2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)

Analysis and Prediction of COVID-19 Timeline and Infection Rates

2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) | 978-1-7281-1056-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/ASONAM49781.2020.9381338

Hisham Al-Mubaid
Dept. of Comptuer Science
Univ. of Houston , Clear Lake
Houston, TX, USA
hisham@uhcl.edu

Izzat Alsmadi
Dept. of Computing and Cyber Security
Texas A&M, San Antonio
San Antonio, TX, USA
Izzat.Alsmadi@tamusa.edu

***Abstract*—The number of new cases of infection of the Coronavirus disease, COVID-19, is alarming in many places in the world. In several world countries, including USA, the infection rates and daily cases numbers are fairly high; and there are even some spike increases in some USA states. Since the USA is experiencing the highest number of daily new cases in the world from May through July, the most important question is when we will witness an effective decline in the number of daily new cases? This paper follows a data-driven approach to induce the disease decline values from two country groups in the world where the disease declined already to less than 25% of its peak daily new cases. We apply these country groups' models to predict the decline to 25% of the US's peak. We compiled, examined, and analyzed pandemic data and statistics of two countries: g1: 42 countries, and g2: 14 countries. We utilize their data in the prediction of the decline timeline of the US. Group g2 consists of 14 countries having a similar number of cases per one million population. The majority of the models predict that the decline to 25% of the US's peak will be around the end of November to the first week of October. The results are significant and impressive as it is highly demanded to have clues and methods for the timeline prediction of this pandemic in the USA.**

***Keywords*—COVID-19 pandemic, Coronavirus disease, timeline prediction.**

1. INTRODUCTION

A novel human coronavirus, more specifically severe acute respiratory syndrome coronavirus 2, (SARS-CoV-2), or COVID-19 was identified in China in late 2019. This disease is highly contagious, and due to many unknowns, such as the incubation period, this has introduced difficulties in surveillance, enforcing effective quarantine measures, and effective contact tracing. This has led to the spread of this novel virus to all parts of the globe, turning this event into a pandemic.

Countries provide their infection rates in the form of confirmed cases, recovered cases, and deaths. Understanding this data, as well as forecasting future spread and mitigation, is dependent on the accuracy of this collected data. This paper will make use of historical data as well as attempt to forecast future spread.

As a worldwide pandemic, COVID-19 impacted the various countries in different timelines. For example, China, South Korea, and Italy can be studied as early countries in COVID-19 historical time series, while the United States, Brazil, and Canada can be studied as late countries.

IEEE/ACM ASONAM 2020, December 7-10, 2020

The assumption here is that countries like China and South Korea, who have taken extreme testing and quarantine measures, will continue to do so, and countries that have not been as strict for whatever reason will continue to do so. External factors such as population density of each country or city, testing available to the public, federal intervention on mitigation, and a plethora of other factors will not be considered in this paper. Individual studies on any of these assumptions may be explored at a later date. In the first week of June 2020, the coronavirus disease (COVID-19) yielded more than seven million cases and over 400,000 fatalities worldwide, with more than 2 million cases and over 110,000 deaths in the USA alone.

We will try to predict the rest of the USA's curve by data analysis compared with the top countries having at least 20,000 cases of the disease by the first week of June (42 countries; this is called group g1) as shown in Table 1. We also apply our analysis with the countries having a similar number of cases per 1M pop (Table 2) and call it group g2. We will utilize the curve of these countries to (extrapolate) and induce the curve of USA for the rest of this pandemic. We will use a number of factors to determine the curve fitting process: culture, economics, diversity, climate, and combination.

As of the first week of Oct. 2020, there are still increasing cases every day in the USA, and some spike increases in several US states. With the USA experiencing the highest number of daily new cases globally and still increasing , the most important question is when we will see the effective decline.

2. BACKGROUND AND RELATED WORK

The coronavirus disease started in China in Wuhan's city in December 2019 and is caused by the SARS-Cov-2 virus as named by the World Health Organization (WHO). The virus transferred to humans, most likely, from bat animals around November of 2019 in Wuhan, China. The virus and the disease spread and transferred into the entire world starting in January 2020. Italy, Iran, and Spain were among the first few countries to get hit hard by it. Around mid-March, COVID-19 became the most difficult and biggest problem afflicting and crippling the whole world and every aspect of our lives and continuing so for at least three months for almost the whole world. The disease was named COVID-19 by WHO, and the virus called the novel coronavirus (novel because it is a new member of the coronaviruses family). Initially, they named it n-Cov-19 or 2019-nCov, then changed the name to SARS-COV-2). Thus, now the virus is SARS-COV-2, and the disease is COVID-19 or the 'Novel Coronavirus Disease 2019'. From the family of human coronaviruses (H-CoV's), three highly pathogenic H-Cov's have been identified so far [10], including: (1) Middle East respiratory syndrome coronavirus (MERS-CoV); (2) severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV), and (3) the 2019 novel coronavirus (or SARS-Cov-2), and previously named 2019-nCoV by the World Health Organization (WHO) [10]. Among these three, the MERS-CoV was responsible for 2494 cases and 858 deaths in 27 countries during the 2012 MERS outbreak; whereas the SARS-CoV pathogen was responsible for >8000 cases and 774 deaths in 37 countries during the 2002 to 2003 SARS outbreak; and SARS-Cov-2 is causing close to half million deaths so far [3].

The SARS-Cov-2 virus is not lab manipulated and naturally produced as a result of many years of changes, as most significant studies showed, but it is not proved as the origin of this virus [1].

In [2], Emameh et al. (2020) used computational modeling and biodata mining to study and find potential domains for the simulation of the human immune system to consider it as a target of vaccine and drug studies [2]. Li et al. (2020), in [3], presented a very good study of the evolutionary history of the virus with the potential animals and species analysis in the context of this disease [3]. Tai et al. (2020) in [4] study and present the complete details of the virus and the disease in the context of its transfer to humans via the spike protein and the sequence of molecular of biological functions involved in the process [10]. In [5], Petropoulos et al. (2020) present objective forecasts for the confirmed coronavirus disease and provide a study and timeline of the disease's potential implications for planning and decision-making. Kristian et al. (2020) present a study of this virus's notable features and origin theories [1]. A comprehensive study and list of proteins in the Coronavirus SARS-Cov-2 are presented by Francis (2020) in [6]. Khailany et al. (2020) presented a study of the genetic variations and mutation comparison of the reported SARS-Cov-2 genetic data over various time frames and locations [7]. They analyzed 95 complete genomes of SARS-Cov-2 submitted to various databases through April 2020 [7].

3. Methods and Techniques

The main goal of this paper is to predict the timeline (and the projection) of disease infection (and spread) rate in the US in terms of daily new cases (DNC), which is specified in *Seven-Day-Moving-Average* 7DMA. We start with the disease spread curve in the number of DNC smoothed with 7DMA, Figure 1, and Figure 2. In these figures, the bars represent the (*raw*) number of daily new cases, whereas the curve represents the *7DMA* of the bars. Figure 2 represents the disease (COVID-19) spread daily counts (bars) and curve (*7DMA*) of *Italy* as it has been one of the first countries to complete its curve [8, 9]. In general, the disease (daily infection) curve is divided into three phases: *increasing* (or *climbing*) phase, *peak* phase, and the *declining* phase. We would like to study and analyze the curve and predict the decline of daily new cases (DNC) curve, which can be applied for countries that have not peaked like the United States. We utilize the data and curves of the countries that already reached and passed the maximum number of daily new cases, which we call the *peak*, and continued to decrease to less than 25% of its *peak*, like Italy, as shown in Figure 1 and Figure 2. We extract and analyze new metrics and stats from the data before versus after reaching the *peak*.

TABLE 1: Top countries having at least 20,000 total cases as of the first week of June: Group ***g1***.

#	Country	Total Cases	Total Deaths	Tot Cases/ 1M pop	Population
	World	7,188,962	408,230	922	
1	*USA*	2,026,112	113,050	6,123	*330,885,824*
2	*Brazil*	710,887	37,312	3,346	*212,467,494*
3	*Russia*	476,658	5,971	3,266	*145,930,871*
4	*Spain*	288,797	27,136	6,177	*46,753,739*
5	*UK*	287,399	40,597	4,235	*67,864,660*
7	*Italy*	235,278	33,964	3,891	*60,466,843*
8	*Peru*	199,696	5,571	6,062	*32,942,259*
9	*Germany*	186,205	8,783	2,223	*83,768,186*
10	*Iran*	173,832	8,351	2,071	*83,924,324*
11	*Turkey*	171,121	4,711	2,030	*84,282,117*
12	*France*	154,188	29,209	2,362	*65,265,090*
13	*Chile*	138,846	2,264	7,267	*19,106,092*
14	*Mexico*	117,103	13,699	909	*128,847,928*
15	*Saudi Arabia*	105,283	746	3,027	*34,778,459*
16	*Pakistan*	103,671	2,067	470	*220,603,292*
17	*Canada*	96,244	7,835	2,551	*37,721,735*
18	*China*	83,040	4,634	58	1,439,323,776
20	*Bangladesh*	68,504	930	416	*164,587,111*
21	*Belgium*	59,348	9,606	5,122	*11,586,627*
22	*S. Africa*	50,879	1,080	859	*59,261,034*
23	*Belarus*	49,453	276	5,233	*9,449,500*
24	*Netherlands*	47,739	6,016	2,786	*17,132,661*
25	*Sweden*	45,133	4,694	4,471	*10,095,462*
26	*Ecuador*	43,378	3,642	2,461	*17,625,612*
27	*Colombia*	40,719	1,308	801	*50,848,887*
28	*UAE*	39,376	281	3,984	*9,882,820*
29	*Singapore*	38,296	25	6,549	*5,847,525*
30	*Egypt*	35,444	1,271	347	*102,204,915*
31	*Portugal*	34,885	1,485	3,421	*10,198,365*
32	*Kuwait*	32,510	269	7,620	*4,266,472*
33	*Indonesia*	32,033	1,883	117	*273,342,397*
34	*Switzerland*	30,972	1,923	3,580	*8,650,765*
35	*Ukraine*	27,462	797	628	*43,747,990*
36	*Poland*	27,160	1,166	718	*37,848,958*
37	*Ireland*	25,207	1,683	5,109	*4,934,313*
38	*Argentina*	23,620	693	523	*45,170,095*
39	*Philippines*	22,474	1,011	205	*109,487,582*
40	*Afghanistan*	20,917	369	538	*38,867,774*
41	*Romania*	20,604	1,339	1,071	*19,244,598*
42	*Dominican*	20,126	539	1,856	*10,841,126*

TABLE 2: Countries having similar total cases per 1M pop (highlighted column) with at least 20,000 total cases (group ***g2***).

#	Country	Total Cases	Total Deaths	Tot Case/ 1M pop	Population (Millions)
1	Kuwait	32,510	269	7,620	4.266
2	Chile	138,846	2,264	7,267	19.106
3	Singapore	38,296	25	6,549	5.847
4	Luxembourg	4,040	110	6,461	0.625
5	Spain	288,797	27,136	6,177	46.753
6	USA	2,026,112	113,050	6,123	330.885
7	Peru	199,696	5,571	6,062	32.942
8	Iceland	1,807	10	5,297	0.341
9	Belarus	49,453	276	5,233	9.449
10	Gibraltar	176		5,224	0.33
11	Belgium	59,348	9,606	5,122	11.586
12	Ireland	25,207	1,683	5,109	4.934

13	Armenia	13,325	211	4,497	2.962
14	Sweden	45,133	4,694	4,471	10.095

That is, the duration of time (number of days) from 25% of the peak to the peak value is a factor that will be utilized to induce the duration to reach (decline) to 25% of the peak after the peak (in the *declining* phase); see Figure 1.

<u>*Example e1:*</u> For example, if the *Peak DNC* (PDNC) which represents the *maximum* daily new cases in Italy, was reached on 3/26 (Figure 2 and Figure 3) with $7DMA$ as 5600, and 3/17 was the 50% of *PDNC* (i.e., *7DMA*=2800 in 3/17); also 25% of *PDNC* was on 3/11 (i.e., on 3/11 the *7DMA* was ~1400 where 1400=75% of *PDNC*).

3.1. Details of the Method

Let n_i be the *raw* number of new confirmed *COVID-19* cases on the day d_i , and let *7-Day-Moving-Average* (7DMA) on day d_i be the average of raw counts of d_i and the preceding six days $\{n_i, n_{i-1}, \dots, n_{i-6}\}$. That is,

$$7DMA \text{ on day } d_i = (\textstyle\sum_{j=i-6}^{i} n_j)/7 \quad ...(1)$$

In this paper, we use $7DMA$ on day d_i to refer to daily new cases, DNC, of day d_i. Define Peak daily new cases (*PDNC*) to be the greatest *7DMA* value that is followed by the *declining* phase; see Figure 1.

$$PDNC = \max_i (7DMA \text{ on } d_i) \quad ...(2)$$

For example, in Italy, the *PDNC* is 5600 (in terms of *7DMA* on March 17). Let *a* and *b* be the points when 25% of *PDNC* and 50% of *PDNC*, respectively, were reached before reaching the *PDNC*; that is, during the *increasing* stage; see Figure 1. For example, in Italy,

a: 1400, point of 25% of *PDNC*

b: 2800, the point of 50% of *PDNC*

PDNC: 5600, peak daily new cases (the day with max number of new confirmed cases in one day, declining stage after that) on 3/17.

The duration (in number of days) takes a country to reach from 25% of *PDNC* to *PDNC* is called d_1; similarly, we recognize d_2 as the number of days from 50% of *PDNC* (*i.e., point b*) to *PDNC* as follows:

d_1: number of days from the point *a* to *PDNC*

d_2: number of days from point *b* to *PDNC*

After reaching the *PDNC*, the *declining* stage starts where the *7DMA* of the number of new daily cases starts to go down. The point when *7DMA* decrease (down) to 50% of *PDNC* we call it b'; similarly the point when it drops down to 25% of *PDNC* (after passing *PDNC*) we call it a' as follows:

a': 1400, point of 25% of *PDNC* during the *declining* stage.

b': 2800, the point of 50% of *PDNC* during the *declining* stage.

With this, we define d_1' and d_2' (Figure 1) as:

d_1': number of days from *PDNC* to point a'

d_2': number of days from *PDNC* to point b'

We also define two terms x_1 and x_2 to draw the relation between d_1 and d_1' (d_2 and d_2' resp.) as follows:

$$x_1 = \frac{d_1'}{d_1} \quad ...(3)$$

$$x_2 = \frac{d_2'}{d_2} \quad ...(4)$$

Therefore, in example *e1* above, d_1= number of days from 3/11, which is *a* (*i.e.,* 25% of PDNC) to 3/26 (which is the *PDNC*); then d_1=15. On the other hand d_1'= number of days from 3/26 to 5/8 (where 5/8 is a' which is the point of decreasing to 25% of *PDNC* after *PDNC* in the *declining* phase); then d_1'=42. This means it took Italy 42 days to move from the *peak* (peak is *PDNC* on 3/25) down to the 25% of peak (on 5/8).

3/26	peak	7DMA(on 3/26)=5600
5/8	25% of peak	7DMA(on 5/8)=1400

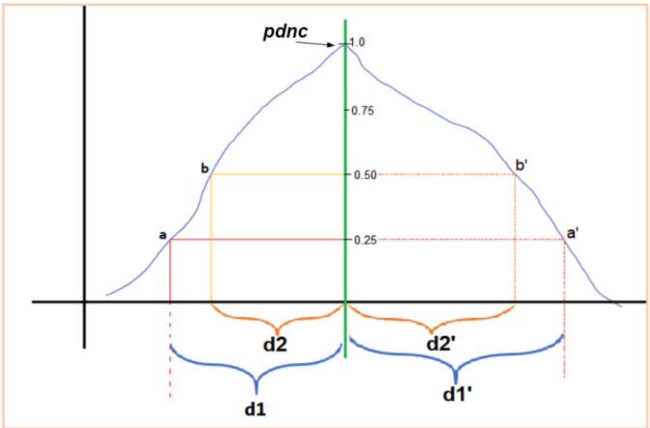

FIGURE 1: 7DMA diagram of daily new cases where d_1, d_1', d_2, and d_2' are illustrated.

TABLE 2b: The disease spread values/measures taken from the UK curve of daily new cases (*DNC*) using 7-day moving average (*7DMA*).

	UK	*Comments*
PDNC	4/14	PDNC is: 5520
a	3/26	1198 (1380) (25% of PDNC)
b	4/1	2850 (2760) (50% of PDNC)
a'	6/12	1377 (1380) (25% of PDNC)
b'	5/22	2789 (2760) (50% of PDNC)
d1	19	No. of days: a and PDNC
d2	13	No. of days: b and PDNC
d1'	59	No. of days: PDNC and a
d2'	38	No. of days: PDNC and b

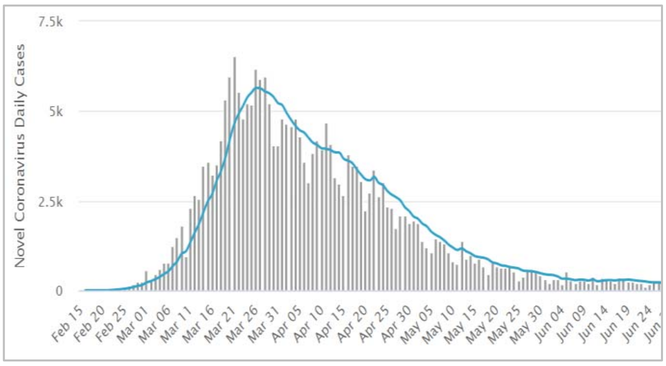

FIGURE 2: The daily new cases with a 7-day-moving-average (7DMA) curve for Italy.

So using equation (3) to calculate x_1 as $x_1 = \frac{42}{15} = 2.80$; similarly we calculate x_2 as in equation (4) above so that, for this (Italy) example, $x_2 = \frac{d_2'}{d_2} = \frac{29}{10} = 2.9$ Next, we utilize both x_1 and x_2 to induce the d_1' and d_2' for the United States, for example, in number of days. We induce d_1' from equation (3) and d_2' from equation (4) as follows:

$d_1' = x_1 * d_1$ ...(5)

$d_2' = x_2 * d_2$ ...(6)

Define two variation factors, f_1 and f_2 , that affect the induction of d_1' (and d_2') from x_i and d_i between two countries such that:

$1 \leq f_1 \leq 3$ and $1 \leq f_2 \leq 3$ ......(7)

These variation factors (equation 7) allow for leeway and flexibility in applying different countries' models. For example, if the USA reaches 75% of *PDNC* on 6/29 (*say 7DMA* = ~41000; *hypothetical*) and starts a decline, only to increase again, reaching up the 75% *PDNC* again on 7/8 then we can infer that the variation factor for the US is fairly large (for example, $2 \leq f_i \leq 3$); see Figure 4

3.2 Italy and United States:

Since Italy is at the end of, and almost completed, the curve, and one of the first countries to reach this stage, we can utilize Italy's daily new cases (DNC) statistics and curves to induce $x_1$and x_2 for the US and predict the decline of the infection in the US. For this example (case of Italy; see Figure 3), we have *a* and *b* are 3/11 and 3/16 respectively, b' and a' are 4/25 and 5/8 resp., and the *PDNC* is on 3/26. Moreover, the value of *PDNC* is 5600, and both a and a' occurs at ~1400 whereas b and b' both occur at ~2800 before and after the *PDNC*. If we examine both Figure 1 and Figure 2, for *Italy*, and Figure 4 (for the USA) where USA *PDNC* 54800 is on 7/9

a = 3/28 (14,000)

b = 6/20 (26,800)

PDNC =7/9 (at 54800)

then, d_1= 103

and d_2= 20

If we apply the model of Italy to the US with the default variation factors $f_1 = f_0 = 1.0$, then we get (for the US) the following predictions: $d_1' = 1.0$*2.8*16 = 44 and $d_2' =$ 1.0*2.9*12 = 35.

3.3 UK and United States:

The curve of daily cases and 7DMA of the UK is shown in Figure 5 and data in Table 2b. To predict the date of 25% PDNC of the US from the UK model; we need to calculate *x1* and *x2* for the UK. *x1* = 59/19 = 3.1 and *x2* = 38/13 = 2.9. Applying these to the US model we get d1'=1.0*2.9*16 = 46.4, which predicts the 25% decline in the US will be almost 46 days after reaching *PDNC* using the default value of f1 variation factor ($f_1 = 1.0$). If we adjust the variation factor of the UK to the US from 1.0 to 2.0, that is: $f_1 = 2.0$, then we get d1' =93, indicating the 25% decline will be after 93 days of the PDNC (4/10), which will be 7/13 (July 13). Similarly, the 50% PDNC predicted for US from the UK model (using *b*, *b'*, x_2, and *f2*) would be June 28 (or 6/28).

C. Group g1 (countries) and the United States:

Table 1 includes 42 countries that we collected in the first group (Group g1); then, we removed countries with less than 10M population (6 countries) remaining 36 countries in group g1. The values of these countries are shown in Table 3. From group g1, we calculate $x_1$and x_2 as x_1 = 55/19 = 2.89 and x_2=38/13 = 2.92. If we apply the model of this group of countries (36 countries) combined to the US we get: $d_1' =$ 1.0*2.89*16 = 46 and $d_2' =$ 1.0*2.89*12 = 35.

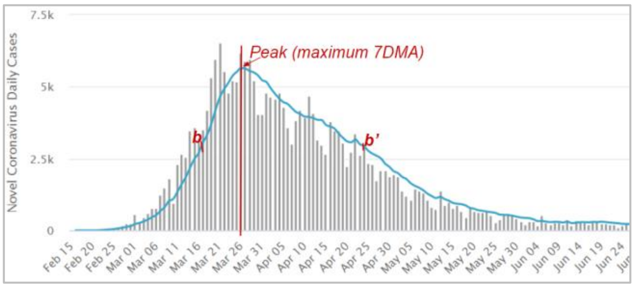

Figure 3: The daily new cases in *7DMA* for Italy. Here *b* and b' represent 50% of the *peak* before and after the *peak* resp. Notice that number of days from pint *b* to the *peak* is denoted as d_2.

3.4 Group-2 and United States: *(countries with similar total cases per 1M pop and >20K total cases)*

The countries in group g_2 are listed in Table 2. This group, g_2, includes 14 countries, which are all the countries that have similar total cases per one million population (1 M pop) with at least 20,000 total cases (Table 2). Ranging from 4400 to 7600 where USA has 6100 cases per one million (USA is the sixth in the list). Table 4 contains the calculated values of the timeline for this group. For this group, g_2, we get: x1 = 59/18 = 3.28 and x2=37/11 = 3.36. If we apply the model of this group of countries (36 countries) combined to the US we get: $d_1' =$ 1.0*3.28*16 = 52 and $d_2' =$ 1.0*3.36*12 = 40.

4. Analysis of Results and Discussion

In general, the recovery time (i.e. the right side of the curve), as in Figures 2 thru 7, takes more than the time taken before the peak (i.e. on the left side of the curve). Also, we noticed that d_1' is always greater than d_1 ($p<0.05$) that is: $x_1 > 0$.

If the point of 25% (or 75%) occurred more than once, we would consider the latest one. The total number of new cases every day, *DNC*, is the best indicator of disease spread and infection rate in any country. In this paper, we take the average of the preceding seven days (*7DMA*). For every day, the DNC value can be in a curve that can be divided into three parts: increasing, peak, and declining. In this section, we will attempt to put the results of section 3 in the context of one country, the USA, and will try to induce a timeline from these models. The timeline of any pandemic or disease outbreak includes several stages. In its simplest form, the curve includes three stages, increase, peak, and decline. The increase and decline stages can also be divided into three stages 25%, 50%, and 75% of peak. We draw the relationship between these climbing and falling parts of the curve in an effort to predict the falling points in a new curve. We need to be careful in determining the peak for any country because a country may reach some peak and then the *7DMA* starts to drop (entering a *declining* stage) then goes back to increase and exceeds the peak. We examined many cases like this, we found that the peak is known/determined when a country reaches the maximum and start dropping (*declining* phase) until the curve reaches down to *0.75* of the peak. For example, suppose a country *C* reaches a *max* value (say, *7DMA*=1000) on date *w* after which the *7DMA* starts to drop and continue to reach 75% (or *7DMA*=750) of that *max* then we can determine that is the peak and *w* is the point in time to predict (count for) d_1' for that country. For example, in Figure 6 and Figure 7, we can see that the peak is on April 10 with *7DMA* of 165 because after that, the *7DMA* dropped to reach 75% of 165 (on April 14). Notice the first maximum on March 30 with *7DMA* of 93 was not taken as the peak value (or PDNC) for this country because it did not decline to 75% of 93.

In this study, we utilized the disease spread curve of Italy, UK, and many other countries to induce the falling (and decline) stages for the US curve so as to predict the complete that the USA will reach 25% of its peak (maximum, which is 32400 cases per day *7DMA*) anywhere between July 25 to August 7, according to the models of the most countries considered in this study as shown in Table 5. Assuming these results are accurate to some extent, this will give insight into planning and decision making for most industries like travel, airlines, tourism, education, and other activities in our lives.

5. Conclusion

The most important issue affecting our lives, as of this writing, is the coronavirus disease *COVID-19*. In many countries in the world and the USA, particularly when we are going to see a real and effective decline, it is the most important or one of the most important questions that anyone can ask. This paper presented a heuristic data-driven model for predicting the decline timeline of disease spread in the US, benefiting from data of similar countries who already passed these stages. Since we are interested in predicting when a country is *almost over* and done with this pandemic, we focus mainly on inducing d_1' The country's point of reaching/drops down to 25% of peak (in the declining phase) means that the country is almost recovered and can be classified as *almost over* and *winning* with this pandemic. For example, according to most models, if the USA reached their peak on June 30, it will be over, decreasing to below 25% of the *peak*, between November 23 and December 1, as shown in Table 5.

References

[1] Kristian G. Andersen, Andrew Rambaut , W. Ian Lipkin, Edward C. Holmes, and Robert F. Garry. The proximal origin of SARS-CoV-2. Nature Medicine 26, 450–452 (2020).

[2] R. Z. Emameh, R. H. Nosrati, and R.A. Taheri, (2020). Combination of Biodata Mining and Computational Modelling in Identification and Characterization of ORF1ab Polyprotein of SARS-CoV-2 Isolated from Oronasopharynx of an Iranian Patient. Biological procedures online, 22, 8. 2020. https://doi.org/10.1186/s12575-020-00121-9

[3] Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B. T., and Chaillon, A. (2020). Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. Journal of medical virology, 10.1002/jmv.25731. Advance online publication. https://doi.org/10.1002/jmv.25731

[4] Wanbo Tai, Lei He, Xiujuan Zhang, Jing Pu, Denis Voronin, Shibo Jiang, Yusen Zhou, and Lanying Du. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine. Cellular & Molecular Immunology (2020).

[5] Petropoulos F, Makridakis S (2020) Forecasting the novel coronavirus COVID-19. PLoS ONE 15(3): e0231236. https://doi.org/10.1371/journal.pone.0231236

[6] Francis K. Yoshimoto. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. Protein J. 2020 May 23 : 1–19. doi: 10.1007/s10930-020-09901-4. PMID: 32447571.

[7] R.A. Khailany, M. Safdar, and M. Ozaslan, (2020). Genomic characterization of a novel SARS-CoV-2. Gene reports, 19, 100682, 2020. https://doi.org/10.1016/j.genrep.2020.100682

[8] JHU https://systems.jhu.edu/research/public-health/ncov-model-2/

[9] https://www.worldometers.info/coronavirus/

TABLE 3: The disease spread values/measures of the 42 countries of the first group (group g1), consisting of 36 countries.

	Group 1 (36 countries)	***Comments***
PDNC	4/17	PDNC is: 5520
a	3/29	1198 (1380) (25% of PDNC)
b	4/4	2850 (2760) (50% of PDNC)
a'	6/11	1377 (1380) (25% of PDNC)
b'	5/25	2789 (2760) (50% of PDNC)
d1	19	No. of days: a and PDNC
d2	13	No. of days: b and PDNC
d1'	55	No. of days: PDNC and a
d2'	38	No. of days: PDNC and b

TABLE 4: The disease spread values/measures of the countries of group g_2 (14 countries).

	Group g2 (*14 countries*)	***Comments***
PDNC	4/12	PDNC is: 5520
a	3/27	1198 (1380) (25% of PDNC)
b	4/2	2850 (2760) (50% of PDNC)
a'	6/15	1377 (1380) (25% of PDNC)
b'	5/21	2789 (2760) (50% of PDNC)
d1	18	No. of days: a and PDNC
d2	11	No. of days: b and PDNC
d1'	59	No. of days: PDNC and a
d2'	37	No. of days: PDNC and b

TABLE 5: This table shows when USA *7DMA* will decrease/decline to 50% and 25% of its *peak* if the *peak* (*PDNC*) is reached on June 30, according to disease spread models of other countries.

	During the declining stage (*after passing the peak*)	
	50% of the peak in USA (or *a*)	25% of peak in USA(or *a'*)
Italy	7/22	9/24
Spain	7/26	9/26
UK	7/24	9/28
Germany	7/26	9/23
Group *g1* countries	7/27	9/25
Group *g2* countries	7/29	10/1
****note: if the USA 7DMA (supposedly) reached its peak on June 30, then according to Italy's model the USA will drop down to 50% of the peak on July 22 and continue to drop down to 25% of the peak on November 24 in the declining*		

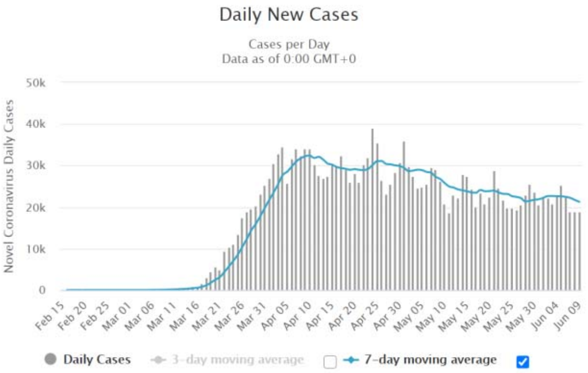

Figure 4**:** The *7DMA* curve for the USA.

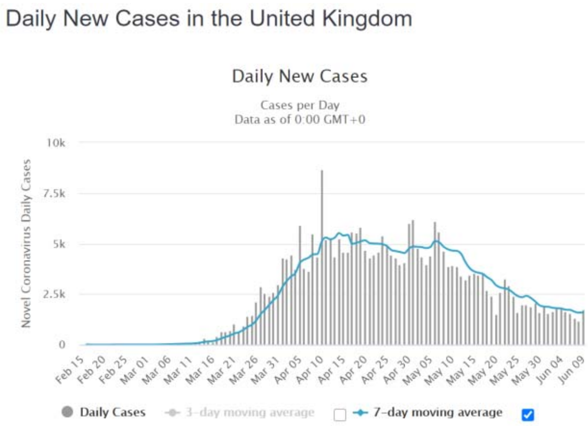

Figure 5**:** This curve illustrates the *7DMA* for the UK

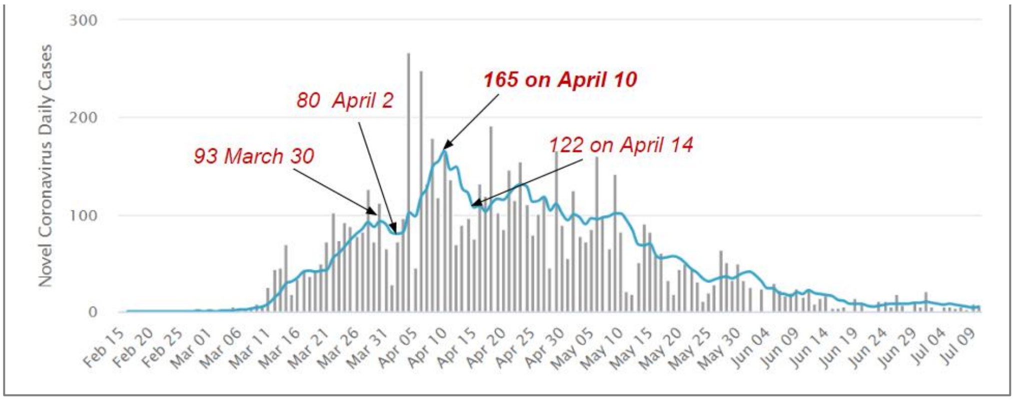

Figure 6: This curve shows the *7DMA* of Finland (from February 15 to July 9). In this curve, four values and dates are illustrated. In this illustration, the *7DMA* reached 90 on March 30 and started dropping to 80 but did not drop to 75% of 93 (which is 69.75) so 7DMA=93 on March 30 is not determined/considered the PDNC. On the other hand, the 7DMA=165 on April 10 then the decline to 122 (which is \75% of 165) on April 14 determines that the 7DMA=165 is the peak PDNC.

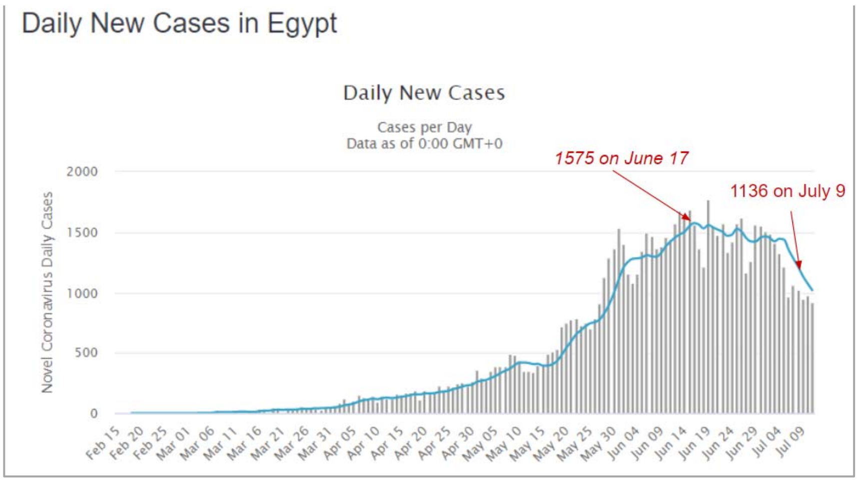

Figure 7: The 7DMA reached a *max* value of 1575 on June 17 and then started to fall till it went down to 1136 on July 9. Since 1136 is less than 75% of 1575 then we can consider this *max* value to be the *peak*.