

When will the Covid-19 Pandemic End? Levitt's Metric on Indian Data

Bhaskaran Raman (✉ br@cse.iitb.ac.in)

Indian Institute of Technology Bombay

Short Report

Keywords: Covid-19, Pandemic End, Linear regression

Posted Date: August 5th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-53018/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

When will the Covid-19 pandemic end? This paper seeks to answer this important question using the metric proposed by Michael Levitt in Mar 2020. We plot the metric for data corresponding to various Indian cities and states, and compare it with the plots for places like Italy and Spain, where the pandemic has “ended”. The projection is that Covid-19 deaths will become “insignificant”, by early Aug in Mumbai and Delhi. Similar projections can be made and are likely to be useful for other regions.

Introduction

The Covid-19 pandemic has brought much of the world to a stand-still. The question foremost on everyone’s minds is “when will the pandemic end?”. Michael Levitt, in his data analysis in Mar 2020 [1] of data from Hubei (China), proposed a simple metric, which shows a linear decreasing trend with time. In this paper, we apply this metric to data from various regions of India, to project as to when the pandemic is likely to “end”.

When can we say that a pandemic has ended? In this paper, we say the disease has “ended” as a pandemic when the number of daily deaths falls to “insignificant” levels (not necessarily zero), such as around 1/20th-1/30th of the daily non-Covid death rate.

Levitt’s Metric: In the data presented in [1], Levitt defines $H(t) = \frac{X(t)}{X(t-1)}$, where $X(t)$ is the cumulative count until day t . The count could be of deaths or cases.

Advantages: Aside from the simplicity of the metric, it has various other advantages. Since it is a ratio, it is independent of population size, and hence it can be used to compare across regions of any size: city, state, district, or country. It is robust to different countries/regions having different testing capacities, or varying definitions of what is a Covid death. It is also robust to systematic under-counting, so long as such under-counting is time invariant.

Deaths vs cases: In this paper, we compute the $H(t)$ metric on Covid-19 deaths rather than cases. Number of cases as a metric is not good for a number of reasons. First, it depends on testing capacity which varies temporally. It also depends on people’s access to tests, which in turns depends on many policies: are private tests allowed, are tests without prescription allowed, do people have transportation to go to test centre, etc. which are all time variant. To add to this, number of tests also depends on social acceptance of tests, which is also highly variable. Also finally, what we most care about is the number of deaths, rather than number of cases.

Linear fall of $H(t)$, $\ln(H(t))$: The original plot of $H(t)$ vs t for Hubei can be found in [1]. It was observed that $H(t)$ falls linearly with t , after about day-55 of the first Covid death. Later, in [2], it was observed that linear fall of $\ln(H(t))$ is a better fit. Note that since $\ln(1 + E) \approx E$ for small E , to a first degree of approximation, either $H(t)$ or $\ln(H(t))$ would do: we observed this uniformly in all our data plots.

Methods

Data Analysis

Plots for European Countries

Before plotting the data for India, we seek to understand a few essential characteristics of $H(t)$ vs t for other regions where Covid-19 has progressed further than in India. We first show the plots for UK and Sweden, where Covid-19 is showing signs of slowing but is still a significant concern: Fig. 1(a) shows the relevant plots.

We see that in both UK and Sweden, $H(t)$ has fallen significantly over time. Like in the Hubei plot in [1], there is a good linear fit to the data points, between day-55 and when $H(t)$ has reached a value of about 1.01 (near day 100). The figure shows both linear fits: both have high R^2 values.

A common concern in the context of $H(t)$ is that such a metric would approach 1 with increasing t , even if the number of deaths were to stay constant over time, i.e. $H(t)$ will simply be $(t+1)/t$ which tends to 1 for large t . In response to this concern, we show a plot of $(t+1)/t$ for comparison, in Fig. 1(a). We see that throughout, the slope of $H(t)$ is significantly steeper than the $(t+1)/t$ plot. This is consistent with the empirical observation that the number of Covid-19 deaths does reduce over time.

Finally, we observe another aspect: we also plotted $\ln(H(t))$ vs t . However, given how close $H(t)$ is to 1, $\ln(H(t)) \approx H(t) - 1$. And hence almost the same linear fit carries over for $\ln(H(t))$ vs t , albeit with an offset of 1. For clarity, we have avoided showing $\ln(H(t))$ altogether in this figure and also in subsequent plots below.

We next plot $H(t)$ for Italy and Spain: two countries where the disease is well past its peak. Fig. 1(b) shows these plots. In both cases, $H(t)$ of 1.01 is reached around day 75; the linear fit shown is for data points between day 45 and day 75. Here too, we see a linear trend, with high R^2 . Looking across the four countries in Fig. 1, it seems likely that we can predict the $H(t)$ in other regions too, based on the observed historical trend.

Plots for India, its States

Fig. 2(a) shows the $H(t)$ plot for India as well as the linear fit for the same. We note that the R^2 value here is low. This could be because of non-uniform policies across various states after lockdown-4. However, despite the low R^2 value, we note that the linear fit has good predictive power: Fig. 2(a) also shows the linear fit, had we done the analysis 15-days back (i.e without the data for the most recent 15 days). We see that the linear fit is almost the same in both cases: with/without the most recent 15-days data.

Fig. 2(a) also shows the $H(t)$ plot and linear fit for Brazil's data. Here we see that the slope is noticeably steeper for Brazil; this merits a causal analysis, but is beyond the scope of this paper.

Next we plot $H(t)$ for two states: Maharashtra and Gujarat, in Fig. 2(b). We note that in the case of Maharashtra, like in the case of India, the R^2 value is significantly low. The reason behind this is likely the fact that the majority of deaths in Maharashtra so far is from Mumbai. Even still, it is pleasantly surprising to see that the linear fit has good predictive power anyway: compare the linear fits with/without the last 15 days of data and observe that they are almost the same. In the case of Gujarat, the linear fit has already reached 1, suggesting that the pandemic has “ended”. This compares well with ground reality, with the entire state reporting only 10-20 deaths per day since more than one week. For comparison, the expected number of deaths for Gujarat’s population (6.3 crores) during “normal” times, as per 2019 death rate (7.3 per 1000 per year) is about 1260. So Covid-19 has already fallen to about just 1% of the overall cause of deaths in Gujarat.

Plots for hotspot cities in India

We next look at the three hotspot cities of Mumbai, Chennai, and Delhi. Fig. 3 shows the plots for each of the cities. In each case, we show $H(t)$, the linear fit with/without the last 15 days of data. In the case of Mumbai, we see that the linear fit before 15 days predicted an earlier end, than what is predicted today. This is because the rate of slowing has lessened. This is not uncommon, and can be observed even in the cases of Italy and Spain, on closer look: the slope of $H(t)$ becomes less steep after reaching about 1.01. So the end is “in sight”, even if a bit further than earlier predicted. In the case of Chennai too, such a difference between with/without last 15 days’ data is seen. And here too, $H(t)$ has fallen to below 1.01, and a gradual lessening of the $H(t)$ slope is to be expected. In the case of Delhi, the linear fit 15-days back is quite close to the linear fit now.

In all three hotspot cities, the $H(t)$ plot suggests that the “end” of the pandemic is near. This matches with ground truth reports too [3, 4].

Discussion

Implications

The implications of the above analysis are many.

1. The analysis is very simple: it does not need sophisticated data scientists, only 11th/12th standard mathematic background and spreadsheet soft- ware. It is thus easily accessible to policy making
2. The linear extrapolation based predictions appear fairly effective, at least until $H(t)$ hits low values like 01.
3. The linear trend is seen irrespective of size of region (city, state, or country), as well as irrespective of
4. The slope itself could depend on the policy (e.g. strictness of lockdown), but the linear decline of $H(t)$ toward 1 in all regions shows clearly what every epidemiologist knows: that the pandemic “ends” through herd

5. Even in places where $H(t)$ suggests that the pandemic is over, we can use the corresponding $H(t)$ plots from regions where the pandemic has long ended (e.g. Italy, Spain), to convince ourselves that likely a similar pattern will occur in other regions where the pandemic has not ended. This can replace baseless fears of repeat outbreaks, with guarded
6. Extending the linear fit for India as a whole, the prediction is that the pandemic will end in about 2 months from now, or toward the end of
7. Extending the linear fit for Maharashtra, the prediction is that the pandemic will end in about 5 weeks from now, or toward the first week of Sep.
8. The linear fit for the hotspot cities as well as Gujarat suggests that the pandemic is almost ended, or about a week away (end of Jul). This matches well with the ground

Conclusion

In this paper, we applied Levitt's $H(t)$ metric on India data, and showed that its linear fit has good predictive power. This has a lot of implications for data-based prediction and corresponding policy making, both in cities/states where the pandemic has almost ended, and also in cities, states, districts, and countries where it is yet to run its course.

Declarations

Acknowledgements I would like to thank the data sources <https://covid19india.org>, <https://worldometer.info>, and Google search engine for misc information. I thank Ramachandran Shankar (IMSc) for sharing Chennai's Twitter data. This analysis would not have been possible without feedback from Mahesh Tirumkudulu (ChE, IITB) and Shashikant Suryanarayanan (ME, IITB), as well as other friends and colleagues at IITB.

Conflict of interest

The author declares that he has no conflict of interest.

References

1. M. Levitt. 14-mar-20 china and world analysis. https://www.dropbox.com/s/ovlsfr53horcmsn/1.The_Corona_Chronologies.%20Part%20I%20-%20China.%20Part%20II.%20Rest_of_World_Michael_Levitt%20Stanford_r.pdf?dl=0 (2020)
2. M. Levitt, A. Scaiewicz, F. Zonta. Predicting the trajectory of any covid19 epidemic from the best straight line. <https://www.medrxiv.org/content/10.1101/2020.06.26.20140814v2.full.pdf> (2020)
3. C. TV18. Mumbai is probably past the covid-19 pandemic, says shashank joshi of lilavati hospital. <https://www.cnbcTV18.com/videos/healthcare/mumbai-is-probably-past-the-covid-19-pandemic-says-shashank-joshi-of-lilavati-hospital-6384781.htm> (2020)
4. B.S. Perappadan. Covid-19 bed occupancy low

Figures

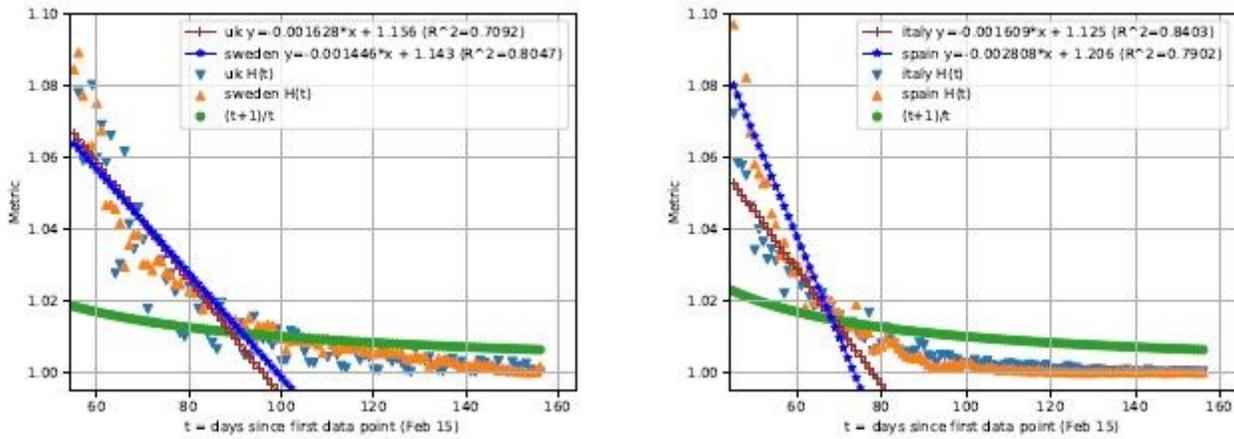


Figure 1

$H(t)$ plots: (a) UK and Sweden, (b) Italy and Spain

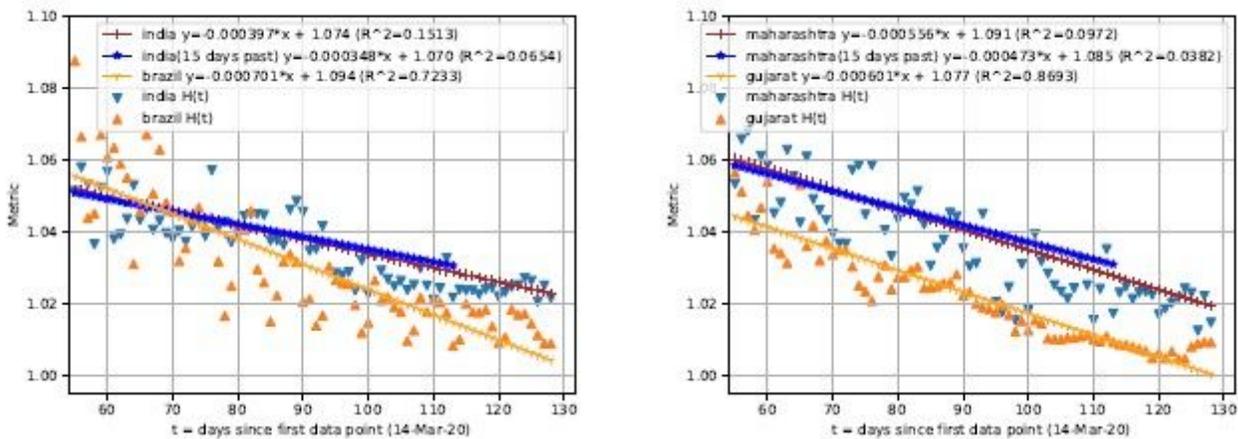


Figure 2

$H(t)$ plots: (a) India (and Brazil), (b) Maharashtra and Gujarat

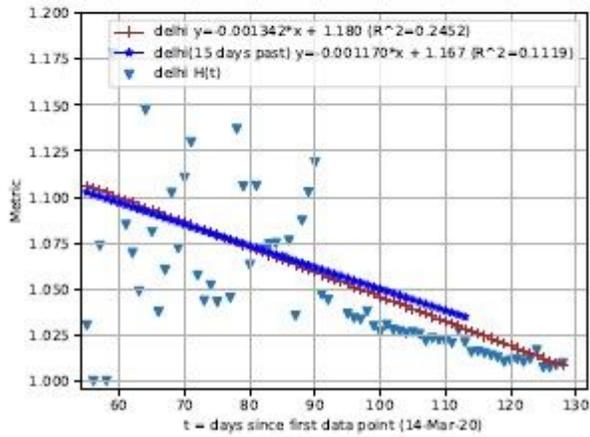
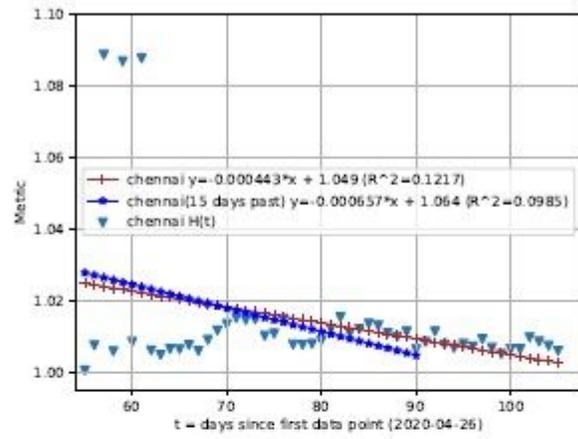
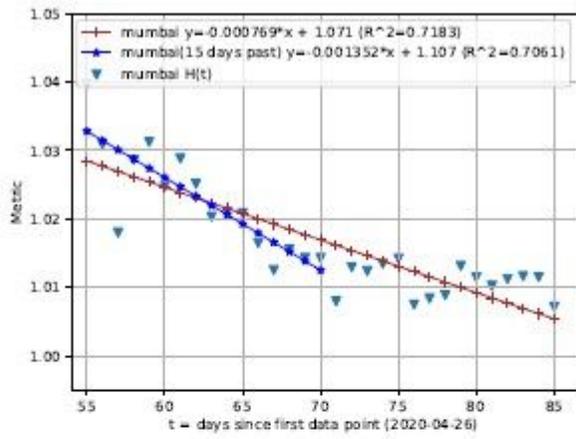


Figure 3

H(t) plots for hotspot cities: Mumbai, Chennai, Delhi