

Do we really know the number of deaths caused by COVID-19 in the USA?

By: Sergei Ananyan, Ph.D.

sananyan@megaputer.com

CEO

Megaputer Intelligence

© 2020 Megaputer Intelligence Inc. All rights reserved.

Introduction: data sources and analysis methodology

To make informed decision on when and to what extent the US economy and social life can be restarted, decision makers need to have at their disposal the accurate numbers reflecting COVID confirmed cases, deaths and recoveries for different locations. The objective of this research is to determine whether there are any significant deviations of the real number of COVID deaths across US states compared to the reported numbers. In order to calculate these, we need to first check if there are any significant deviations in the numbers of non-COVID-related deaths numbers in different states (in the following we will be calling these Regular Deaths to differentiate them from COVID deaths).

For this purpose, we downloaded the data about deaths per month for individual US states during fourteen years (2005-2018) from the site of Centers for Disease Control and Prevention (CDC) (<https://wonder.cdc.gov/>). We used this data to calculate the mean mortality rate for each of the US states for different months and, assuming for simplicity that the distribution of mortality rate is normal, the corresponding dispersion during the fourteen-year time period, so that we can calculate the expected numbers of Regular Deaths (non-COVID) forecasted for the three full months when COVID was present in the USA: February through April 2020. In this calculation we used the numbers for the population of individual US states provided by US Census.

Then we downloaded from CDC the provisional data covering the deaths from COVID, as well as deaths from all causes, across all US states for three months of 2020: February through April (<https://www.cdc.gov/nchs/nvss/vsrr/covid19/index.htm>). This resource provided some data for the month of May too, but we deliberately excluded the data for the first weeks of May 2020 because the numbers for the recent weeks were incomplete. CDC is providing the deaths on a weekly basis. To convert this to monthly death numbers, we had to make an assumption that deaths are evenly distributed during the weeks falling on the border of two sequential months. The resulting death numbers for every month represented the actually reported deaths. We calculated the number of Actual Regular Deaths per month for each State by subtracting the number of COVID deaths from the number of deaths from all causes.

The reader should keep in mind that the deaths data for 2020 (both all deaths and COVID-related deaths) are still provisional as the CDC is making daily corrections to these data. However, these numbers are gradually settling with time. Our calculations are based on the latest batch of data provided by CDC (uploaded on May 22, 2020). We assume that it is relatively safe now to consider the

numbers for February – April 2020. As a cross-check we will be verifying the obtained results against data on COVID deaths taken from another reputable source.

First, we observed that some States take really long time to report their deaths data to CDC. We are specifically pointing out Connecticut and North Carolina whose deaths data for February – April 2020 reported to CDC, even on all deaths, appears to be seriously incomplete. Correspondingly, we excluded these States from our current analysis, planning to return to the analysis of data from these states later when they complete the reporting of their deaths data.

Now we were ready to compare the Actual Regular Death numbers per month reported by each state to the Expected Regular Death numbers forecasted based on the data covering 2005 through 2018. We calculated the ratio *Actual Regular Deaths / Expected Regular Deaths*: the resulting number larger than 1.0 implies that the Actual Regular Deaths are higher than Expected. In order to isolate only the most prominent deviations from Expected, we took into account only those States for which the number of Actual Regular deaths lied outside of $3.0 \cdot \text{dispersion}(\text{Expected Regular Deaths})$ interval around the mean value of the number of Expected Regular deaths for the corresponding month of the year over the past fourteen years. The reason for considering data points deviating from the mean value by three dispersions is that there is close to zero chance that this could have happened as a statistical fluctuation. In fact, during those fourteen years that we used as the benchmark of historical Regular Deaths data, no single State ever had the deviation of that size or more from the mean value of Regular Deaths for the corresponding month.

Discussion of results

Displaying the results on a geo-map, we assign the red color to those states that demonstrate significant excess or Actual Regular Deaths against Expected Regular Deaths, and blue color to those States that on the contrary have significant shortage of Actual Regular Deaths. The more intense the saturation of the color, the stronger is the corresponding deviation.

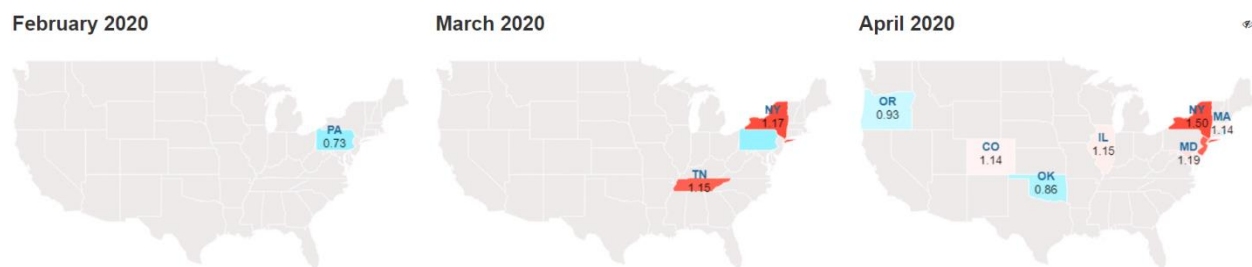


Fig. 1. States with significant (over $3.0 \cdot \text{dispersion}$) excess of Actual/Expected Regular Deaths are colored with progressive saturations of red, based on CDC data for COVID-related deaths. States with significant shortage of Actual/Expected Regular Deaths are colored with progressive saturation of blue.

These results are worth taking a closer look. First, as one could expect, during the month of February there was no single state in contiguous USA that had a significant upward deviation in the Actual Regular deaths over Expected Regular deaths for that month, which corroborates with our knowledge that there

were almost no COVID deaths during that month. Interestingly though, there was such an upward deviation in the number of regular deaths in the State of Hawaii. Could this be an indication that Hawaii was the first US state hit by COVID as early as February? This question requires further investigation, and we leave it outside of the current analysis.

Out of the three months we were analyzing, April was the month with the largest number of reported confirmed cases and deaths related to COVID. Correspondingly, let us focus first on the results for April.

April 2020

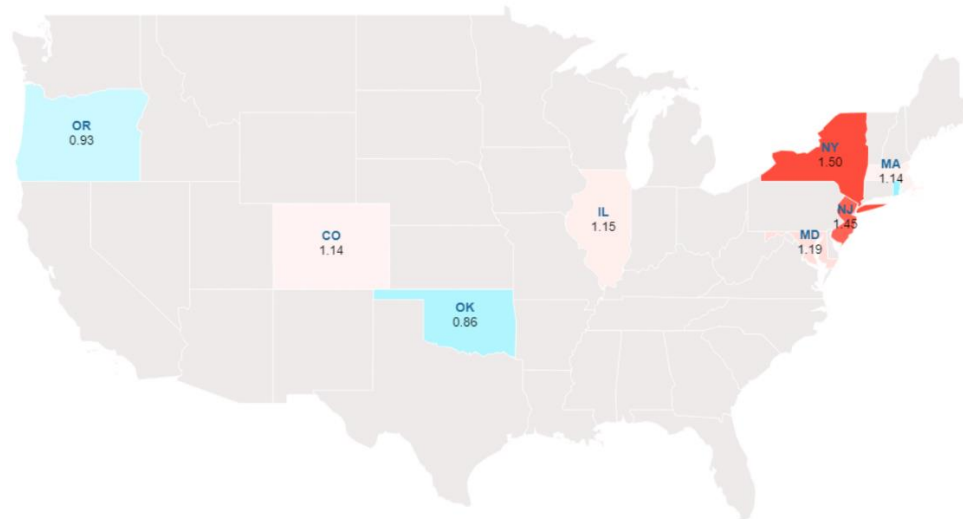


Fig. 2. States with significant deviations in excess (red) or shortage (blue) of Actual/Expected Regular Deaths during April 2020, based on CDC data for COVID-related deaths.

Since the deaths data for the last three months might be still incomplete, we can postpone the analysis of shortages in Actual Regular Deaths. These shortages can be due to the fact that not all deaths have been reported yet. We will be focusing on those states where we see a significant excess in the numbers of Actual Regular Deaths over Expected. The recorded death numbers only increase over time typically.

The two states that have the most significant excess in the number of regular deaths are New York (where the number of Actual Regular Deaths in April was 50% higher than expected: 18,805 actual non-COVID instead of 12,522 expected) and New Jersey (with 45% excess of Actual Regular Deaths over Expected: 8,768 actual non-COVID instead of 6,034 expected). Correspondingly, the numbers of excess in Actual Regular Deaths are: 6,418 – for New York, and 2,628 – for New Jersey. We should remember that these two States were the epicenter of COVID outbreak and of course there might be some unknown reasons for some excess in the number of regular deaths. However, it is going to be hard to explain the deviations in regular deaths of 50% and 45%, correspondingly. It appears much more plausible that a significant share of deaths due to COVID were not correctly identified as such and were reported as regular deaths. In order to obtain more accurate estimates for the upper limit of the number of deaths caused by COVID, we can assume that the excess in Regular Deaths is due entirely to missing COVID as the probable cause of the excess deaths. Then to calculate the total numbers of COVID deaths, we need to add to the already reported COVID deaths (19,375 - for New York and 7,559 – for

New Jersey in April), the numbers of excess Regular Deaths calculated above. This puts the Total numbers of COVID deaths in April to 25,658 in New York (a 32% increase) and 10,293 in New Jersey (a 36% increase).

As you can observe in Fig. 2, New York and New Jersey are not the only States with significant excess in Actual Regular deaths over Expected in April 2020. The states of Maryland, Illinois, Massachusetts, and Colorado are closely following them with the excess in Actual Regular deaths ranging 14-19%. The corresponding numbers for excessive regular deaths for all these States are provided in Table 1 below.

Regular deaths - CDC

	2-March			3-April		
	Actual Regular deaths	EXCESS Regular deaths	Expected Regular deaths	Actual Regular deaths	EXCESS Regular deaths	Expected Regular deaths
Colorado				3,617	449	3,168
Illinois				9,730	1,287	8,443
Maryland				4,628	723	3,905
Massachusetts				5,601	703	4,898
New Jersey				8,768	2,734	6,034
New York	15,738	2,293	13,445	18,805	6,283	12,522
Tennessee	6,828	891	5,937			

Table 1. Comparison of the numbers of Actual vs. Expected non-COVID deaths for US states that have significant upward deviations during months of March and April. The calculations are based on CDC data for COVID-related deaths.

While in March 2020 the pandemic was claiming only the first officially recognized death toll in the USA, for that month too, as we can see in Fig. 3, there was a more than three-dispersions excess in Actual Regular deaths over Expected for two States: New York (17% excess) and Tennessee (15% excess). The excess in regular deaths in March in the State of New York equals to 2,293 people.

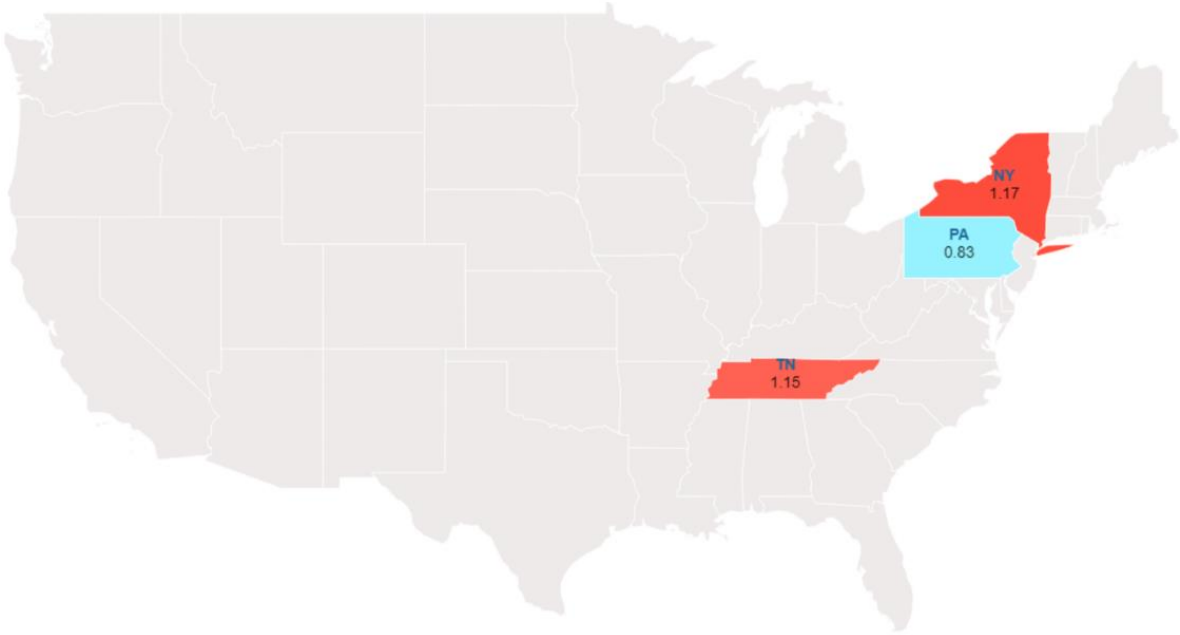


Fig. 3. States with significant deviations in excess (red) or shortage (blue) of Actual/Expected Regular Deaths during March 2020, based on CDC data for COVID-related deaths.

Recalling our assumption that the significant excess in Actual Regular deaths over Expected is most probably due to unrecognized COVID cases, we calculate the corresponding Total COVID-related deaths for those states that demonstrate significant upward spike in regular deaths (see Table 2).

COVID deaths - CDC

	2-March			3-April		
	COVID deaths	EXCESS Regular deaths	Total COVID deaths	COVID deaths	EXCESS Regular deaths	Total COVID deaths
Colorado				795	449	1,244
Illinois				2,155	1,287	3,442
Maryland				1,162	723	1,885
Massachusetts				3,594	703	4,297
New Jersey				7,559	2,734	10,293
New York	3,061	2,293	5,354	19,375	6,283	25,658
Tennessee	22	891	913			

Table 2. Reported COVID deaths numbers for US states that have significant upward deviations in Actual over Expected Regular deaths during months of March and April. The calculations are based on CDC data for COVID-related deaths.

New York is the only state that had significant upward deviations in the reported regular deaths during both months. It has the highest total number of excessive regular deaths (or under-reported COVID deaths) too: 8,576 when added over two months, which amounts to 38% addition to the reported

COVID deaths of 22,436 recorded by CDC during March and April. While New York had the highest absolute number of excessive regular deaths, for New Jersey excessive regular deaths equal 2,734 (or 36% of the reported COVID deaths), for Illinois – 1,287 (or 60% of COVID deaths), for Maryland – 723 (or 62% of COVID deaths), for Massachusetts - 703 (or 20% of COVID deaths), for Colorado – 449 (or 56% of COVID deaths).

Altogether, based on the numbers provided by CDC, we assessed that the total number of under-reported COVID deaths for the USA was 15,531, with the pair of States of New York and New Jersey responsible for over two thirds of these: 11,310 under-reported COVID deaths jointly. Based on COVID deaths data from CDC, New York and New Jersey combined reported 29,995 COVID deaths. Correspondingly, the under-reported COVID deaths for these two states combined represent 38% of the reported COVID deaths based on CDC data.

Cross-checking the results

As with any research, it would be most useful to try cross-validating the obtained results by using alternative data obtained from some other sources recording COVID-related deaths across the USA.

The only known credible source of data on US deaths from all causes during February – April 2020 is CDC. At the same time, the data for COVID-related deaths are simultaneously tracked by several other reputable institutions. To make sure our results and conclusions are trustworthy, we are going to cross-validate the results we obtained based on CDC data alone results against the COVID-related deaths data from John Hopkins University (JHU). It appears that JHU collects data on COVID deaths much faster, while CDC takes more time to perform more thorough investigation of individual cases before including them in their database. It will be interesting to see how much our results will change when we use COVID deaths data from these two different sources. Fig. 4 below compares the number of COVID deaths per week, as calculated by CDC and JHU. The investigated time period of February through April corresponds to weeks 5 through 18 on this chart. One can see that JHU reported more COVID deaths for weeks 16-18.

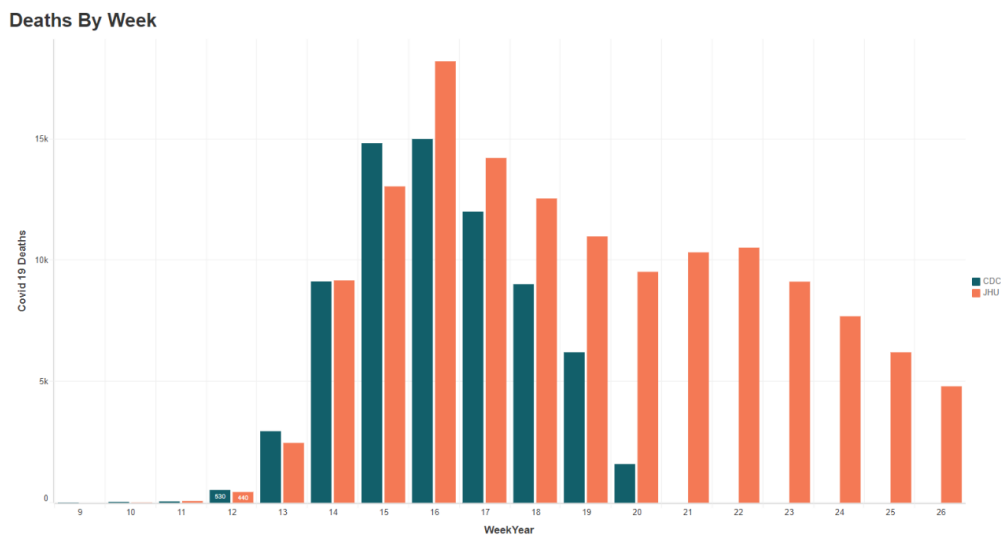


Fig. 4. Comparison of weekly COVID deaths numbers recorded by CDC (teal color) and JHU (corral color).

Repeating on the JHU data for COVID deaths the same calculations we did on the data recorded solely by CDC, we obtain the results depicted in Fig. 5.

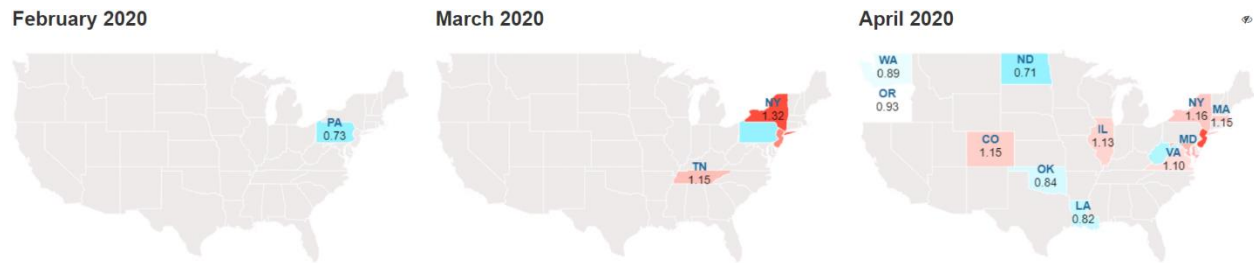


Fig. 5. States with significant deviations in excess (red) or shortage (blue) of Actual/Expected Regular Deaths during April 2020, based on JHU data for COVID-related deaths.

First, we observe that by and large the new result captures the same states that we saw in the calculations based on CDC data alone. We see a few more states that have significant shortages of the number of Actual Regular deaths against what was expected based on historical data (over three dispersions lower). While potentially this might be a signal of over-reporting COVID deaths in these states, we are going to postpone the discussion of these situations till the time when CDC finalizes its data reflecting deaths from all causes for the considered time period. Focusing on the states with the significant excess in regular deaths, we note that we gained only one more state with such excess in April: Virginia, which had a 10% excess in reported regular deaths in April. Also, we see that the absolute numbers of the excess in regular deaths had expectedly changed for all states. Let consider the new results in more detail (Figs. 6-7).

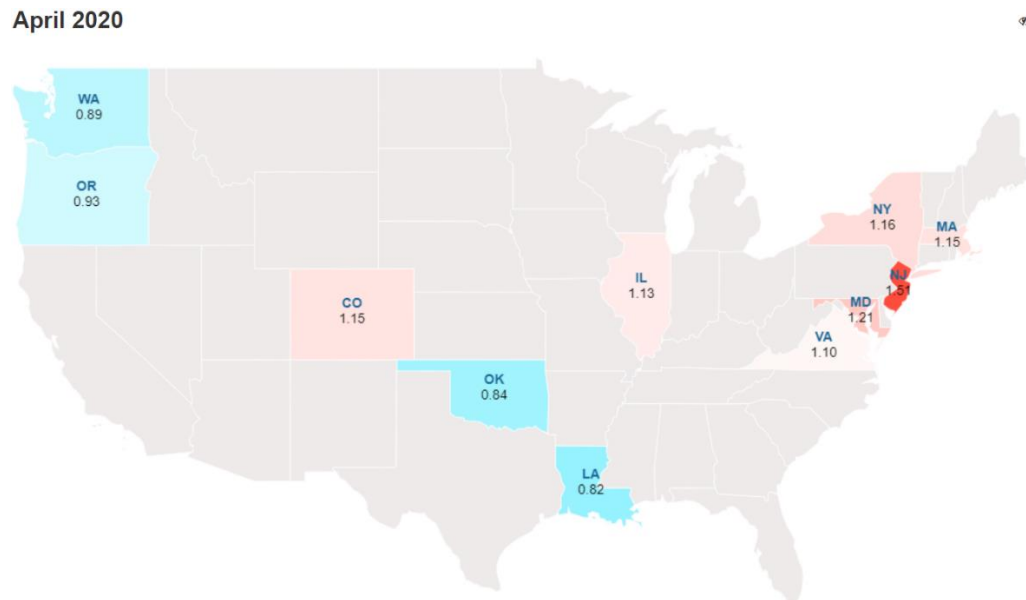


Fig. 6. States with significant deviations in excess (red) or shortage (blue) of Actual/Expected Regular Deaths during April 2020, based on JHU data for COVID-related deaths.

It is interesting that outside of gaining Virginia, the excess in regular deaths changed by more than 2% of Expected Deaths in only two states: New York (decreased from 50% to only 16%) and New Jersey (increased from 45% to 51%).

March 2020

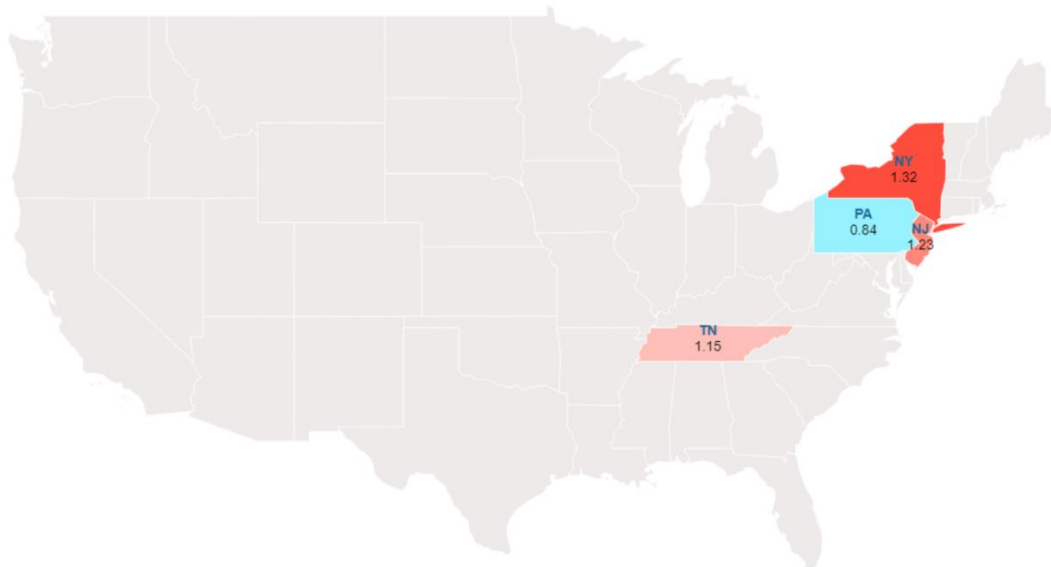


Fig. 7. States with significant deviations in excess (red) or shortage (blue) of Actual/Expected Regular Deaths during March 2020, based on JHU data for COVID-related deaths.

Excessive regular deaths for March 2020 changed significantly only for the same two states again. Using JHU data we see the 32% excess for New York (instead of 17% we calculated based on CDC data). In addition, based on JHU data on COVID deaths we see a significant spike in regular deaths in New Jersey in March (23% excess), which we did not see when relying on CDC data alone.

The results for all states with significant excesses in regular deaths for March-April based on JHU data are summarized in Table 3 below.

Regular deaths - JHU

	2-March			3-April		
	Actual Regular deaths	EXCESS Regular deaths	Expected Regular deaths	Actual Regular deaths	EXCESS Regular deaths	Expected Regular deaths
Colorado				3,635	467	3,168
Illinois				9,531	1,088	8,443
Maryland				4,733	828	3,905
Massachusetts				5,638	740	4,898
New Jersey	8,014	1,512	6,502	9,099	3,065	6,034
New York	17,722	4,277	13,445	14,588	2,066	12,522
Tennessee	6,838	901	5,937			
Virginia				5,991	565	5,426

Table 3. Comparison of the numbers of Actual vs. Expected non-COVID deaths for US states that have significant upward deviations during months of March and April. The calculations are based on CDC data for COVID-related deaths.

Correspondingly, we can calculate the total number of under-reported COVID deaths for the same states based on JHU data.

COVID deaths - JHU

	2-March			3-April		
	COVID deaths	EXCESS Regular deaths	Total COVID deaths	COVID deaths	EXCESS Regular deaths	Total COVID deaths
Colorado				777	467	1,244
Illinois				2,354	1,088	3,442
Maryland				1,058	828	1,886
Massachusetts				3,557	740	4,297
New Jersey	109	1,512	1,621	7,228	3,065	10,293
New York	1,078	4,277	5,355	23,592	2,066	25,658
Tennessee	12	901	913			
Virginia				554	565	1,119

Table 4. Reported COVID deaths numbers for US states that have significant upward deviations in Actual over Expected Regular deaths during months of March and April. The calculations are based on CDC data for COVID-related deaths.

To summarize, using COVID deaths reported by JHU we see two states that had significant upward deviations in reported regular deaths during both months: New Jersey joined New York. New York still demonstrates the highest total number of excessive regular deaths (or under-reported COVID deaths): 6,343 when added over two months, which amounts to 26% addition to the reported COVID deaths of 24,670 recorded by JHU during March and April. For New Jersey excessive regular deaths equal 4,557 (or 62% of the reported COVID deaths of 7,337), for Illinois – 1,088 (or 46% of reported COVID deaths of 2,354), for Maryland – 828 (or 78% of reported COVID deaths of 1,058), for Massachusetts - 740 (or 21% of reported COVID deaths of 3,557), for Colorado – 467 (or 60% of reported COVID deaths of 777), for Virginia – 565 (or 102% of reported COVID deaths of 554).

It is interesting that based on the COVID deaths numbers provided by JHU, we see almost the same total number of under-reported COVID deaths of 15,667 (compare with 15,531 based on CDC data alone – less than 0.9% difference). And again, the pair of states of New York and New Jersey is responsible for over two thirds of these: 10,920 under-reported COVID deaths jointly (compare with 11,310 based on CDC data alone – less than 3.5% difference). Based on COVID deaths data from JHU, New York and New Jersey combined reported 32,007 COVID deaths. Correspondingly, the under-reported COVID deaths for these two states combined represent 34% of the reported COVID deaths based on JHU data.

Fig. 8 below summarizes the complete set of results illustrating the under-reported COVID deaths numbers for the USA as a whole, as well as split by individual states that have the highest deviations of the actually reported Regular deaths against the expected Regular deaths.

Under-reported COVID deaths in the USA: **15,531** (based on CDC data)

Under-reported COVID deaths by State (based on CDC data)



Under-reported COVID deaths in the USA: **15,677** (based on JHU data)

Under-reported COVID deaths by State (based on JHU data)



Fig. 8. The total number of under-reported COVID deaths for the USA and the distribution of under-reported COVID cases by states. The results based on COVID deaths data from CDC (top) and JHU (bottom), correspondingly.

Conclusion

Our calculations based on two alternative sources of COVID deaths data (CDC and JHU) revealed that during the months of February - April 2020 throughout the USA there were over 15,500 unreported COVID deaths.

Over two thirds of all COVID death under-reporting instances are coming from the pair of states of New York and New Jersey: about 11,000 COVID deaths appear to be misreported there as non-COVID related deaths. This comprise 38% of all COVID deaths reported by these two states during March and April based on CDC data (or 34%, based on COVID deaths data recorded by JHU).

The significant number of under-reported COVID-19 deaths in New York and New Jersey calls for more scrupulous investigation of the causes of all deaths that occurred in these states during March and April 2020. This investigation should include the analysis of medical records of those people who had symptoms indicative of COVID prior to their deaths. Given the large number of medical charts that has to be investigated, it would be useful to employ specialized text analytics tools automating the process of information extraction from electronic medical records.

As time goes by and different institutions are finalizing their data on the number of deaths and the causes of these deaths for the time period covering COVID-19 pandemic, we will be updating our calculations aimed at obtaining the correct number of deaths caused by COVID. Megaputer Intelligence Inc. is providing the most up-to-date results for corrected COVID-19 death numbers online in the form of an interactive graphical report available at the following address:

<https://USCovidDeaths.megaputer.com>