1 June 2020

Analysis of Covid 19 data for US and North Carolina

I have analyzed available covid 19 data through 31 May 2020. Sources for this analysis are: (1) data on number of new cases, cumulative number of cases, and new deaths, for US and NC- https://coronavirus.1point3acres.com/en ; (2) data on daily number of tests in US- https://ourworldindata.org/grapher/full-list-total-tests-for-covid-19?time=2020-03-14.. ; (3) data on daily number of new cases per test for NC- https://www.ncdhhs.gov/divisions/public-health/covid19/covid-19-nc-case-count .

I make some projections here about when the first wave of infections will have passed. These projections are not based on epidemiological models. Rather, they are based on statistical models. In particular, I fit non-linear regressions to the current data and make projections based on the best fit regression equation. It should be kept in mind that such projections assume that the processes that have been occurring up to now will continue on into the future. That assumption, of course, might not be true.

I. Analyses

In this document I report several statistics: number of new (reported) cases per day, number of new tests per day, number of new cases/test, and number of new deaths/day.

The number of new cases per day in the US leveled off around 1 Apr. Because the number of tests per day has increased steadily increased in this period (Fig. 2), one would expect to see an increase in the number of new cases per day if the true number of new cases were steady. Since we don't see that, it suggests that that one needs to correct the number of new cases by dividing by the number of tests.

This can be explained by a simple example. Suppose on day 1 there were 1,000 tests and 100 of them came back positive, while on day 2 there were 2,000 tests and 100 of them came back positive. Assuming that there could have been 2,000 tests done on day 1, we would expect to see 200 positive cases on that day (twice as many as observed because twice as many tests). In that case, instead of seeing the same number of new cases on the two days, we would have seen 200 on day 1 and 100 on day 2.

Another way of looking at this is to calculate the proportion of tests that are positive (which I will call pp for "proportion positive"), which is just the number of new cases/number of tests. For day 1, that would be 100/1,000 = 0.1, whereas for day 2 it would be 100/2,000 = 0.05. In other words, this measure says that there were twice as many new cases on day 1 than on day 2, which is the same result as above. One can thus the value pp to follow the *relative* number of new cases over time. It's ok to use relative numbers, because what we're interested in here is the time course of infections.

One caveat is that this analysis assumes the probability of detecting an infection with a test remains constant as the number of tests increases. One possibility to keep in mind is that early

on, tests were given primarily to those who were very sick or who were suspected to have come into contact with someone carrying the virus. It is possible that as tests become more available and anyone can get a test at any time, the probability that a test will yield a positive result could decline because a greater proportion of people being tested are less likely to have come into contact with the virus. My suspicion, though, is that the amount of testing that has been done so far has not reached the point at which this effect may become more than negligible.

The cumulative pp (Cum(pp)) for a given day is just the sum pp's up to that date. I report the fit of this data to a modified logistic growth curve:

$$Cum(pp) = a + K/(1 + e^{-r(day-b)})$$

Using the estimated parameters (a, K, r, and b), this curve is used to project when the cumulative number of reported infections levels off.

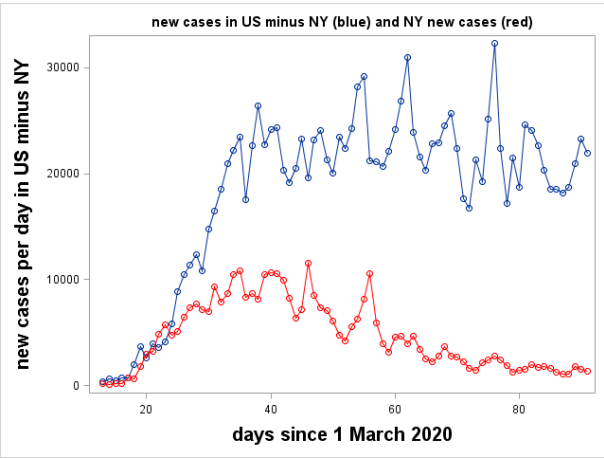
II. Number of new cases in US

In this section I break out statistics from NY from the rest of the US because the number of cases in NY is a substantial part of the total in the US, and it is of interest to know how infections are occurring outside of NY. Outside of NY, the number of new reported cases since day 50 (40 days ago) is showing a significant downward trend ($P < 0.0001$) (Fig. 1). Given that the number of new tests continues to increase (Fig. 2), this is a real sign that the actual number of new cases has been declining.

This conclusion must be tempered, however, by a trend in the number of new cases per test (Fig. 3). This had been declining steadily for about 30 days. However, for the past 20 days, this number has been statistically flat ($P = 0.3456$), suggesting that the actual number of new cases per day has stopped decreasing. Interestingly, the beginning of this period coincides with the beginning of "opening up" and lessening of social-distancing practices.

The data on cumulative new cases per test for the US outside of NY still fits a modified logistic growth curve (Fig. 4), and projections from this curve indicate a passing of most of the first wave of infections around day 95 (Fig. 5). Specifically, for the US outside of NY, on days 93, 121, and 159 (1 June, 29June, and 6 Aug, respectively—these dates are about 1 – 4 days later than projected last time), 95%, 99%, and 99.9% of the total infections during the first wave of the virus will have occurred. On these dates, the minimum number of infections yet to come are 84,166, 18,592, and 1,956. On days 112 and 146, the expected minimum number of new cases will be 1,128 and 27.

It is very evident that NY has done an excellent job in controlling the virus. The number of new cases and new cases per test have been declining steadily since around day 30 (30 March) and continues to decline. They are clearly doing something right.



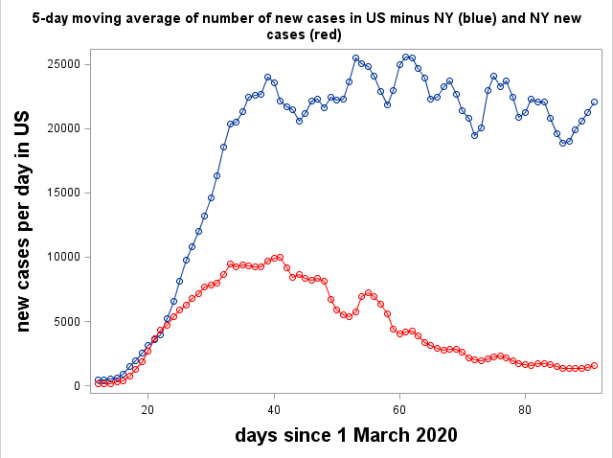
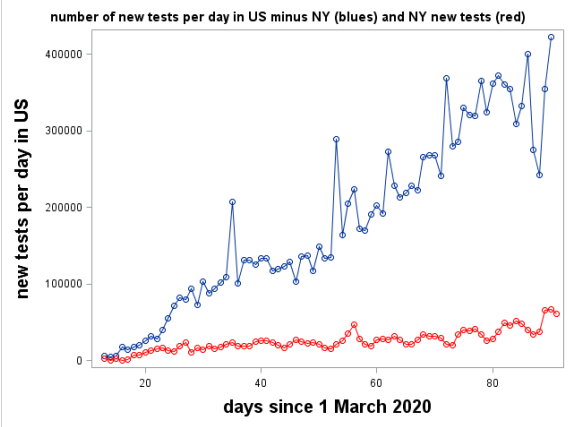


Fig. 1



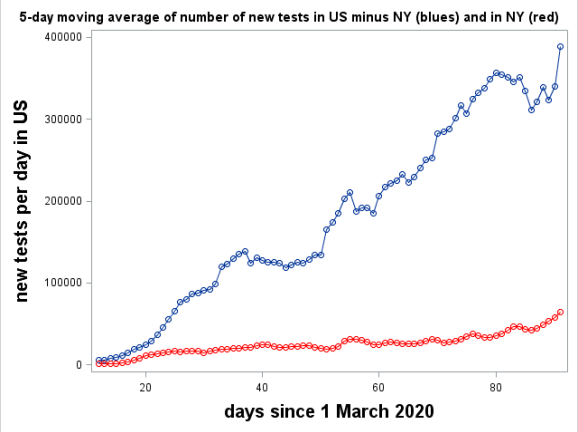
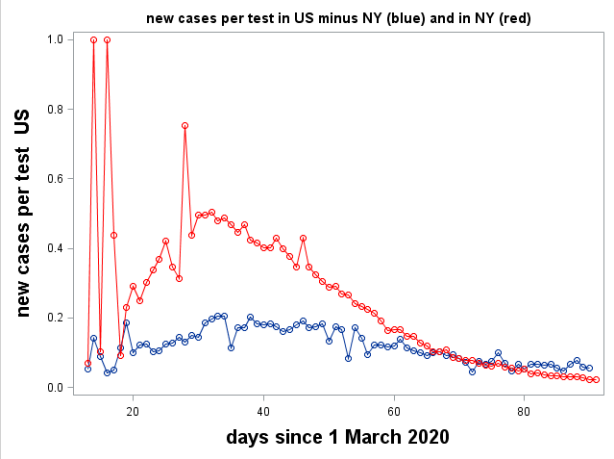


Fig. 2



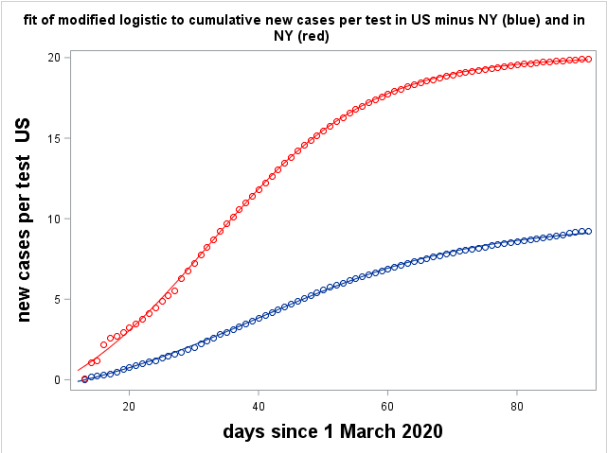


Fig. 3

Fig. 4

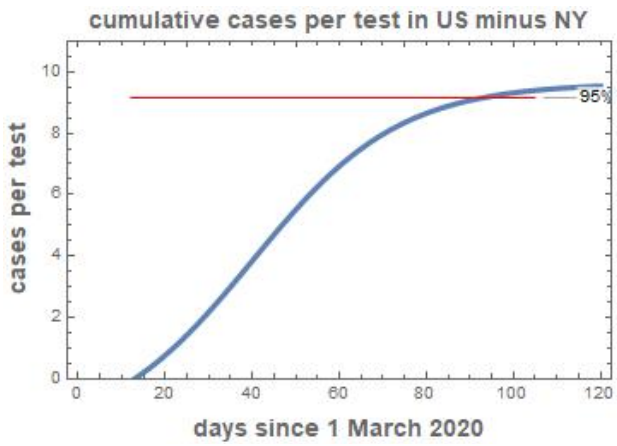


Fig. 5

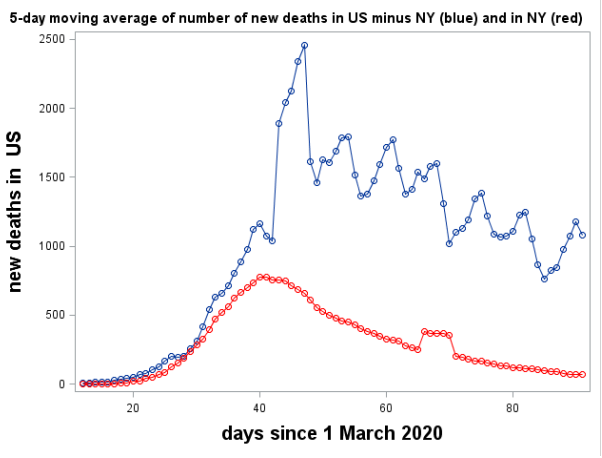


Fig. 6

III. Number of new deaths in US.

The good trend of a decline in death rates continues in the US outside of NY (and in NY) (Fig. 6). For the US, this trend is highly statistically significant ($P < 0.0001$) after day 45 (45 days ago). Extrapolation of this linear trend puts the "zero deaths" day around day 145 (23 June). As always, though, there is a substantial error associated with this projection—a zero death day of 266 (21 Nov) is within the margin of error.

A peak in deaths is expected to lag behind a peak in new cases, simply because there is time between detection and death. This expectation is seen in the data: the peak in new deaths per test occurs a week later than the peak in new cases. In that sense, the data from infections and deaths are consistent, reinforcing belief that the trends are real. In addition, this consistency suggests that the correction of new cases for number of tests performed is reasonable.

IV. Number of New Cases in NC

In North Carolina, the number of new reported cases per day continues to climb (Fig. 7). When this statistic is corrected for number of tests, the number of new cases per test has taken an ugly turn upward (Fig. 8). This trend began about 20 days ago, is highly significant ($P = 0.0018$), and coincides with "opening up" and relaxation of social distancing—a really bad trend.

Not surprisingly, given this trend, the data on cumulative new cases per test continues to deviate systematically from a modified logistic growth curve (Fig. 9)—the actually data is more linear from day 71 than the fit. Consequently, no projections will be made at this time.

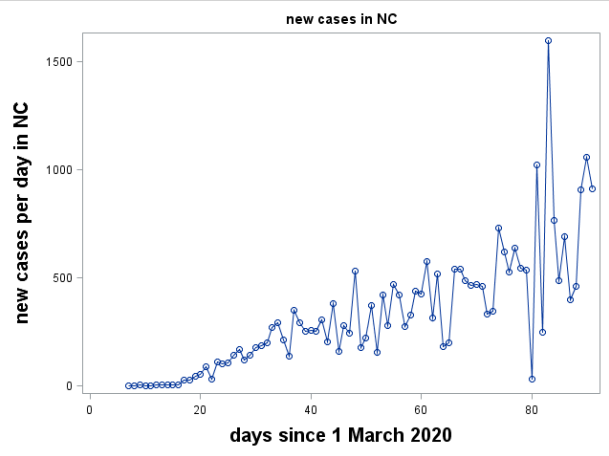


Fig. 7

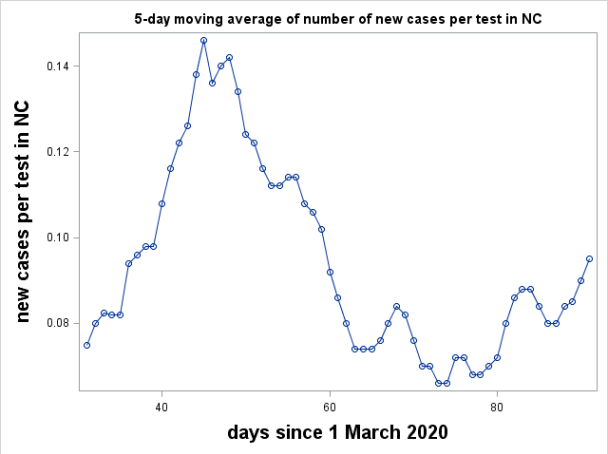


Fig. 8

V. Number of New Deaths in NC

The number of new deaths per day in NC had plateaued beginning around day 43. In the last week, however, there has been a spike in new deaths (Fig. 9). This is a lag of about a week and a half from the upturn in number of new cases per test and suggests that "opening up" in NC is leading to an increase in both new infections and new deaths.

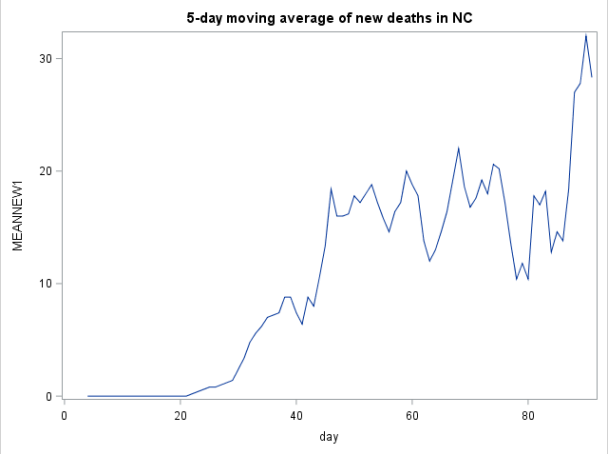


Fig 9

In conclusion, on the National level, there is room for some optimism: New deaths in the US continue to decrease. However, there is also substantial evidence for pessimism, both in the US and in NC. In the US, while the number of new cases per test had been declining, it seems to have reached a plateau. We will need to watch to determine whether it will turn down again or begin to increase. In NC, the number of new cases per test—which is probably a good estimate of actual number of new cases—appears to have taken an upturn, which is not a good sign for the foreseeable future. The same trend is evident in new deaths per day. These trends may constitute evidence that relaxation of social distancing practices is beginning to contribute to an increase in infection and death rates.