

Data-driven estimation of change points reveal correlation between face mask use and accelerated curtailing of the COVID-19 epidemic in Italy

Morten Gram Pedersen^{1,2*}, Matteo Meneghini¹

¹ Department of Information Engineering, University of Padova, Italy

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

* pedersen@dei.unipd.it

Abstract

Italy was the first Western country to be seriously affected by COVID-19, and the first to implement drastic measures, which have successfully curtailed the epidemic. To understand which contain measures effected disease dynamics, we estimate change points in COVID-19 dynamics by fitting a compartmental model to official Italian data. Our results indicate that lockdowns managed to cause the epidemic to peak in late March 2020. Surprisingly, we found a change point during the decay from the peak, which does not correspond to obvious drastic legal interventions, but may be explained by widespread promotion and mandatory use of face masks. We confirm these interpretations at regional levels, and find that the gradual reopening of society since early May has caused no change in disease dynamics. We speculate that widespread use of face masks and other protective means has contributed substantially to keeping the number of new Italian COVID-19 cases under control in spite of society turning towards a new normality.

Introduction

The COVID-19 disease due to the SARS-CoV-2 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 COVID-19, 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 one of the most affected European country, with ~ 240.000 confirmed cases and nearly 35.000 COVID-19 related deaths, and was the first to implement drastic contain measures. The imposed restrictions, culminating with complete lockdowns, have turned out to be effective in controlling the epidemic in Italy; the number of new daily cases peaked in late March 2020 at ~ 6000 and declined to $\sim 200 - 300$ by early June. The limitations in activities were followed by milder orders and direct invitations to behavioural change, such as distribution of face masks accompanied by their mandatory use, first in the most hit regions and later nationwide. During the month of May 2020, the country reopened many activities without compromising the decay in the number of newly infected individuals. Analyzing the Italian data carefully may therefore provide important insights into the epidemiology of COVID-19, and in particular to investigate if, how and which limitations in activities and other actions affected the disease dynamics in Italy.

Effective measures slow diffusion of the disease. By identifying such change points in COVID-19 spreading [1], it is therefore possible to associate interventions that were able to modify the course of the epidemic without assuming any effect a priori. Further, such an approach may reveal whether e.g. reopening of society lead to changes in disease dynamics, or could hint at change points apparently unrelated to regulations that deserve further investigations. To find such change points, it is advantageous to use relatively simple mathematical models of infectious diseases, which compared to more complex models, can be fitted to data with a minimum number of assumptions on model parameters [1–3]. 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 COVID-19 cases, a SIQR (susceptible – infectious – quarantined – recovered) model [4] 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 .

By fitting a modified SIQR model to official Italian data, we find change points that correspond well with general lockdowns. Our results indicate further that the mild restrictions imposed in Italy during the first two weeks of the outbreak had negligible effects on the disease dynamics. Surprisingly, we found an acceleration in the decay from the peak, which does not correspond to obvious drastic legal interventions, but may be explained by widespread promotion and mandatory use in face mask use. We confirm these interpretations at regional levels, and find that the gradual reopening of society since early May appears not to have caused any change in the disease dynamics. To the best of our knowledge, this is the first demonstration of a correlation between widespread face mask use and a reduction in COVID-19 transmission dynamics.

Methods

We use a SIQR model [4] to describe COVID-19 dynamics in Italy. Since we will be fitting the cumulative number of official cases, we unite the Q and R compartments in a “cases” compartment $C = Q + R$. The model equations are

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

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

$$\frac{dC}{dt} = \eta I, \quad (3)$$

where β is the rate of infection, η models the average rate with which infectious individuals become tested and quarantined, and eventually appear in the official statistics, and α is the rate with which unidentified infectious individuals recover or die from the disease. We do not explicitly model the number of recovered or deceased non-identified COVID-19 patients, but only the rate α with which these patients 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. Note that $S + I + C \neq N$. To obtain such a conservation law, an additional compartment R_I that model recovered but non-identified patients could be added with dynamics $dR_I/dt = \alpha I$.

Since there is evidence that COVID-19 can be transmitted in the absence of symptoms [5, 6], we do not include an explicit exposed-but-noninfectious (E) state, i.e., we do not consider a SEIQR model [7]. Further, from the Diamond Princess cruiseship and from the Italian village Vo' Euganeo, it has been found that $\sim 50\%$ of COVID-19-positive individuals do not develop symptoms [8, 9]. 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.

In spite of the fact that almost 300,000 individuals have been found COVID-19 positive, and possibly a few millions of unidentified cases have occurred, the total number of persons that have had the infection constitute a relatively small fraction of the Italian population of ~ 60 million. Thus, it is a reasonable approximation that most of the population is still susceptible, $S \approx N$, and hence, as well known, the number of infected individuals follows at any time exponential growth or decay, unless parameters change.

To identify change points, we allow β to be a piecewise constant function of time, thus modelling how contain measures may affect the rate of COVID-19 transmission. We estimate both the time points (change points, T_1, T_2, T_3) where β changes and the values of $\beta = \beta_i$ in the intervals $(T_{i-1}, T_i]$ with $T_0 = 0$ (Feb. 21, 2020) and $T_4 = 122$ (June 22, 2020; last data point). The assumption of piecewise constant β is equivalent to $I(t)$ being a piecewise exponential function (under the assumption $S \approx N$).

As mentioned, the cumulative number of identified COVID-19 cases corresponds to C . With the above assumptions, we obtain from (2) with $S = N$ and by integrating (3),

$$C(t) = C_0 + \begin{cases} \frac{\eta I_0}{\rho_1} (e^{\rho_1 t} - 1), & t \leq T_1, \\ \frac{\eta I_0}{\rho_1} (e^{\rho_1 T_1} - 1) + \frac{\eta I_0 e^{\rho_1 T_1}}{\rho_2} (e^{\rho_2 (t-T_1)} - 1), & T_1 < t \leq T_2, \\ \frac{\eta I_0}{\rho_1} (e^{\rho_1 T_1} - 1) + \frac{\eta I_0 e^{\rho_1 T_1}}{\rho_2} (e^{\rho_2 (T_2-T_1)} - 1) \\ + \frac{\eta I_0 e^{\rho_1 T_1} e^{\rho_2 (T_2-T_1)}}{\rho_3} (e^{\rho_3 (t-T_2)} - 1), & T_2 < t \leq T_3, \\ \frac{\eta I_0}{\rho_1} (e^{\rho_1 T_1} - 1) + \frac{\eta I_0 e^{\rho_1 T_1}}{\rho_2} (e^{\rho_2 (T_2-T_1)} - 1) \\ + \frac{\eta I_0 e^{\rho_1 T_1} e^{\rho_2 (T_2-T_1)}}{\rho_3} (e^{\rho_3 (T_3-T_2)} - 1) \\ + \frac{\eta I_0 e^{\rho_1 T_1} e^{\rho_2 (T_2-T_1)} e^{\rho_3 (T_3-T_2)}}{\rho_4} (e^{\rho_4 (t-T_3)} - 1), & t > T_3, \end{cases} \quad (4)$$

where $\rho_i = \beta_i - (\alpha + \eta)$, $i = 1, 2, 3$, and C_0 is the initial value of identified cases. We fit this expression to the Italian COVID-19 data from February 22, 2020 through June 18, 2020 (Fig. 1), using the `nls` function in R [10].

We note that the values of η and α are irrelevant for the fitting procedure and the main findings in this manuscript regarding the identification of change points, but they permit us to estimate e.g. the (time varying) basic reproduction number $R_0 = \beta_i / (\alpha + \eta)$ by calculating $\beta_i = \rho_i + \alpha + \eta$. To estimate α and η we use previous findings. The probability that a COVID-19 positive individual is tested and quarantined is $\delta = \eta / (\eta + \alpha)$. It has been estimated that the average incubation time is ~ 5 days [11, 12] and the duration of the milder cases of disease is 5-10 days [5]. Identified cases are mostly symptomatic patients, which we assume are tested and isolated a few days after the incubation time is over and first symptoms appear, i.e., after ~ 10 days. We also assume an average time of duration from infection to recovery or death of non-isolated cases of 10 days, i.e., on average infectious individuals are removed from compartment I with rate 0.1/day, either because they recover (milder cases) or become

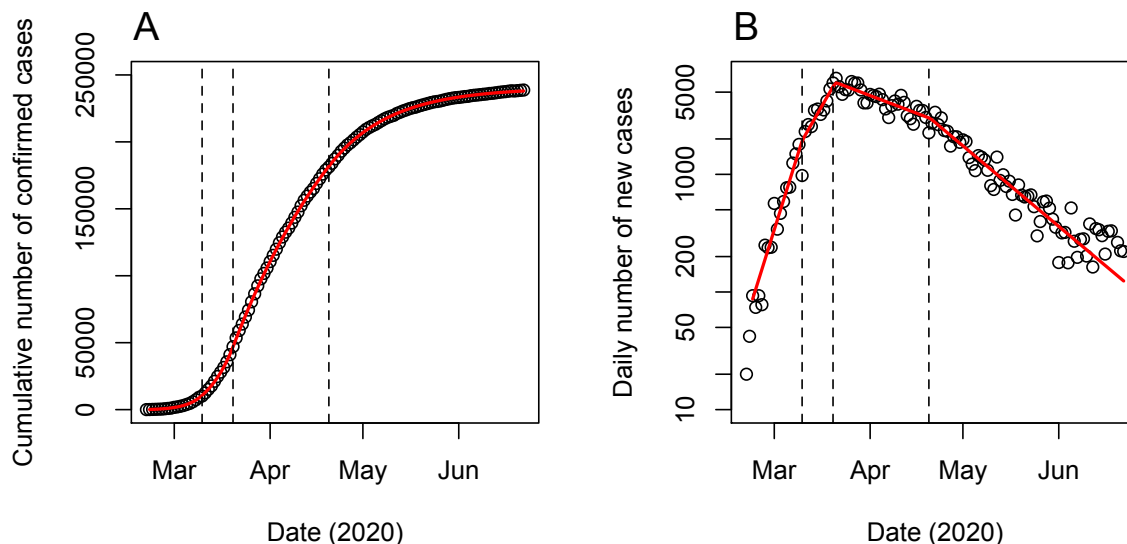


Fig 1. Model fit to national Italian data. *A*: The cumulative number of confirmed COVID-19 cases (circles) and the best fit (curve) of expression (4) with $\eta I_0 = 66.3/\text{day}$ (SE 11.8/day), $T_1 = 17.8$ days (March 7, 2020; SE 1.5 days), $T_2 = 28.4$ days (March 20, 2020; SE 0.3 days), $\rho_1 = 0.192/\text{day}$ (SE 0.015/day), $\rho_2 = 0.104/\text{day}$ (SE 0.008/day), and $\rho_3 = -0.023/\text{day}$ (SE 0.001/day). The vertical dashed lines indicate the change points T_1, T_2 and T_3 . *B*: The daily number of new confirmed cases on logarithmic scale (circles) and the derivative of the fitted curve in panel A (red lines), i.e., $dC/dt = \eta I(t)$. Data from https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series.

tested and enter compartment C . Hence, $\alpha + \eta = 0.1$. Since $\sim 50\text{-}75\%$ of the population is 93
asymptomatic [8, 9], but some milder cases may also go unnoticed and not end in isolation, we 94
assume that $\delta = 1/3$ of infectious individuals are eventually tested. We thus set $\alpha = 0.067/\text{day}$ 95
and $\eta = 0.033/\text{day}$. 96

Results 97

We obtained an excellent fit to the Italian data of cumulative COVID-19 cases assuming three 98
change-points (Fig. 1A). Our results indicate that the early Italian containment measures 99
introduced in late February (school closings, hygiene indications, etc.) had negligible, if any, 100
effect on the disease dynamics. The first change points T_1 was estimated to fall on March 10, 101
2020, corresponding very well to the lockdown of the Northern regions on March 8, 2020, which 102
was followed by complete lockdown of Italy within a few days. These interventions lowered the 103
growth rate ρ by approximately one third. However, only later around March 20, 2020, did the 104
growth rate become negative, and the number of infected individuals started to decline. This 105
corresponds reasonably with the more stringent control measures and closure of all non-essential 106
work places introduces around this date. Indeed March 20, 2020, was the last work day before 107

the wider lockdown of all non-essential companies. Another reasonable interpretation is that the two change points reflect a gradual change in dynamics caused by a distributed, delayed effect of the major lockdown of March 8, as discussed further in the Discussion. A third change point was found at April 20, 2020 where the decay was further accelerated.

From the plot of the cumulative number of cases it may be hard to see the abrupt changes in dynamics. In particular, the assumption of piecewise constant ρ (and β) – and hence exponential growth or decay of the number of infectious individuals in each subinterval defined by the change points – is difficult to verify graphically from this figure. Therefore, we plotted the daily number of new cases on logarithmic scale (Fig. 1B) with the derivative of the fit from Fig. 1A. Indeed, straight lines, corresponding to exponential growth or decay in $I(t) = (dC/dt)/\eta$, are easily identified, justifying the assumption of piecewise constant parameters.

The latter change point identified at April 20, 2020, is apparently not related to any specific containment measures. We speculated that the introduction of mandatory face mask use also outdoors in Lombardy (from April 4, 2020) and Veneto (from April 13, 2020) might be the cause of the acceleration of the decline occurring from this date. In addition, other hard-hit regions such as Tuscany, Emilia-Romagna, Piedmont and Liguria had distributed free face masks during April, encouraged their use, and some municipalities had made them mandatory in public spaces. This promotion of face masks anticipated mandatory use from April 20, 2020 in Tuscany and from May 4, 2020 nationwide. We thus proceeded to fit the data from these six regions with the largest number of official COVID-19 cases. For Lombardy and Veneto, we were able to identify three change points as for the agglomerated Italian data, whereas for the other four regions only two change points were identified, likely due to the relatively low number of cases in these regions during the first week of March.

We found that all regions had a change point in late March (Lombardy and Veneto had two, similar to what we found for the national data) and another in late April. The change in dynamics at this latter change point was very mild in Lombardy, but in other regions it was very strong and clearly seen in the raw data (Fig. 2). Lombardy introduced mandatory face mask use on April 4, 2020, very close to the epidemic peak where the regional health care system was close to a collapse. The data may therefore be unreliable near the peak, masking the effect of face mask use, or the effect of face masks may be hidden behind the other regional strict containment measures, lockdown and high level of alert in the region. For Veneto, the introduction of the lockdowns in March, which correspond well to the first two identified change points, stabilized the number of daily new cases at ~ 400 with a small decline until the change point found at April 20. The number of new cases then shown a marked exponential decline. The identified change point corresponds well to the mandatory use of face masks from April 13 in Veneto.

For Tuscany and Piedmont, we found a similar pattern with a near-constant plateau during late March – early April followed by exponential decline. The corresponding latter change points were, respectively, April 14 and April 19. For Tuscany, the date corresponds well with regional order of April 6, 2020, regarding the distribution of face masks in Tuscany from April 7, anticipating mandatory use in single municipalities once the distribution had completed, and from April 20 in the entire Region. Piedmont did not require face masks until May 4, 2020, but the regional government announced April 15 that masks would become mandatory and started their distribution soon after. Similar patterns and explanations hold for Emilia-Romagna (change point April 28) and Liguria (April 26).

Our chosen values for α and η yield an initial infection rate $\beta_1 = 0.294/\text{day}$ and consequently

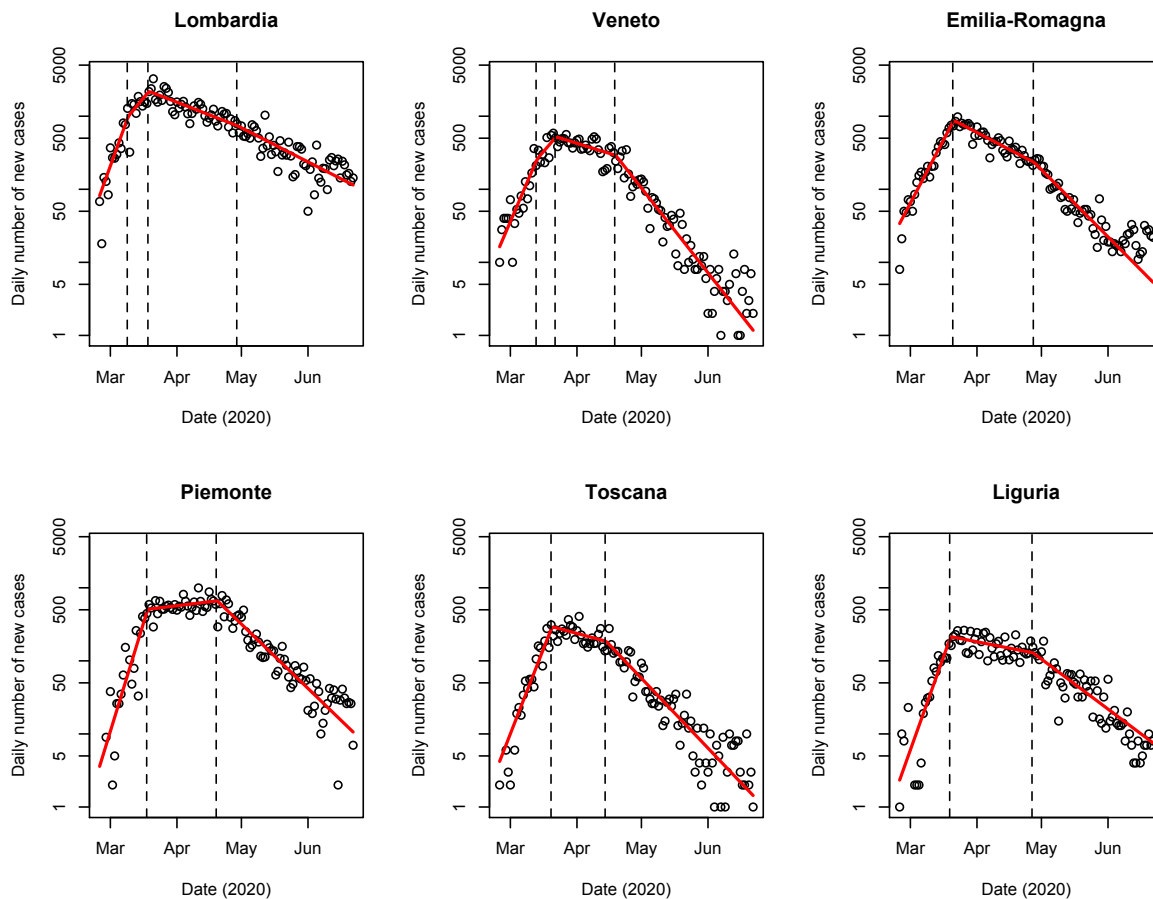


Fig 2. The daily number of new confirmed cases in the six Italian regions with most COVID-19 cases, corresponding to Fig. 1B. For Emilia-Romagna, Piedmont, Tuscany and Liguria only two change points were identifiable. Data from https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data/csse_covid_19_time_series.

the basic reproduction number is estimated to be $R_0 = \beta/(\alpha + \eta) = 2.94$, in line with previous estimates of R_0 falling between 2 and 4 [11, 13–16]. Further, we obtain an estimate of the number of infectious individual at the moment of the outbreak of $I_0 = 65.0/0.033 \approx 2000$, although with a large uncertainty because of the large standard error on the estimate of ηI_0 , and because the calculated I_0 obviously depends on the value of η . Based on the dynamics before the outbreak ($\delta = 0$), we can estimate that the first infectious case appeared in Italy $\log(I_0)/(\beta_1 - 0.1) \approx 40$ days before the outbreak, i.e., around January 12, 2020.

Discussion and Conclusion

Our data fitting procedure identified three change points where the dynamics of the COVID-19 epidemic changed. The first two correspond with the major lockdowns imposed in Italy during March 2020. We therefore conclude that these were effective in halting the spreading of the disease, which lead to the peak in new cases seen in late March. More surprisingly and in

contrast to our expectations, we found a change point around April 20, which at first glance did not correspond to any specific interventions. The major new containment measure introduced in April were the orders of mandatory face mask use in Lombardy, Veneto and Tuscany, and the distribution of free face masks in many regions. Analyzing regional data, we were able to distinguish change points for each region that correspond well to the introduction of face mask distributions or orders in the individual regions. Face masks have been suggested to be important mainly for reducing COVID-19 transmission from asymptomatic individuals [17] that may be responsible for the major number of newly infected cases due to the large fraction of hidden-to-quarantined SARS-CoV-2 positive individuals (~10-fold difference) [17–19].

We considered alternative explanations for the acceleration of the decline of new cases seen in late April. The number of COVID-19 tests did not decline in correspondence to the identified change point, and the fraction of positive tests showed an acceleration in the decline similar to the number of new cases. Mobility data (<https://www.google.com/covid19/mobility/>) showed, if anything, increased activity in late compared to early April, excluding that reduced activity underlie the change. The weather was mild and dry in April in Italy (<https://www.3bmeteo.com>), with little variation in temperature during the month. Virtually no rain fell during the entire month until April 27. Thus, no abrupt change in weather was seen around mid-April, which might have been the cause of the change in dynamics. In summary, no obvious alternative explanations for the acceleration of the end of the epidemic were found.

Our model assumes that the rate of infection changes instantaneously, which is not completely realistic, but simplifies the model. We verified (not shown) that assuming that the change in growth rate occurred in a linear fashion over 2, 5 or 10 days [1] did not change the conclusions. The more gradual change in dynamics smoothed the curve, which made it harder to distinguish the two first change points. The data was well fitted with a first change point shortly after the lockdown on March 8 and a second one around April 15.

We estimated that there were ~2000 of infectious but undetected 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 COVID-19, and the first infected person died from COVID-19 on the following day, wide testing and isolation efforts started. By backward interpolation, we estimated that the first infectious individual (“patient zero”) appear in Italy around January 12, 2020. This estimate suggests that COVID-19 was present in Italy even earlier than a presumed patient zero, suggested to be a German citizen linked to a cluster in Munich [6] visiting Italy around January 25, 2020 [20].

In conclusion, identifying change points confirmed that strict lockdown measures were infective in slowing the spread of the epidemic, leading to its peak in late March. Face mask use appeared as a plausible explanation for the further acceleration of the decline in the number of new cases in April. The reopening of society in May did not lead to change in the decay rate (Fig. 1), also when analyzing the individual regions (Fig. 2), which may be due to the mandatory use of face masks nationwide from May 4. Our results thus lend further support to the importance of face mask use for controlling COVID-19 during the reopening of societies [17, 21, 22].

Acknowledgments

206

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

207

208

209

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.

210

211

212

213

214

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

215

References

1. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*. 2020;.
2. Hethcote HW. The mathematics of infectious diseases. *SIAM review*. 2000;42(4):599–653.
3. Eubank S, Eckstrand I, Lewis B, Venkatramanan S, Marathe M, Barrett CL. Commentary on Ferguson, et al., "Impact of Non-pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand". *Bull Math Biol*. 2020;82(4):52. doi:10.1007/s11538-020-00726-x.
4. Hethcote H, Zhien M, Shengbing L. Effects of quarantine in six endemic models for infectious diseases. *Mathematical biosciences*. 2002;180(1-2):141–160.
5. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. *Jama*. 2020;.
6. 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.
7. 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.
8. National Institute of Infectious Diseases J. type;. Available from: <https://www.niid.go.jp/niid/en/2019-ncov-e/9417-covid-dp-fe-02.html>.
9. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio C, et al. Suppression of COVID-19 outbreak in the municipality of Vo, Italy. *medRxiv*, doi:10.1101/2020041720053157v1. 2020;.
10. R Core Team. R: A Language and Environment for Statistical Computing; 2016. Available from: <https://www.R-project.org/>.

11. 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.
12. 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.
13. 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;13(1):3–7. doi:10.1111/jebm.12376.
14. 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.
15. 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;.
16. 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;.
17. Prather KA, Wang CC, Schooley RT. Reducing transmission of SARS-CoV-2. *Science*. 2020;.
18. 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.
19. Pedersen MG, Meneghini M. A simple method to quantify country-specific effects of COVID-19 containment measures. *medRxiv*. 2020;.
20. Kupferschmidt K. Genome analyses help track coronavirus' moves. *Science*. 2020;367:1176–1177.
21. Chu DK, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, et al. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. *The Lancet*. 2020;.
22. Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, Kuang Y, et al. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infectious Disease Modelling*. 2020;.