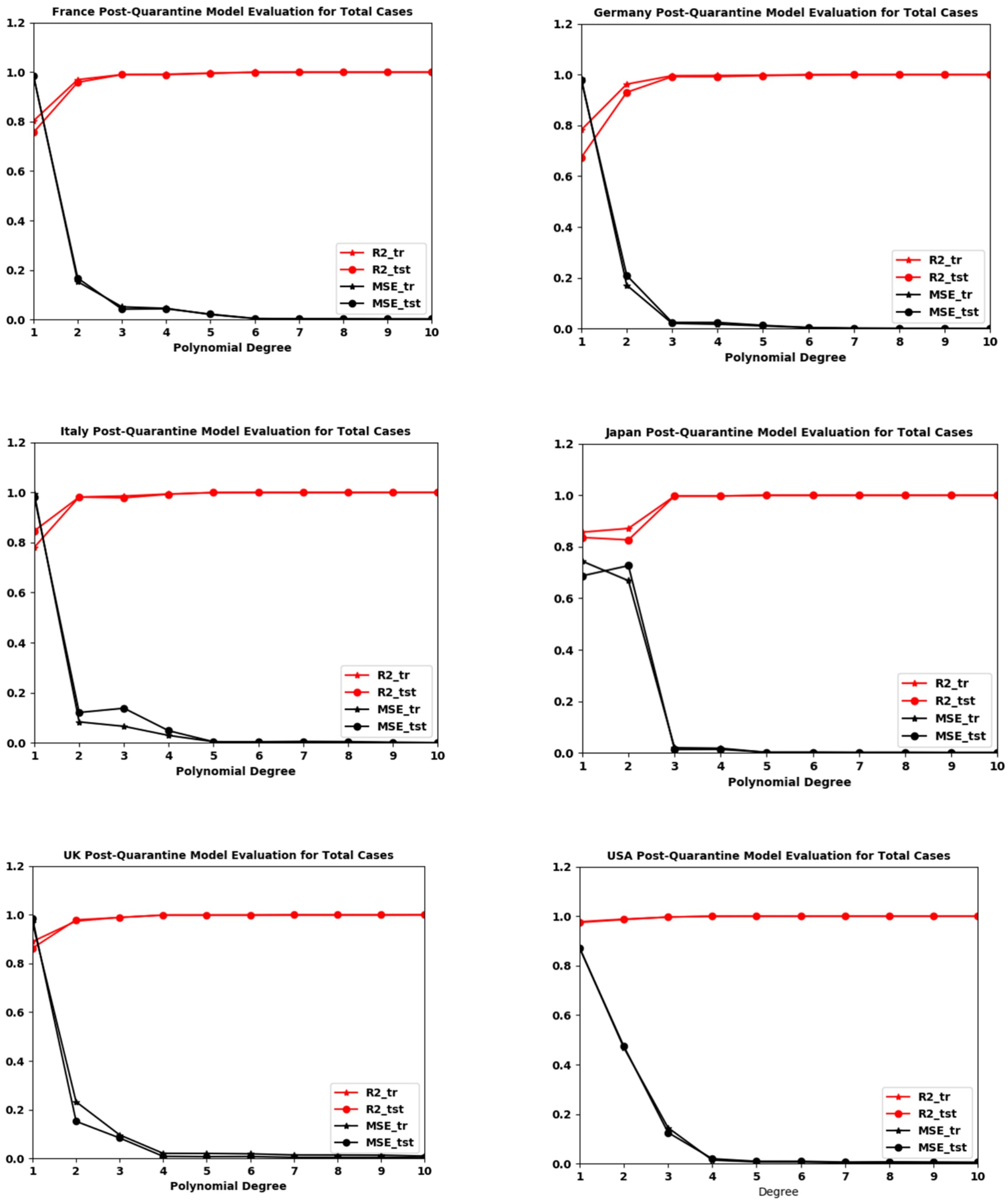


Fig. 28. Post-quarantine model evaluation for total cases for G7 countries except for Canada.

The details and mathematical formulas of polynomials are summarized in Table 3, where *x* denotes the time. Polynomial functions with different degrees were selected for each country in the G7 group.

Table 3. The post-quarantine mathematical models for number of total confirmed cases in G7 countries.

Country	Pre-Quarantine Formula for Total Cases
Canada	$p(x) = 0.002182x^4 - 1.122x^3 + 201.2x^2 - 13720x - 312500$
France	$p(x) = 0.0000017x^6 - 0.001311x^5 + 0.4008x^4 - 63.33x^3 + 5400x^2 - 231400x + 3873000$
Germany	$p(x) = 0.0000016x^6 - 0.001219x^5 + 0.3733x^4 - 58.82x^3 + 4970x^2 - 208300x + 3360000$
Italy	$p(x) = 1.97e - 09x^7 - 9.351e - 07x^6 + 1.6323e - 05x^5 + .06887x^4 - 16.95x^3 + 1729x^2 - 7.6e + 04x + 1.213e + 06$
Japan	$p(x) = 0.000016x^5 - 0.01121x^4 + 3.051x^3 - 418.3x^2 + 28930x - 791100$
UK	$p(x) = 4.853e - 08x^7 - 3.945e - 05x^6 + 0.01337x^5 - 2.437x^4 + 256.1x^3 - 15380x^2 + 487800x - 6319000$
USA	$p(x) = -1.722e - 07x^6 - 0.00021x^5 + 0.2214x^4 - 65.97x^3 + 8739x^2 - 509200x + 1.065e + 07$

The predicted values of total number of confirmed cases versus the actual reported values for Canada were plotted, Fig. 29. The performance of mathematical models to predict the changes in values of confirmed cases with time for all remaining countries can be observed in Fig. 30. All predicted values are in a great accordance with the reported values. Even with the abnormality in the datasets the mathematical models captured the nature of the relationship between number of cases and time (c.f. 30. E).

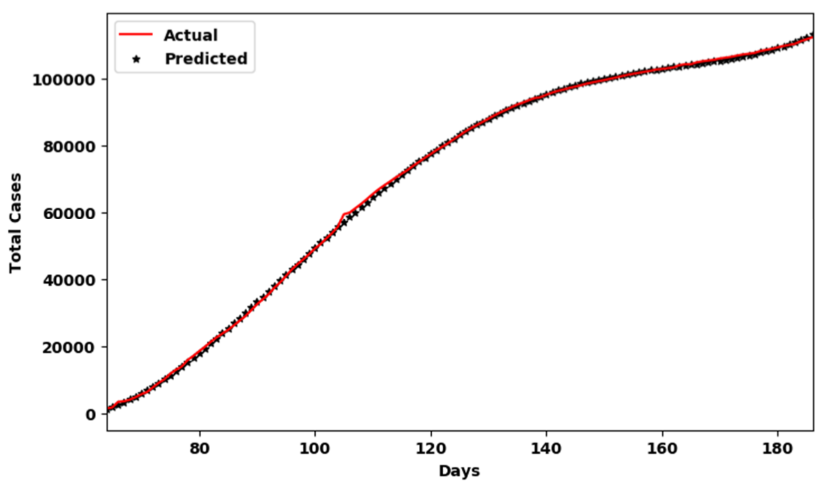
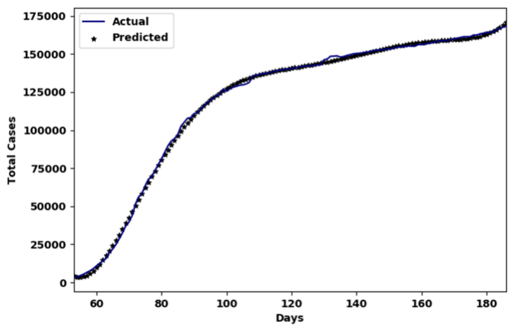
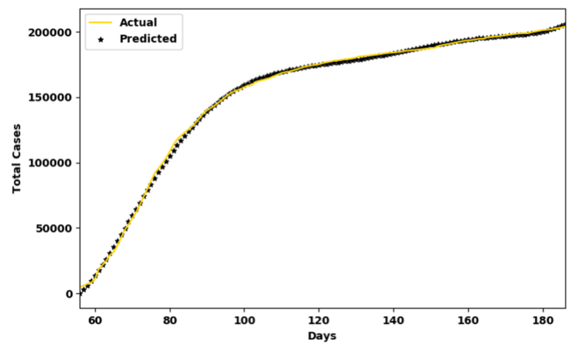


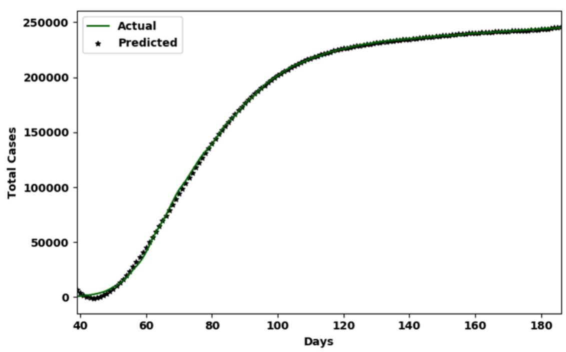
Fig. 29. The predicted number of confirmed cases vs. actual reported values for Canada.



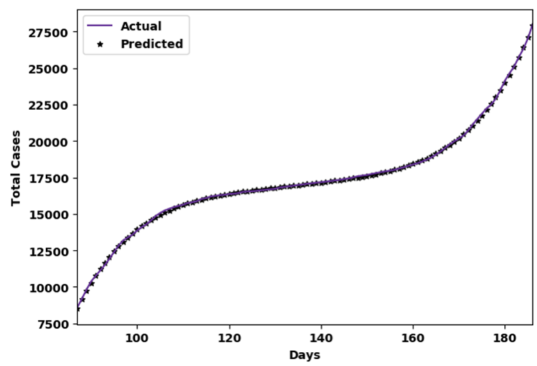
A)



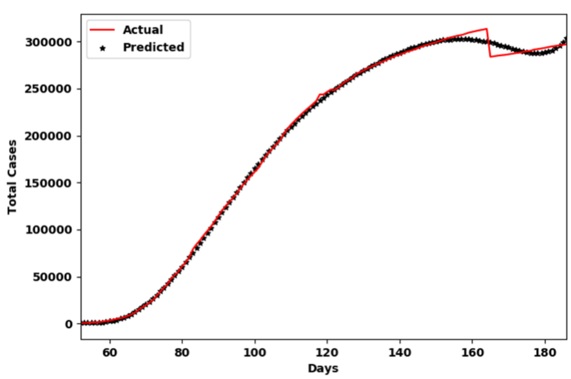
B)



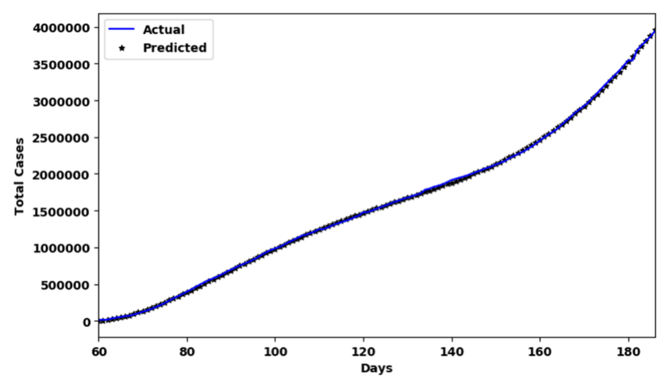
C)



D)



E)



F)

Fig. 30. The predicted number of confirmed cases vs. actual reported values in G7 countries, A) France, B) Germany, C) Italy, D) Japan, E) United Kingdom and F) United States.

The performance of mathematical models to predict the number of deaths was examined in Fig.31-32. The proposed mathematical models captured the overall nature of changes in the number of deaths with time. However, the predicted values deviated from actual reported values in early stages.

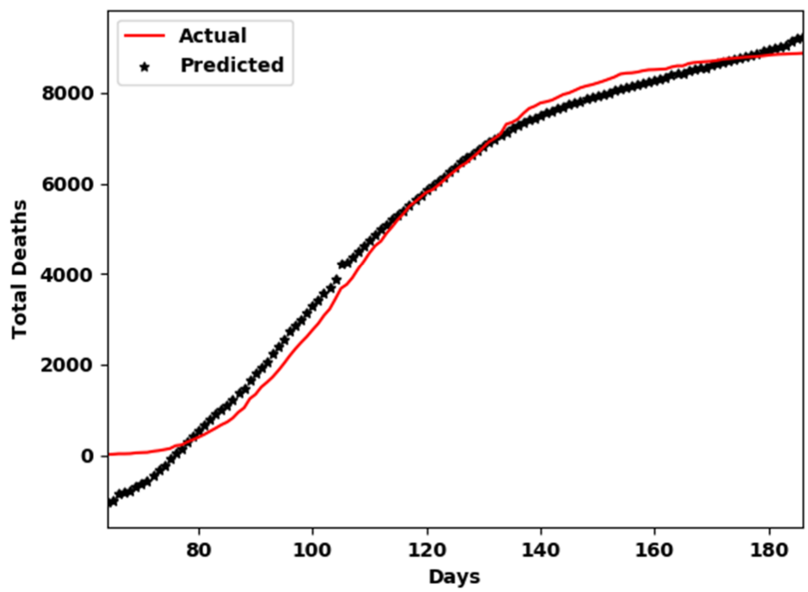
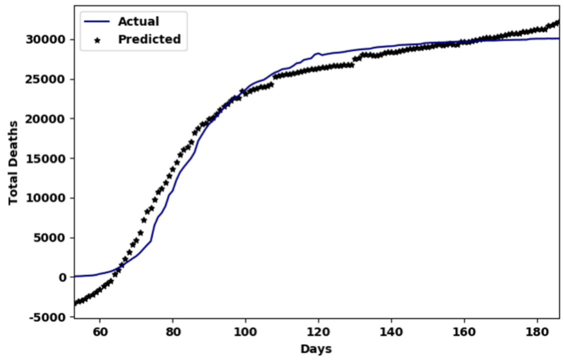
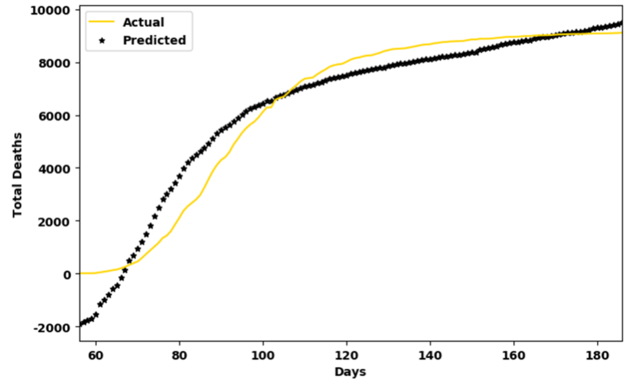


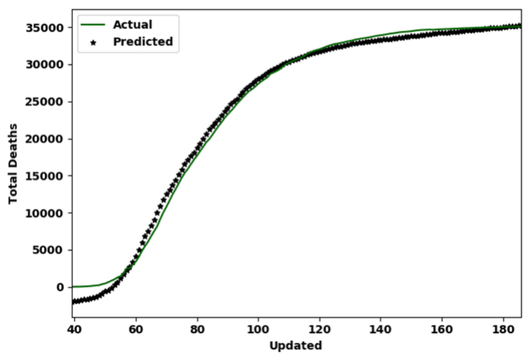
Fig. 31. The predicted number of deaths vs. actual reported values for Canada.



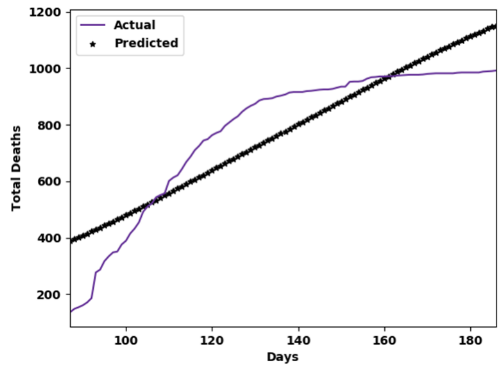
A)



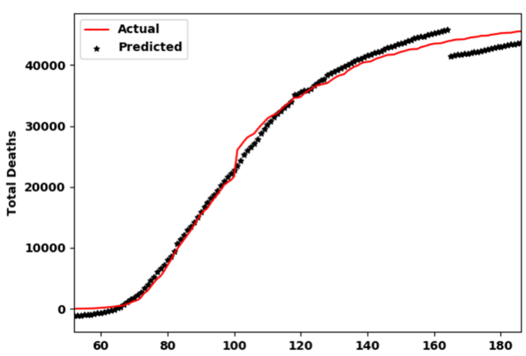
B)



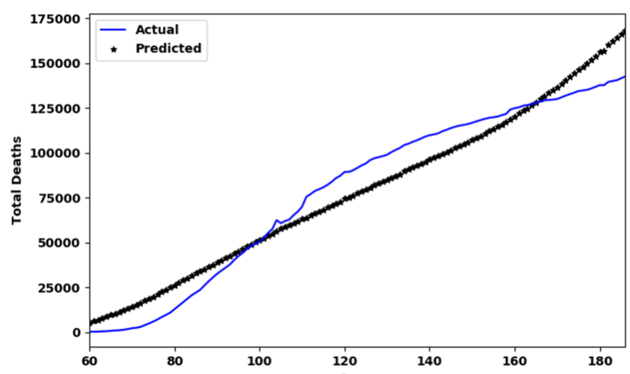
C)



D)



E)



F)

Fig. 32. The predicted number of deaths vs. actual reported values in G7 countries, A) France, B) Germany, C) Italy, D) Japan, E) United Kingdom and F) United States.

Conclusion

In conclusion, simplified mathematical models suggested in this report clearly confirmed the effectiveness of the controlling measurements like quarantine in reducing the intensity of Coronavirus outbreak. The reliability of simplified models decreases significantly to predict the number of deaths. This indicates the sensitivity and complexity of changes in number deaths with several known and unknown factors and requires further investigation.