

# **An extended SIR model for COVID which illustrates the impact of shrinking Susceptible population on waves of infection.**

Michael P.C. Watts; Retired

Austin TX

## **Abstract**

As a highly infectious disease, the waves of Covid have been managed by a combination of cycles of social distancing and vaccination. As the pandemic progresses, a larger fraction of the population is effectively protected by prior infection or vaccination. We have modeled the dynamics of the pandemic with an extended SIR model that includes the effects of vaccination and asymptomatic infections. The simplest models of the daily case data in TX and NY were obtained when the asymptomatic infection rate was approximately 2.3x the symptomatic infection rate. We then use the model to illustrate that when social distancing controls the Reproduction Number, there will be a point at which the infection will naturally decay. Conversely, when either society starts to reopen or a new, more infectious variant arrives, the infection rate will increase until a new natural decay point is achieved consistent with the new Reproduction Number. Eventually cases in unvaccinated individuals drop and breakthrough cases constitute the majority of cases. Breakthrough infections occur throughout the both the vaccinated and the unvaccinated ex-infected populations. We model the breakthrough infection rate as if it were a new infection distributed across the entire protected population with an appropriately low Reproduction Number. Data from the UK shows trends consistent with almost complete protection in the unvaccinated by Sept 2021. The observed changes in Reproduction Number in TX are consistent with Google Mobility as a proxy for distancing and the spread of the Delta variant. We fit the daily death data with a time varying Case Fatality Rate and lag between cases and deaths that appears to be consistent health care improvements. .

## **Introduction**

The waves of the Covid pandemic have had a major economic and social costs around the world. These costs have led to extensive efforts to model and predict the evolution of the pandemic. The most widely used models, IMHE<sup>i</sup> and MRC<sup>ii</sup>, use a Bayesian analysis of the parameters in the infection rate models. This analysis yields statistical estimates of both the model parameters and their inferred distributions. With these distributions in hand, Monte Carlo simulations are then used to construct the range of likely scenarios which in turn guides both commercial and governmental decision making.

The approach in this extended SIR is to create a much simpler model that can be fitted to the daily case history, with the goal of providing some conceptual understanding of the underlying process. It is not an attempt at statistical forecasting.

The core deterministic model for infectious disease is the Susceptible, Infection and Recovery Model (SIR)<sup>iii</sup>. This extended SIR model attempts to capture 4 real life complications: social distancing and masking; the impact of Covid variants; the presence of asymptomatic infections; and ultimately the arrival of vaccines that reduce the size of the Susceptible population.

## SIR Model

The SIR model identifies 3 populations in a pandemic: the **Susceptible** who have not been infected; the **Infected**; and the **Recovered** who are now effectively immune. The new cases per day are given by the product of Susceptible and Infected and the infections rate constant  $K_i$ . The recovered per day is given by the product of Infected and the recovery rate constant  $K_r$ . The new infections per day are given by the new cases per day minus the newly recovered per day.

The ratio of the infection rate constant to the recovery rate constant is the Reproduction Number or the number of new infections each person passes on. The Reproduction Number ( $R_n$ ) is a measure of the “infectiousness” of the virus and is often used to characterize a disease; flu is mildly infectious  $R_n = 1.5-2$ . Measles is highly contagious and has an  $R_n = 12-18$ .

Any disease with a  $R_n$  value greater than 1 will grow exponentially just like compound interest or a nuclear chain reaction. The simplest way to analyze an exponential growth is to plot the values on a log scale, so that periods of constant growth appear as straight-line segments. For this reason, all data will be shown using log y axis versus time.

## Extended SIR model

To account for social distancing and variants that affect the infection rates, the infection rate constant is allowed to vary over time ( $K_i(t)$ ). Observations of social distancing during the pandemic would suggest that people adjust to changes and then maintain consistent behavior, so  $K_i(t)$  was modelled as a series of constant steps with an exponential transition between steps with a  $1/e$  width of 7 days. The  $K_i(t)$  values were then used to drive the calculation of daily cases. The Reproduction Number was calculated from  $R_n(t) = K_i(t)/K_r$ .

Since the start of the pandemic, it has been clear that there is a significant fraction of the population that get infected but are asymptomatic and therefore do not get tested. The model fit worked best with asymptomatic infections = 2.3x infections. Assuming that asymptomatic and symptomatic infection are equally effective in providing immunity, then S, I and R are divided into asymptomatic and symptomatic fractions.

Vaccination also reduces the Susceptible fraction of the population. If one injection provides protection equivalent to infection, then we can reduce S by the fraction of the population that was vaccinated at least once which we lag by 10 days to allow immunity to develop. Both protected and unprotected may choose to get vaccinated, so the daily vaccinated must multiplied by the fraction Susceptible so that there is no double counting of the protected.

Eventually cases in the unvaccinated will drop to the point that breakthrough cases dominate. Breakthrough infections will occur both among those who are unvaccinated but have already been infected, and among the fully vaccinated. We model these breakthrough infections as if they were new infections distributed across the entire protected population with an appropriately low Reproduction number.

The final model was simple enough to be programmed in an Excel spreadsheet.

### Fit to Texas data

The daily cases for Texas are shown in Figure 1a with the Reproduction Number ( $R_n(t)$ ) that gives the best “eyeball” fit to the waves of infection. The goal was the fewest changes in  $R_n(t)$ , so for example a single  $R_n(t)$  level that reproduced both the rise and fall of a wave was preferred over 2 or more  $R_n(t)$  levels.

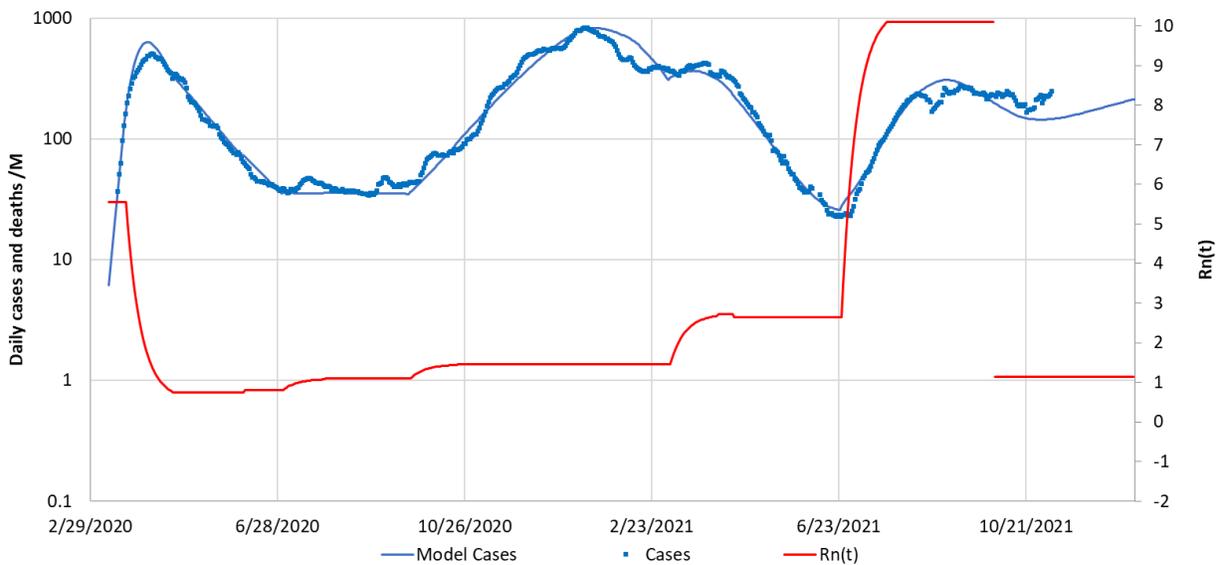
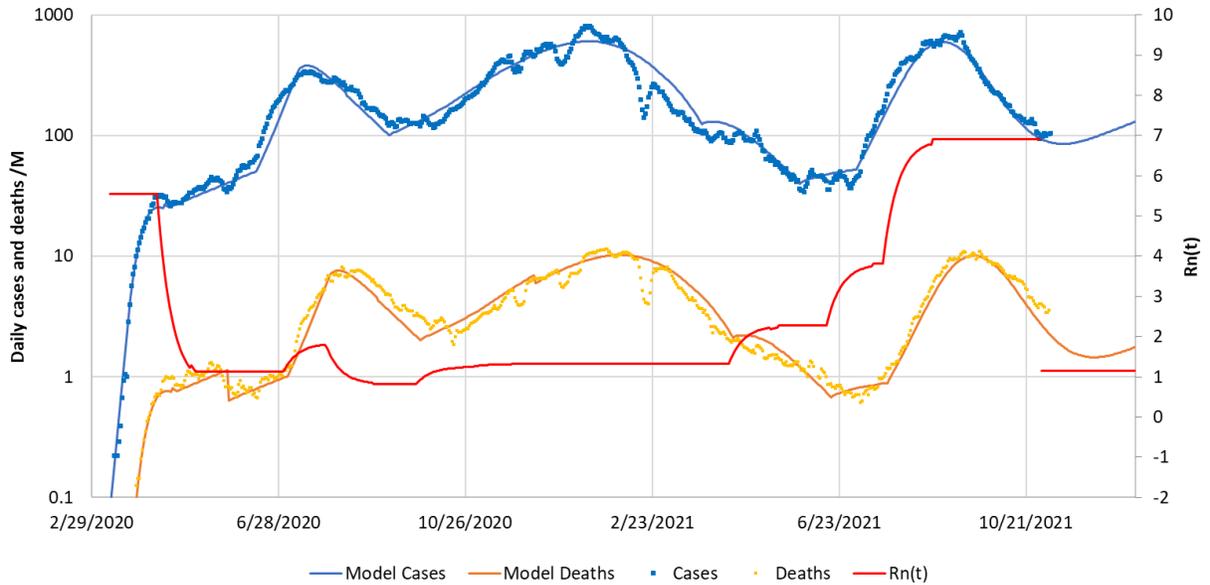


Figure 1 Case and Deaths per day per Million 7 day rolling average for TX a) and NY b). The Asymptomatic infections = 2.3x the Symptomatic infections. Population vaccinated at least once, lagged by 10 days are used to

*correct Susceptible population. The model uses a time varying Reproduction number ( $R_n(t)$ ). By Nov 2021, the fraction of previously infected individuals combined with the fraction of vaccinated individuals is so large that the progress of the infection is now dominated by breakthrough infections with a much smaller  $R_n(t) = 1.2$ . The deaths in TX were modelled by a time and lag varying Case Fatality Rate.*

There are many notable features in the TX data:

- 1) At the start of the pandemic, the daily case rate climbed by 10-fold in just 7 days. This rapid increase in infections implies that when the population is completely susceptible and does not follow social distancing, the Reproduction Number is 5.5.
- 2) A Reproduction Number of this size would in short order lead to a total collapse of the health care system, prompting strict social distancing, masking, working from home, and the closure of many public venues. The daily case rate fell after the introduction of these changes in social behaviors.
- 3) In July 2020, the first attempt to reopen resulted in a second wave with a Reproduction Number of 1.8. Again, social distancing and masking sharply reduced the daily case numbers.
- 4) In Sept 2020, after social distancing restrictions started to relax, a third slower wave began with an associated Reproduction Number of 1.3.
- 5) In Jan 2021, the modelled case rate started to fall without an accompanying change in the Reproduction Number. After the contribution of vaccinated people were included, further reducing the Susceptible population, the modelled cases rate matched the observed daily case rate.
- 6) In Sept 2021 saw a very rapid rise in cases that was best fit with a Reproduction Number of 7, higher than the initial virus. This coincided with the Delta variant taking over the US.
- 7) One month later, in October 2021, daily cases fell just as quickly as they had risen without any change in Reproduction Number. This behavior implies the Delta variant had burned through much of the remaining Susceptible population, or equivalently there were not enough unprotected people left to sustain the infection.
- 8) As of this writing (November 2021) the daily case rate has level off. We suggest this level case rate arises mostly through breakthrough infections in the already protected. Again, we model these breakthrough infections as if they were distributed across the whole population with a Reproduction number of 1.2. This  $R_n$  is consistent with UK data and will be refined as we accumulate more data.

Figure 1b shows a similar fit to NY data. NY had had a larger first wave, a correspondingly smaller September 2021 wave and is missing the second TX wave. Once we account for these obvious differences, the implied models are surprisingly similar and illustrate the life cycle of the pandemic.

By now, many of the unvaccinated will have been infected either knowingly or unknowingly, so there will be increasing immunity in the nominally unvaccinated. We used our model for to compare the number of unprotected ( $1-S$ ) and the number of unvaccinated. As of September 2021, roughly 90% of the unvaccinated are in fact protected. Under the assumption that vaccination and infection confer similarly immunity, a vaccine with 95% efficiency will appear to be only 60% efficient.

We note that similar indications of increasing protection in the unvaccinated have already appeared in the UK. As of September 2021,<sup>iv</sup> the vaccinated now represent about half of those hospitalized, the cases per day per million vaccinated are similar to the cases per day per million unvaccinated, and 98%

of blood donated has Covid antibodies. Effectively, nearly everyone in the UK is protected to at least some degree. It is important to note the unvaccinated are twice as likely to be hospitalized or die as the fully vaccinated, so vaccination is still highly beneficial.

### Root Causes for changes in Reproduction Number

The Reproduction Number starts high, falls with social distancing, has small peaks associated with re-openings, and then rises back high as the infection spreads through a progressively smaller susceptible population, suggesting that the pandemic has run its course. Social distancing and masking are difficult to measure, so early in the pandemic the Google Mobility metrics<sup>y</sup> were started in which people allowed access to their phone's location information. By identifying where the individuals went, Google created metrics for travel to different locations. Google then compared this metric from just prior to the pandemic as a baseline to obtain a % reduction for each metric.

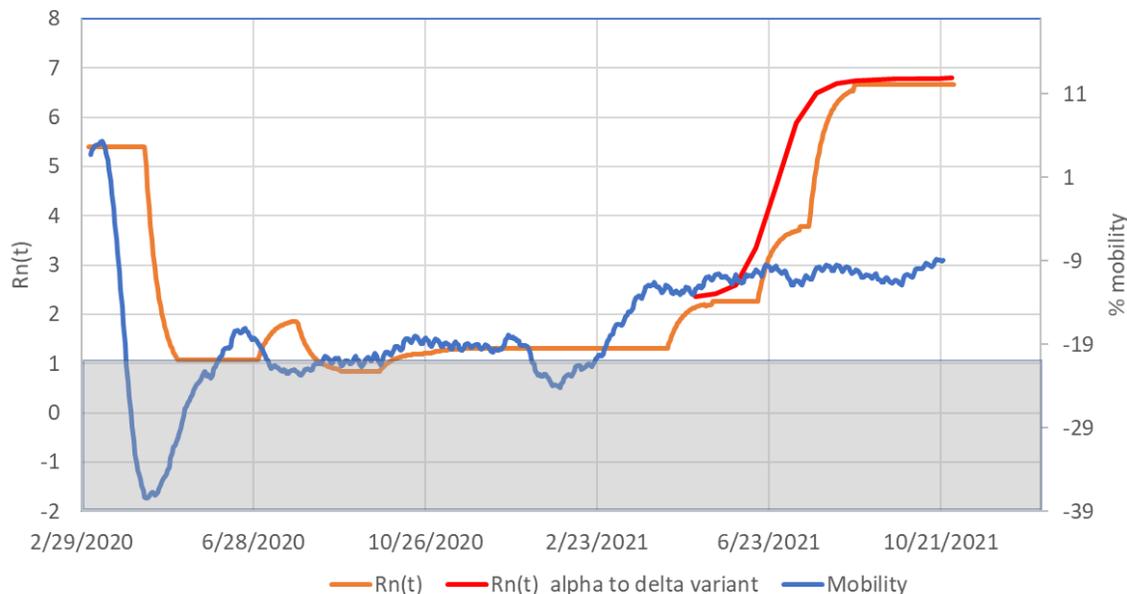


Figure 2 Google Mobility as a proxy for social distancing and onset of Delta variant as root causes of the changes in  $R_n(t)$  in TX. Mobility of -20% effectively stalls the infection at  $R_n(t) = 1$  indicated by the shaded area.

We averaged and smoothed the TX metrics for shopping, transit, and workplace to produce our own Mobility metric. Our smoothing reduced or eliminated the large, rapid fluctuations in Mobility associated with the holidays and Super Bowl.

The Reproduction Number and mobility metrics are plotted in Figure 2, with the mobility scale adjusted to match % mobility to  $R_n(t)$ . The shaded area indicates where  $R_n(t)$  is less than 1, meaning that the virus is not propagating. If the virus is unable to propagate, then further social distancing will have no

effect. As a result,  $R_n(t)$  less than 1 will be independent of the mobility changes, indicated in the shaded area when % mobility is less than -20%. The correlation coefficient between  $R_n(t)$  and % mobility (lag 10 days) was statistically significant ( $R=0.82$ ) for the data up to 5/3/21 and  $R_n(t) > 1$ .

The data on the spread of the Delta variant<sup>vi</sup> was added to Figure 2 as an addition to  $R_n(t)$  assuming Delta variant is 2x as infectious as the Alpha.

The combination of Mobility and Delta variant appears to explain the changes in  $R_n(t)$ , suggesting that Mobility is a useful proxy for social distancing. It is worth noting that in Sept 2021, Mobility is still around -10% of the baseline, indicating that not everything has reopened to pre-pandemic levels.

### **Fatality Data**

The daily case and death data in Figure 3a, show that deaths in wave 1 lagged cases by 5 days. By August 2020, the lag had increased to 20 days. The Case Fatality Rates (CFR), at different lags are shown in Figure 3b, from May 5 to June 2020 CFR = 3% at a lag of 5 days. After June, CFR = 1.7% as a lag of 20 days. This seems consistent with improvements in health care after the initial wave of the infection. These CFR and lags were applied to the daily case data in Figure 1a to produce the model result that tracks the daily death data. Breakthrough infections are expected to result in a further reduction in CFR as shown in Figure 1a.

### **Conclusions**

There is significant protection in the nominally unvaccinated in TX and NY because there has been a large population of asymptomatic infections. When most of the unvaccinated are protected by prior infection, the breakthrough infections possibly form a second infection with much smaller Reproduction Number. Google Mobility Metrics are an effective proxy for social distancing, and in combination with variant spread seem to explain the major features of the pandemic. Once health care strategies had stabilized, Case Fatality Rate on a 20-day lag has been 1.7%.

### **About the author**

Michael P.C. Watts PhD is a retired material scientist who spent his career in semiconductor technology. This project is an extension of his long-term interest in using simple models to understand complex processes.

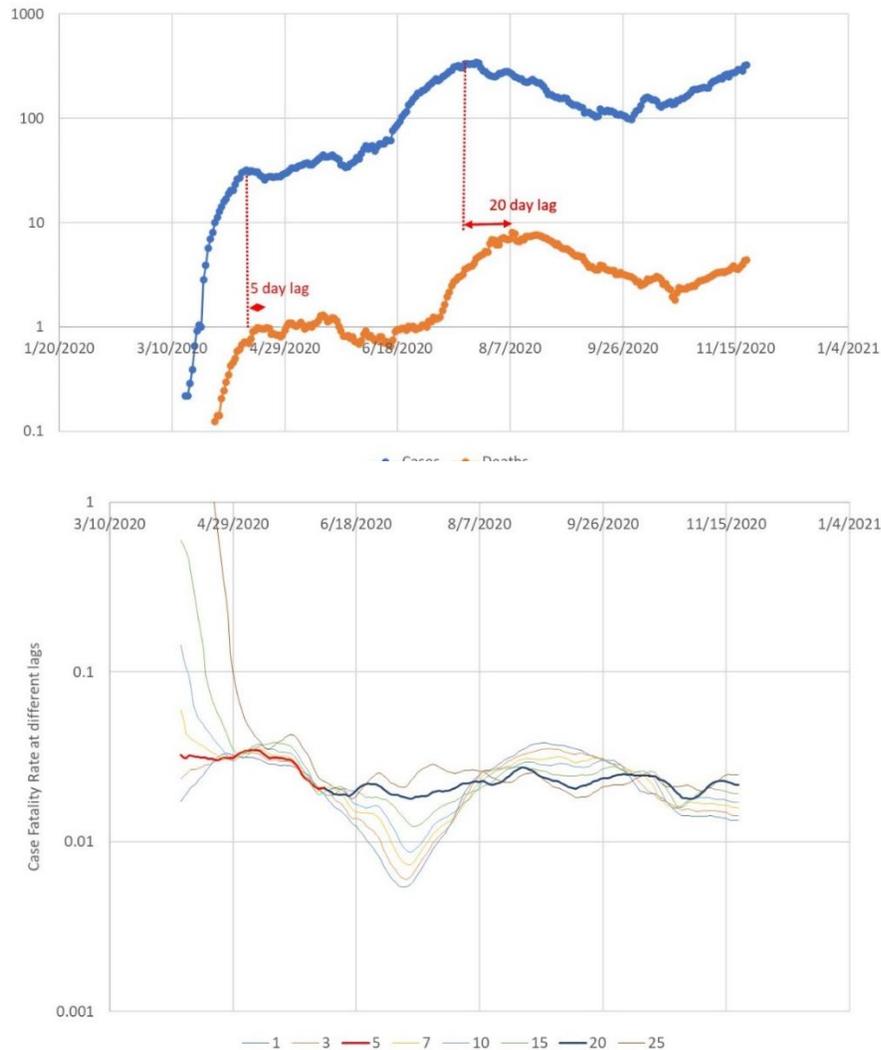


Figure 3 Varying lags between cases and fatalities. a) Cases and Deaths per day in 2020. The peaks show the changes in lag as health care improves. b) Case Fatality rates at the different lags, the best fit lags are shown in bold lines; CFR = 3% lag 5 days with a transition to CFR = 1.7% lag 20 days in July 2020.

<sup>i</sup> IMHE <http://www.healthdata.org/research-article/mapping-global-trends-vaccine-confidence-and-investigating-barriers-vaccine-uptake>

<sup>ii</sup> MRC <https://www.mrc-bsu.cam.ac.uk/>

<sup>iii</sup> SIR model <https://digitalcommons.usf.edu/cgi/viewcontent.cgi?article=1027&context=ujmm>

<sup>iv</sup> UK Vaccination report week 38

[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1029606/Vaccine-surveillance-report-week-38.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1029606/Vaccine-surveillance-report-week-38.pdf)

<sup>v</sup> Google Mobility <https://www.google.com/covid19/mobility/>

<sup>vi</sup> Delta variant data from "COVID-19\_Vaccinations\_in\_the\_United\_States\_Jurisdiction"