

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/339915690>

# Quantifying undetected COVID-19 cases and effects of containment measures in Italy

Preprint · March 2020

DOI: 10.13140/RG.2.2.11753.85600

CITATIONS

0

READS

5,203

2 authors:



[Morten Gram Pedersen](#)

University of Padova

84 PUBLICATIONS 1,183 CITATIONS

[SEE PROFILE](#)



[Matteo Meneghini](#)

University of Padova

402 PUBLICATIONS 5,663 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Potential Induced Degradation of Bifacial Solar Cells [View project](#)



A multicellular computational model of glucagon secretion from pancreatic islets [View project](#)

**Title:**

# Quantifying undetected COVID-19 cases and effects of containment measures in Italy

**Short title:** Undetected COVID-19 cases and containment in Italy

**Authors:**

Morten Gram Pedersen<sup>1,2,\*</sup>, Matteo Meneghini<sup>1</sup>

**Author Affiliated Institutions:**

1) Department of Information Engineering, University of Padova, Italy

2) Department of Mathematics “Tullio Levi-Civita”, University of Padova, Italy

\*) Corresponding author: pedersen@dei.unipd.it, (+39) 049 827 7961

**Abstract**

Italy is the Western country most affected by the COVID-19 disease, and the first to implement drastic containment measures. We propose to model COVID-19 dynamics with a SIQR (susceptible – infectious – quarantined – recovered) model, since the confirmed positive cases are isolated and do not transmit the disease. We estimate model parameters by fitting model expressions to official Italian data and predict that the number of unidentified SARS-nCov2-positive individuals is  $\sim 4$  times the number of confirmed cases. Our analysis predicts that recent drastic restrictions have reduced virus spreading modestly but insufficiently to halt the epidemic. We then simulate possible scenarios to predict how strict new containment measures must be in order to invert the trend of COVID-19 spreading. We conclude that extreme social distancing is needed to contain the disease because of the large amount of undetected, infectious individuals in the absence of social-wide testing for SARS-nCov2.

**Introduction**

The COVID-19 disease due to the SARS-nCov2 coronavirus is spreading rapidly across the globe since its outbreak in China, and was declared a pandemic by the WHO on March 11, 2020. After the first severe patient was brought to the hospital of Codogno, Italy on February 20, 2020, and subsequently tested positive for SARS-nCov2, a rapidly increasing number of patients have been identified, initially in Northern Italy and later in the rest of the country and Europe. Italy is the most affected Western country, with more than 41.000 confirmed cases, a death toll that exceeds China, and the first to implement drastic measures in an attempt to contain the disease. It is therefore necessary to analyze the Italian data carefully, and in particular to investigate if and how the limitations in activities affect the disease dynamics.

Simple mathematical models of infectious diseases are useful for providing insight into disease dynamics, and compared to more complex models, can be fitted to data with a minimum number of assumptions on model parameters. However, even simple models should respect that nature of the data. There is thus a compromise between using a parsimonious model but sufficiently complex to be based on correct underlying assumptions.

In our setting, to fit the data on identified SARS-nCov2 positive cases, a SIQR (1) is appropriate. In this model, infected individuals may be isolated, entering the “quarantined” subpopulation Q, so that these individuals no longer transmit the disease. Since Italian positive cases have been put in isolation (in hospitals or at home) immediately, the revealed data of active cases thus correspond to the number of individuals in state Q. It would not be correct to confront the infectious subpopulation of a SIR/SEIR (susceptible – infectious – [exposed] – recovered) model with the recorded data, since it is unknown how many infectious but undetected individuals are circulating in the population. As will be clear in the following, besides being conceptually clearer, the explicit quantification of the quarantined individuals will allow us to provide estimates of the infectious cases at the very beginning of the epidemic outbreak as well as of the current number of undetected infectious individuals.

## Results

To parameterize the SIQR model, we fitted analytically derived expressions for  $Q + R$  to the data of total COVID-19 cases in Italy (Fig. 1). Although a model with constant parameters was found to give a reasonable fit, it overestimated the last data points, which indicated that the containment measures might have changed for example the transmission rate  $\beta$ . We therefore fitted a model with a piecewise-constant  $\beta$ , taking one value before a given date and another value after that date. We found that the best fit was obtained when  $\beta$  changed March 8, 2020, in excellent agreement with the restrictions in activities imposed on this date (DPCM 8 marzo 2020) in particular for the Northern, most affected Italian regions. Before March 8, 2020,  $\beta - \eta - \alpha$  was estimated to be 0.213/day (SE 0.003/day), which was lowered by 0.093/day (SE 0.011/day,  $p < 0.0001$ ) to  $\beta - \eta - \alpha = 0.120$  from March 8, 2020. The other free parameter in the fit was estimated to be  $\eta I_0 = 54.8/\text{day}$  (SE 1.5/day).

FIGURE 1 about here

Our model has only two free parameters to be determined from other data. The estimated values for the cure rate of non-identified cases,  $\alpha = 0.067/\text{day}$ , and the rate of isolation,  $\eta = 0.067/\text{day}$ , (see Methods) yield  $\beta = 0.347/\text{day}$  before March 8, 2020, and  $\beta = 0.254/\text{day}$  from March 8, 2020. Consequently, the basic reproduction number is estimated to  $R_0 = \beta/(\alpha + \eta) = 2.59$  before the strict measures were imposed, in line with previous estimates of  $R_0$  falling between 2 and 4 (2–6). Since March 8, 2020, the basic reproduction number is  $\sim 27\%$  lower ( $R_0 = 1.90$ ). Further, we obtain an estimate of the number of infectious individual at the moment of the outbreak of

$I_0 = 54.8/0.067 = 822$ , but this number obviously depends on the value of  $\eta$ , which reasonably falls within  $(0.02, 0.10)$ , yielding  $I_0 \in (550, 2740)$ . Based on the dynamics before the outbreak ( $\delta = 0$ ), we can estimate that the first infectious case appeared in Italy  $\log(I_0)/(\beta - 0.1) \approx 27$  days before the outbreak, i.e., around January 26, 2020. In the absence of isolation, our model reduces to a SIR model with the natural (i.e., in the absence of isolation and other measures)  $R_0 = \beta/0.1 = 3.47$ , which falls within the range of previous estimates.

We simulated the dynamics of the SIQR model (Fig. 2) and found that the Italian data was very well fitted when assuming this piecewise-constant time-variance in  $\beta$ , indicating that the early Italian containment measure (school closings, hygiene indications, etc.) had negligible, if any, effect on the disease dynamics until March 8, 2020, whereas the stronger limitations introduced on this date have had some effect, but appear to be insufficient to invert the propagation of COVID-19 in Italy, since  $R_0$  is still greater than one.

The model predicts that, after the initial transient and up to March 8, 2020, for each patient in quarantine approximately four infectious individuals are present in the populations ( $Q/I \approx 0.24$ ). From Eq. 3 and Eq. 5 it can be shown analytically that following the transient and during the exponential growth,  $Q/I = \eta/(\gamma + \beta - (\alpha + \eta)) = \eta/0.28 = 0.239$ . In other words, the relation between  $Q$  and  $I$  follows directly from the fit to the data up to the factor  $\eta$ , while it is independent of the assumptions on the other individual parameters. With the range  $\eta \in (0.02, 0.10)$  used above, we obtain that  $I$  is 3–13 fold larger than  $Q$ . This range is in agreement with a recent study on the number of unidentified infectious individual during the epidemic in China (7). Following the measures of March 8, 2020, the  $I$ -to- $Q$  ratio has fallen to  $\sim 2.8$  (1.9–9). Thus, there are still many unidentified cases that do not appear in the official statistics.

FIGURE 2 about here

We then simulated alternative scenarios that may be the result of the drastic measures imposed from March 8, 2020 (Fig. 3). If no parameters of the model had changed since March 8, 2020, the exponential growth would have continued and the number of identified SARS-nCoV2 positive cases ( $Q$ ) would have reached almost 1 million by April 1, 2020. If interactions, and hence  $\beta$ , are by reduced 10–50% the dynamics is slowed down, but  $I$  and  $Q$  are still increasing (Fig. 3). In particular, with the parameters estimated and used for the simulation in Fig. 2,  $Q$  will arrive at 185,000, whereas  $I$  will reach 430,000, by April 1, 2020. To stop the epidemic spread,  $\beta$  must be reduced by  $\sim 65\%$  compared to the pre-March 8 value, so that  $R_0 < 1$ . However, to reduce the number of infectious individuals  $I$  in a reasonable timeframe,  $\beta$  must be reduced by 90%. Nonetheless, even if such a strong reduction had been the result of the limitations introduced on March 8, 2020, three weeks from the introduction of the restrictions, i.e. by April 1, 2020, the number of positive cases would not have dropped and be  $\sim 20,000$ .

FIGURE 3 about here

Do our conclusion carry over to other countries? Without changing parameters the model provides excellent fits to data from Germany, Spain and France, where no strict measures have been imposed only very recently (Fig. 4). To fit the data from China, we assumed that  $\beta$  changed from the default value to 0.1/day ( $\sim 70\%$  reduction) on day 12, which fits reasonably with the fact that China imposed restrict containment measures in late January 2020.

FIGURE 4 about here

## Discussion and conclusions

From our analysis it is clear that the early decrease in slope seen in the log-transformed data (Fig. 1) is not due to the measures due to school closings etc. imposed during the first weeks after the outbreak in Italy, but a direct consequence of the model structure. The stronger limitations imposed from March 8, 2020, are needed to lower  $R_0$  through a decreased transmission rate  $\beta$  in order to reduce the exponential growth rate. From the current data we estimated that  $R_0$  has been lowered by  $\sim 30\%$ , reducing the doubling time from  $\sim 4$  to  $\sim 6$  days, but insufficiently to halt the progression of the epidemic. With nearly 10% of identified patients requiring intensive care, and  $\sim 5000$  intensive care beds in Italy before the COVID-19 outbreak, it is clear that it is utmost important that the number of infected individuals needs to be contained very soon, in particular to lower the pressure on intensive care units and hospitals (8).

We estimate that a reduction of  $\sim 70\%$  of  $R_0$  was obtained in China by a combination of serious restrictions and widespread testing. Such a reduction in  $R_0$  would be sufficient to contain the disease in Italy. It has to be seen whether further restrictions in Italy obtain the same reduction of  $\beta$ , but in our opinion this seems unlikely in the absence of wider testing for COVID-19. Such testing would correspond to an increase in  $\eta$  and a lowering of  $\alpha$  in our model. Our model approach can be used to estimate the effect of these restrictions as data appears. These conclusions for Italy carry over to other Western countries (Fig. 4).

With our model approach, which notably has only two free parameters to estimate, we estimate that the number of infectious individuals  $I$  is approximately 4 times larger than the number of isolated patients  $Q$  (Fig. 2). This fact follows to a large extent directly from the data, and therefore hold for other countries where our parameters give excellent fits to the public data (Fig. 4).

Based on estimated confirmed case fatality rates of 5–8 % based on data from China (9) this implies that the overall fatality rate is 1–2 %. However, it should be noted that the case fatality rate appears to be higher in Italy, since the number of COVID-19 fatalities in Italy has surpassed the Chinese death toll in spite of a lower number of confirmed COVID-19 cases in Italy.

We were also able to estimate that there were several hundreds of infectious but undetected infectious individuals in Italy at the time of the outbreak around February

21, 2020. Only when a patient with severe symptoms was hospitalized and tested for SARS-nCov2, and the first infected person died from COVID-19 on the following day, wide testing and isolation efforts started. We estimate that isolation has reduced  $R_0$  from  $\sim 3.5$  to  $\sim 2.6$ . By backward interpolation, we estimated that the first infectious individual (“patient zero”) appear in Italy around January 26, 2020. This estimate coincides remarkably well with recent reports on “patient zero” being a German citizen linked to a cluster in Munich (10) visiting Italy around January 25, 2020 (11).

## Methods

We use a SIQR model (1) to describe COVID-19 dynamics in Italy. The model equations are

$$\frac{dS}{dt} = -\beta SI/N, \quad (1)$$

$$\frac{dI}{dt} = \beta SI/N - (\alpha + \eta)I, \quad (2)$$

$$\frac{dQ}{dt} = \eta I - \gamma Q, \quad (3)$$

$$\frac{dR}{dt} = \gamma Q, \quad (4)$$

where  $R$  models SARS-nCov2-positive, isolated individuals that recover or die from the disease. We do not explicitly model the number of recovered or deceased non-identified COVID patients, but only the rate  $\alpha$  with which these patient become non-infectious. Further,  $N$  is the total number of individuals in the population, assumed constant since we are studying the early phase of the epidemic.

Since there is evidence that the SARS-nCov2 virus can be transmitted in the absence of symptoms (10, 12), we do not include an explicit exposed-but-noninfectious (E) state, i.e., we do not consider a SEIQR model (13). Further, from the Diamond Princess cruiseship, it has been found that  $\sim 50\%$  of SARS-nCoV2-positive individuals do not develop symptoms (14). We assume that such positive but asymptomatic individuals can transmit the disease, i.e., the  $I$  state includes both individuals that will not develop symptoms, cases that did not develop symptoms yet, and symptomatic patient that still have not been tested positive and isolated.

So far a relatively small fraction of the Italian population has been found positive for COVID-19. Thus, we are still in the early phase of the epidemic where  $S \approx N$ , and as well known in the case of constant parameters, the number of infected individuals follows exponential growth

$$\frac{dI}{dt} \approx (\beta - (\alpha + \eta))I, \quad \text{i.e.,} \quad I(t) = I_0 \exp\{(\beta - (\alpha + \eta))t\}, \quad (5)$$

where  $I_0$  is the number of infectious individuals at the beginning of the outbreak  $t = 0$ .

As mentioned, the number of individuals that have been found SARS-nCov2 positive and put in isolation does not correspond to  $I$  but to  $Q + R$ . From Eq. 5 it follows that  $Q + R$  follows

$$\frac{d(Q + R)}{dt} = \eta I = \eta I_0 \exp\{(\beta - (\alpha + \eta))t\}, \quad (6)$$

which yields

$$(Q + R)(t) = \frac{\eta I_0}{\beta - (\alpha + \eta)} (\exp\{(\beta - (\alpha + \eta))t\} - 1). \quad (7)$$

Similar expressions are readily obtained for a piece-wise constant  $\beta$ .

Using the nls function in R (15), we fitted these expression to the log-transformed Italian COVID-19 data from February 22, 2020 through March 19, 2020 (Fig. 1), and estimate  $\beta - \eta - \alpha = 0.187/\text{day}$  (SE 0.004/day) and  $\eta I_0 = 63.1/\text{day}$  (SE 2.8/day) in case  $\beta$  is constant. With piecewise constant  $\beta$ , we fitted the expression for every possible change-date, and chose the change-date (March 8, 2020) that resulted in the fit with the lowest residual standard error. Before March 8, 2020,  $\beta - \eta - \alpha$  was estimated to be 0.213/day (SE 0.003/day), which was lowered by 0.093/day (SE 0.011/day,  $p < 0.0001$ ) to  $\beta - \eta - \alpha = 0.120$  from March 8, 2020, whereas  $\eta I_0$  was estimated to be 54.8/day (SE 1.5/day). Thus the model with piecewise constant  $\beta$  fitted the data significantly better.

The recovery/death rate  $\gamma$  was fit by linear regression of the day-to-day differences in the number of recovered/deceased cases ( $\Delta R_i = R_i - R_{i-1}$ ) on the number of active cases on the corresponding day ( $Q_{i-1}$ ). We estimate  $\gamma = 0.036/\text{day}$  (SE 0.003/day), indicating that isolated and hospitalized patients recover or die in approximately 25 days, which is consistent with the median time-to-death being 18.5 days (15.0–22.0) in Wuhan, China (16).

To identify the other individual parameters we use previous findings. It has been estimated that the average incubation time is  $\sim 5$  days (2, 17) and the duration of the milder cases of disease is 5-10 days (12). We assume an average time of duration from infection to recovery or death of non-isolated cases of 10 days, corresponding to a rate of 0.1/day. If a fraction  $\delta$  of infectious individuals is tested positive and put in isolation, we obtain  $\alpha = (1 - \delta) \times 0.1/\text{day}$ .

Similarly  $\eta$  related to the time until patients are tested positive and isolated, but also to the fraction of all infectious individuals that are tested positive. These are mostly symptomatic patients, which we assume are isolated soon after the incubation time is over and first symptoms appear, i.e., after  $\sim 5$  days. Letting  $\delta$  denote the fraction of infectious individual entering  $Q$ , we obtain  $\eta = \delta \times 0.2/\text{day}$ .

Since  $\sim 50\%$  of the population is asymptomatic (14), but some milder cases may also go unnoticed and not end in isolation, we assume that  $\delta = 1/3$  of infectious individuals are tested after an average of 5 days. We thus set  $\alpha = \eta = 0.067/\text{days}$ .

## References

- [1] Hethcote H, Zhien M, Shengbing L. Effects of quarantine in six endemic models for infectious diseases. *Mathematical biosciences*. 2002;180(1-2):141–160. doi:10.1016/S0025-5564(02)00111-6
- [2] Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med*. 2020. doi:10.1056/NEJMoa2001316
- [3] Zhou T, Liu Q, Yang Z, Liao J, Yang K, Bai W, et al. Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *J Evid Based Med*. 2020 Feb;13(1):3–7. doi:10.1111/jebm.12376
- [4] Zhao S, Lin Q, Ran J, Musa SS, Yang G, Wang W, et al. Preliminary estimation of the basic reproduction number of novel coronavirus (2019-nCoV) in China, from 2019 to 2020: A data-driven analysis in the early phase of the outbreak. *International Journal of Infectious Diseases*. 2020;92:214–217. doi:10.1016/j.ijid.2020.01.050.
- [5] Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of travel medicine*. 2020. doi:10.1093/jtm/taaa021
- [6] Lai A, Bergna A, Acciarri C, Galli M, Zehender G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *Journal of medical virology*. 2020. doi:10.1002/jmv.25723
- [7] Li R, Pei S, Chen B, Song Y, Zhang T, Yang W, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science*. 2020. doi:10.1126/science.abb3221
- [8] Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet*. 2020. doi:10.1016/S0140-6736(20)30627-9
- [9] Jung SM, Akhmetzhanov AR, Hayashi K, Linton NM, Yang Y, Yuan B, et al. Real-Time Estimation of the Risk of Death from Novel Coronavirus (COVID-19) Infection: Inference Using Exported Cases. *J Clin Med*. 2020 Feb;9(2). doi:10.3390/jcm9020523
- [10] Rothe C, Schunk M, Sothmann P, Bretzel G, Froeschl G, Wallrauch C, et al. Transmission of 2019-nCoV infection from an asymptomatic contact in Germany. *New England Journal of Medicine*. 2020;382:970-971. doi:10.1056/NEJMc2001468
- [11] Kupferschmidt K. Genome analyses help track coronavirus' moves. *Science*. 2020 March;367:1176–1177. doi:10.1126/science.367.6483.1176



- [12] Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. *Jama*. 2020. doi:10.1001/jama.2020.2565
- [13] Jumpen W, Wiwatanapataphee B, Wu Y, Tang I. A SEIQR model for pandemic influenza and its parameter identification. *International Journal of Pure and Applied Mathematics*. 2009;52(2):247–265.
- [14] National Institute of Infectious Diseases Japan. Field Briefing: Diamond Princess COVID-19 Cases, 20 Feb Update. Last accessed March 12, 2020. Available from: <https://www.niid.go.jp/niid/en/2019-ncov-e/9417-covid-dp-fe-02.html>.
- [15] R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. Available from: <https://www.R-project.org/>.
- [16] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet*. 2020. doi:10.1016/S0140-6736(20)30566-3
- [17] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. *Ann Intern Med*. 2020. doi:10.7326/M20-0504

## Acknowledgments

**Funding:** this work was supported by MIUR (Italian Minister for Education) under the initiative "Departments of Excellence" (Law 232/2016).

**Authors contributions:** MGP and MM conceived research and discussed all results. MGP developed the theoretical framework and performed parameter estimation. MM developed the simulation framework and compared real data to model results. MGP wrote the paper. MM commented drafts and approved the final version of the paper.

**Competing interests:** The authors declare to have no competing interests.

## Figure legends

Figure 1: The total number of COVID-19 positive cases (circles) and the best fit of Eq. 7 with  $\beta$  constant (black curve;  $\beta - \eta - \alpha = 0.187/\text{day}$ ,  $\eta I_0 = 63.1/\text{day}$ ) and the corresponding expression with piecewise constant  $\beta$  (red curve;  $\beta - \eta - \alpha = 0.213/\text{day}$  (before March 8, 2020, indicated by the dashed line)  $\beta - \eta - \alpha = 0.120$  (from March 8, 2020),  $\eta I_0 = 54.8/\text{day}$ ). Data from <https://github.com/pcm-dpc/COVID-19/blob/master/dati-andamento-nazionale/dpc-covid19-ita-andamento-nazionale.csv>.

Figure 2: Simulation of the SIQR model with default parameters as given in the main text and  $N = 60$  million, starting from date of the Italian outbreak February 21, 2020 (red:  $S$ ; green:  $I$ ; pink:  $Q$ ; blue:  $R$ ). Recorded data are shown with dots (blue: active, isolated cases; red: sum of recovered and deceased cases).

Figure 3: Simulation of the SIQR model as in Fig. 2, but assuming that  $\beta$ , and hence  $R_0$ , is reduced by 0, 10, 20, 30, 40, 50, 60, 70, 80 or 90% (upper to lower curve) on March 8, 2020 (day 16).

Figure 4: Simulation of the SIQR model as in Fig. 2, changing only  $N$ ,  $I_0$ , and redefining  $t = 0$  to correspond to the beginning of the outbreak in France, Germany, Spain, and the Hubei province in China. For Hubei  $\beta$  was lowered from the default value to  $0.1/\text{day}$  on day 12. Data from [https://github.com/CSSEGISandData/COVID-19/tree/master/csse\\_covid\\_19\\_data/csse\\_covid\\_19\\_time\\_series](https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series).

Figure 1

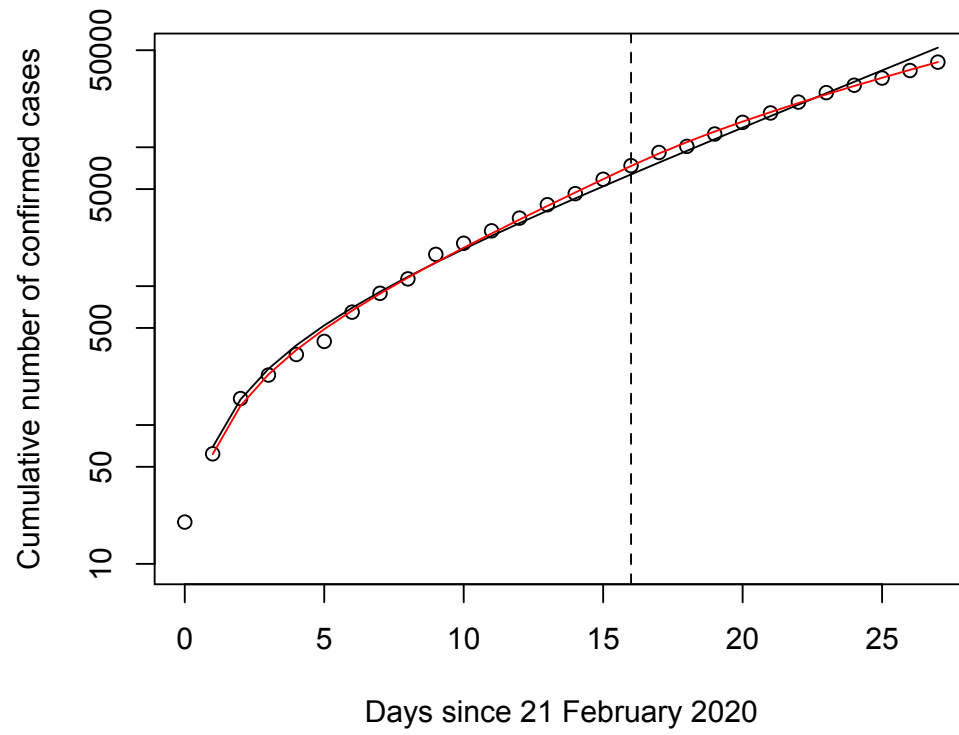


Figure 2

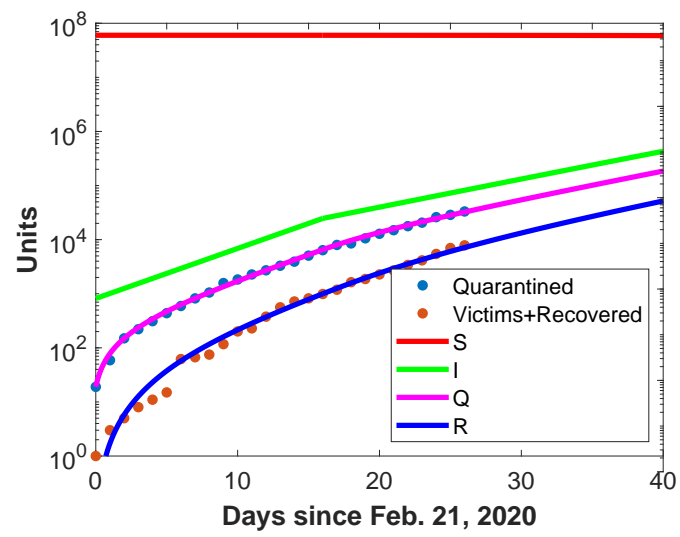


Figure 3

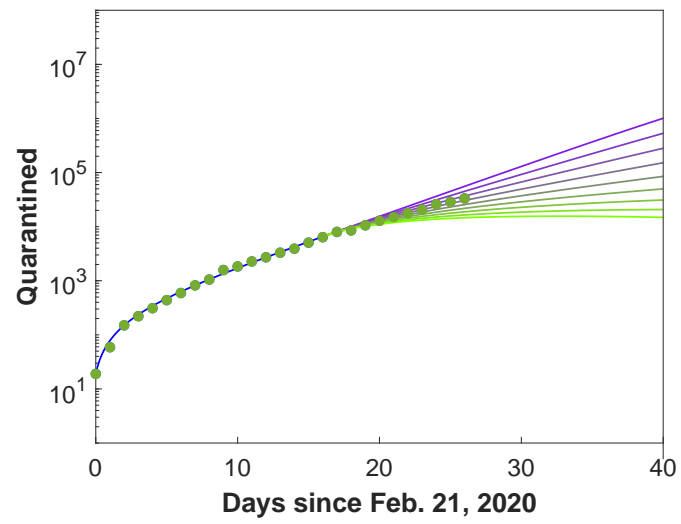


Figure 4

