

Estimating the number of infected by COVID-19 in Italy

Cintra, P. H. P.^{1,*} and Fontinele, F. N.²

¹Institute of Physics, University of Brasilia, Brazil

²Institute of Physics, University of Brasilia, Brazil

*pedrohpc96@hotmail.com

ABSTRACT

Italy suffered heavily with the new pandemic crisis caused by the novel coronavirus SARS-CoV-2. Given the low number of tests performed on the early stages of the outbreak, Italy lost track of most of infections. We use a modified SEIR model to reconstruct the most realistic infection curve using the hospitalization curve of the registered data. Using this method we estimated that, by the end of the first infection wave, about 3-4% of the population will have been infected by the virus. Following the same process, the number of deaths is projected to be between 100000 to 115000. We also find a significant correlation between the number of tests performed, the fraction of undocumented infections and the rate of change dI/dt of the real infection curve. We conclude that herd immunity is not enough to contain further spread of the disease inside the country.

Keywords - SEIR, Italy, infection curve, estimative

Introduction

Since the outbreak of the novel coronavirus SARS-CoV-2, on Wuhan, the world has suffered the biggest pandemic crisis of the century. The virus, which causes the disease named COVID-19 (Coronavirus Disease - 2019), presents an \mathcal{R}_0 between 2-3, meaning that each infected person causes from 2 to 3 secondary infections^{1,2}. The novel coronavirus, presents a lot of similarities with the SARS-CoV virus, which caused the SARS epidemic crisis back on 2003³, belonging to the *coronaviridae* family, the *orthocoronavirinae* subfamily and the *betacoronavirus* genus, having mutated from bats to human⁴. The first cases of the novel virus date back to December 2019 at the Food Market of Wuhan, China⁵, where bats are sold among other exotic animals, since then the virus has been spreading throughout China, later Asia, Europe, Africa and America, causing a global scale economic crisis and being notified by the World Health Organization as a pandemic on March 11th, with human-to-human sustained transmission confirmed⁶.

As an individual is exposed to the virus, the incubation period begins, with no symptoms and a small chance of contaminating others. When the symptoms onset, the infected individual show symptoms in a varied range of intensities and may develop severe acute respiratory syndrome. COVID-19 has a general case fatality rate (CFR) bellow 5%⁷, with an average of 2.3%. The behavior of the disease is age dependent, with the higher risk group being older populations, that present a CFR rate of 8% for individuals between 70-79 years and 14.8% for people older than 80 years⁸. However, even with a low CFR, the number of case hospitalizations rate (CHR) is quite high, with 5% of the cases being critical and 14% being severe⁹, presenting a challenge to health care systems of some countries.

Italy suffered heavily with the pandemic, having registered the first case at January 31th with two Chinese tourists on Rome, and having declared a lock-down at March 11th¹⁰ in order to prevent further spread of the virus. With a low testing rate on the early stages of the local outbreak, Italy lost track of the infections and experienced a collapse on the health system, without hospital beds to attend the population needing it. The region of Lombardy quickly became the local epicenter of the pandemic, accounting for more than one third of the total cases registered. With a case fatality rate (CFR) around 14%, much higher than the world average and the infection fatality rate (IFR) of 0.7%¹¹, Italy suffers from a huge undernotification of infections.

Mathematical models predicted the potential for an international outbreak early on¹² and described how Wuhan became the center of an epidemic crisis on China. The outbreak quickly spread throughout mainland China and other countries. Although the pandemic crisis began on Wuhan, the United States of America is nowadays the epicenter of the pandemic with the biggest number of reported cases of COVID-19.

With such a critical situation in the world, it is crucial to determine the shape and the peak of the infected population during the outbreak, several models have already been developed, one particular model that had a great impact on intervention politics

during the pandemic was developed by the Imperial College London¹³. We then, use a modified SEIR model to reconstruct the more realistic infection curve on Italy, finding the date of the infection peak, the real number of infections and deaths. Aiming to investigate the possibility of herd immunity in Italy, knowing the percentage of the population already infected is crucial for decision making by local authorities on the subject.

Model

As a novel strain of the *coronaviridae* family and the *betacoronavirus* genus, it is not expected that any individual possesses antibodies against it, which causes the entire population to be susceptible to infection.

We make use of a modified SEIR mathematical, in which we consider the population N of a given region as divided in 5 groups. At time t , there are those who are susceptible to get infected $S(t)$, the ones who have already been exposed to the virus but does not present symptoms yet $E(t)$, people who are already infected and present the symptoms $I(t)$, the ones that have already recovered from the disease $R(t)$ and those who are dead due to the infection $D(t)$. We also take into account the hospitalizations due to COVID-19 by proposing the hospitalized population $H(t)$, suggesting the name SEIHRD for the model. This model is a good approximation to a short epidemic, so the population of a region is roughly constant throughout the epidemic period. Also, since this is a deterministic model, we assume N to be a big number compared to the number of people associated with the infection of a single person. The final consideration is that we also assume that people that are recovered from the disease acquire immunity and does not become susceptible to become infected again for some years. The last is justified by the behavior of immunity on the SARS epidemic crisis of 2003, where infected patients acquire antibodies against the SARS-CoV virus for a period of 2-3 years¹⁴. The SARS-CoV presents large number of similarities with the new SARS-CoV-2 virus, belonging even to the same genus *betacoronavirus*. We also have evidence that 44% of secondary infections happens due to contact with an individual on presymptomatic stage (i. e. during the incubation period)¹⁵.

The rate of infection λ is proportional to the number of people infected and the number of exposed, $\lambda(t) = (1 - P_{exp})\beta I(t) + P_{exp}\beta E(t)$, where the constant β represents the effectiveness of the infection and P_{exp} is the percentage of total infections due to contact with exposed individuals and not infected ones. The rate of cure $\gamma = (1 - P_{IFR})\tau_r^{-1}$, where P_{IFR} is the infection fatality rate and τ_r is the average time taken for an infected person to recover. Similarly the rate of death is $\mu = P_{IFR}\tau_d^{-1}$, where τ_d is the average time taken for an infected person to die. Once the symptoms start, there is a probability of being hospitalized given by the infection hospitalization rate (IHC) P_h , inside the hospital, the patient now possesses a different value of γ and μ given as γ_h and μ_h . Figure 1 carries a visual representation of the SEIHRD model.

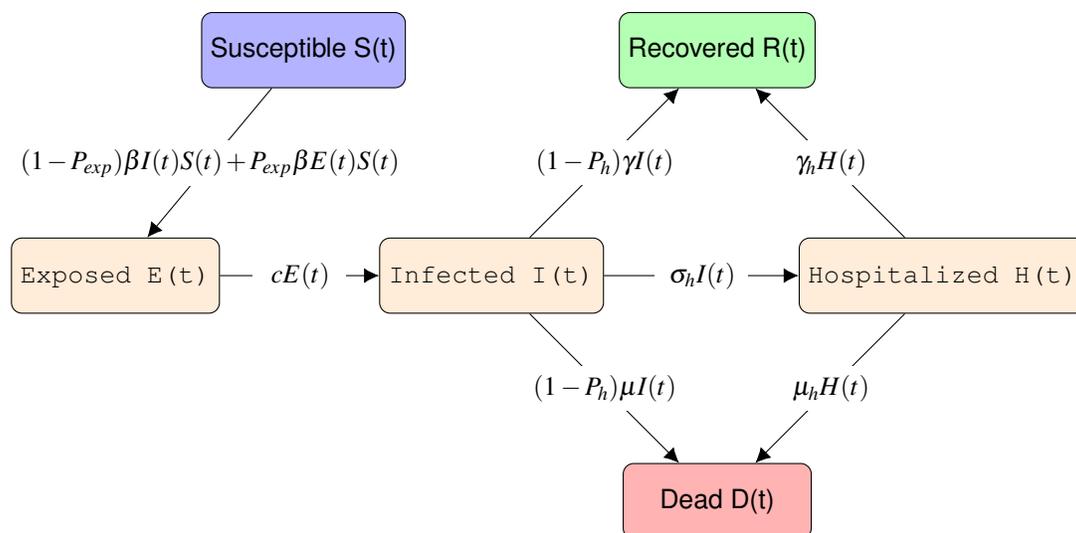


Figure 1. Representation of a SEIRD model, a susceptible person gets exposed to the virus, being infected afterwards and either dies or recovers from the disease.

The differential equations representing the evolution of the populations are given by

$$\frac{dS}{dt} = -\frac{(1-P_{exp})\beta}{N}I(t)S(t) - \frac{P_{exp}\beta}{N}E(t)S(t), \quad (1)$$

$$\frac{dE}{dt} = \frac{(1-P_{exp})\beta}{N}I(t)S(t) + \frac{P_{exp}\beta}{N}E(t)S(t) - cE(t), \quad (2)$$

$$\frac{dI}{dt} = cE(t) - (1-P_h)\gamma I(t) - (1-P_h)\mu I(t) - P_h I(t), \quad (3)$$

$$\frac{dH}{dt} = P_h I(t) - (\gamma_h + \mu_h)H(t), \quad (4)$$

$$\frac{dR}{dt} = \gamma I(t) + \gamma_h H(t), \quad (5)$$

$$\frac{dD}{dt} = \mu I(t) + \mu_h H(t), \quad (6)$$

where the rates γ , γ_h , μ and μ_h are expressed in terms of the infection fatality rate (IFR), the infection hospitalization rate (IHR), the average time from symptoms onset to hospitalization, from symptoms onset to recovery and death and from hospitalization to recovery and death:

$$\sigma_h = \frac{P_h}{\tau_h}, \quad (7)$$

$$\gamma = \frac{1 - P_{IFR}}{\tau_r}, \quad (8)$$

$$\mu = \frac{P_{IFR}}{\tau_d}, \quad (9)$$

$$\gamma_h = \frac{1 - P_{dh}}{\tau_{rh}}, \quad (10)$$

$$\mu_h = \frac{P_{dh}}{\tau_{dh}}. \quad (11)$$

Equations (1) to (6) are solvable numerically given the initial conditions $S(0)$, $E(0)$, $I(0)$, $H(0)$, $R(0)$, $D(0)$. We assume $R(0) = D(0) = H(0) = 0$ at the beginning of the outbreak and $S(0) \approx N$ since $N \gg I(0)$ or $E(0)$. Therefore, N here plays the role of an effective population which has become susceptible to the spread of the virus during this first wave. Since the population is not homogeneously distributed, N cannot be understood as the total population; therefore we choose to leave it as a free parameter for the data adjustment. N also plays a vital role on finding the population of infected after the first wave of the outbreak. Other than N , $I(0)$ and $E(0)$ were left free for the adjustment as well.

The lockdown acts on the β parameters, decreasing it to a new value β_f which is a small fraction of the initial value β_i . We propose a logistic function to model the decrease of β with time upon the declaration of lockdown.

$$\beta(t) = \frac{R\beta_i}{1 + \tau e^{(t-t_c)}} + (1-R)\beta_i, \quad (12)$$

where t_c is the day the lockdown is imposed, τ is a constant for adjusting the start of the lockdown to t_c and R is the reduction of β due to lockdown effect.

We did not consider the proportion of asymptomatic infections, on the argument that it would add many more parameters to the adjustment which are not measured and could increase the probability of finding a local minimum by the fitting algorithm, making the success of the model highly dependent on the initial conditions. We do not know yet the mechanism governing asymptomatic infections and several studies find different proportions of asymptomatic carriers; on the diamond princess ship about 18% of infections were asymptomatic¹⁶, on an Italian village, about 50 to 75% of infections were asymptomatic¹⁷, on the previous MERS-CoV outbreak, provoked by a coronavirus belonging to the same *betacoronavirus* genus of the SARS-CoV-2 virus, 12.5% of infections were estimated to be asymptomatic¹⁸. On another study with citizens of an airplane flight, 11.2% of them were considered asymptomatic¹⁹. Even though most studies conclude an asymptomatic fraction between 10-20%, we do not know clearly how the parameters $\tau_{r,d}$ behave for these populations or the percentage of infections they are responsible for. For that reason, the curve of infection obtained here contains both symptomatic and asymptomatic cases.

The values for the parameters used for the adjustment and construction of the infection curve are shown on Table 1.

Methods

For the adjustment of data, we choose each parameter related to the spread of the disease according to various clinical studies and international averages shown on Table 1. On total, four parameters were left free for the fitting, N , $I(0)$, $E(0)$ and β .

Parameter	Value	Reference
c	5.1 days ⁻¹	20 ¹⁵ 21 ²² 23
τ_r	18 days	24
τ_d	10 days	25
P_{exp}	0.44	15
β	Fitting	--
P_h	0.036	11
P_{IFR}	0.007	11
P_{dh}	0.23	26
τ_{rh}	13.5	26
τ_{dh}	5 days	25
τ_h	5 days	25
N	Fitting	--
R	0.77	11

Table 1. The parameters marked as *Fitting* are the ones left for the adjustment of with the data.

When performing the fitting of data, in order to find the best N values that would represent the situation in the country, we varied N from 0 to 30% of the total population on steps of 0.1%. At each step, we fix N and fit the simulation with the data. The best values of N were chosen as being the ones that maximize the χ^2 value of the fit and minimize the error on the parameters fitted. To take into account the undernotification of hospitalizations as well, we varied N to higher values than the optimal one and used them to estimate a margin of error for N .

Results

After the adjustment, the values of β , $I(0)$, $E(0)$ and N were used to reconstruct the real infection curve with a full simulation of the SEIHRD model using the parameters from 1 and 2

N	χ^2	β	$I(0)$	$E(0)$
3-4%	0.991 - 0.995	0.538-0.680	34-908	16-347

Table 2. Fitting parameters for Italy

The infection curve generated with those parameters was plotted alongside the data reported by Italy as a comparative tool.

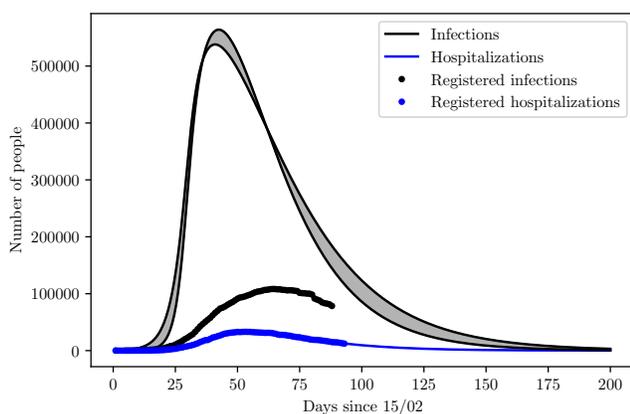


Figure 2. More realistic curve of infections in Italy.

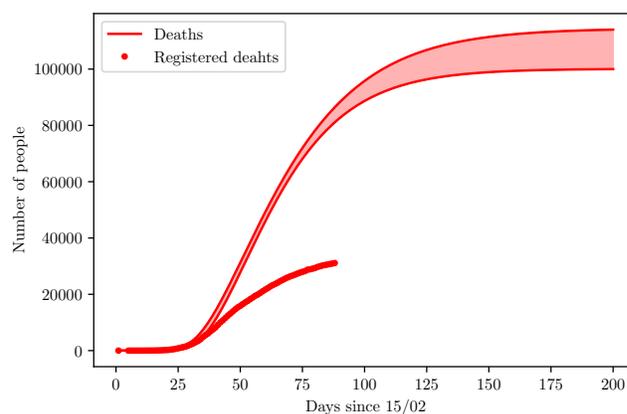
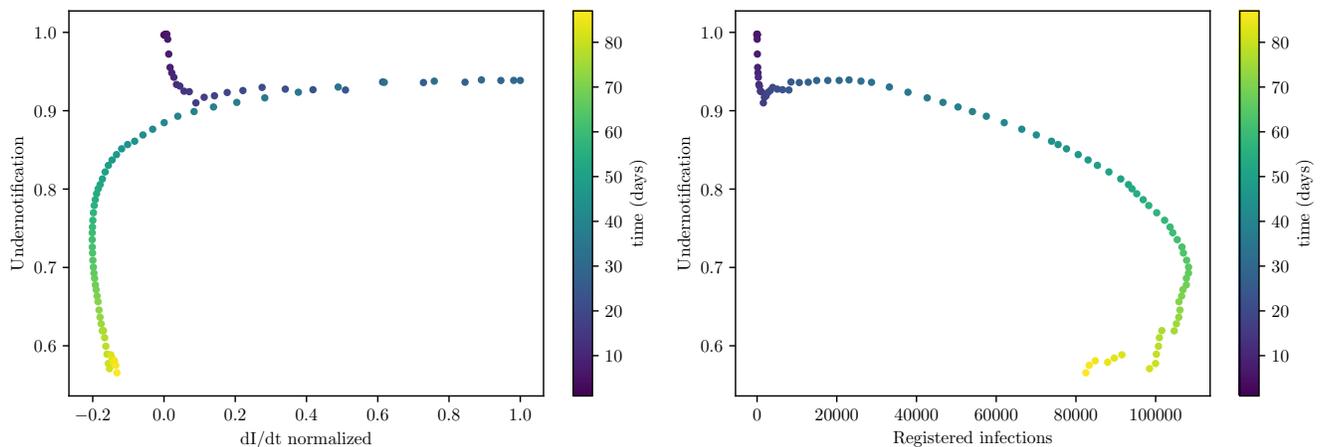


Figure 3. More realistic curve of deaths in Italy.

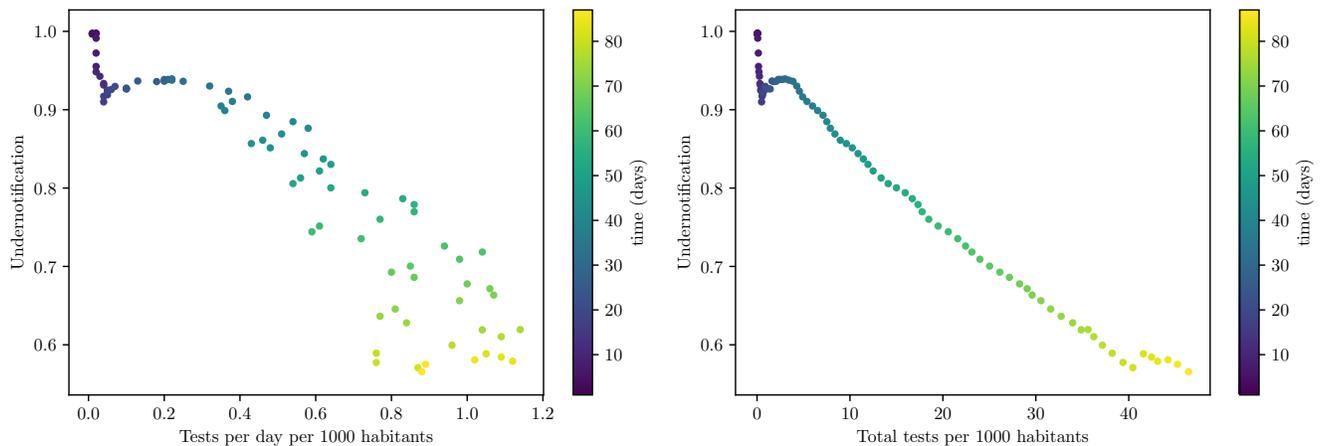
From the reconstruction, we estimate a total of 1.6 to 2 millions of infections by the end of this first pandemic wave, representing 2.5 to 3.2% of the population. We did also observe the date of the more realistic peak of the outbreak, about 25 to 22 days before the actual registration (i. e. between March 28th and March 31th). That is perfectly reasonable by considering the increase of tests performed daily by Italy. Although the infection peak passed, there were many more infections than the registered value, so the increase in testing notified more cases, giving the impression that the peak was later. The peak of infections is a bit higher than 500000 infected, more than the cumulative number registered until now. The number of deaths estimated after this first wave is from 100000 to 115000.

After acquiring the more realistic curve of infection, we estimated how the undernotification varied in Italy as days passed. We also did investigate the correlation of the undernotification with the rate of change of the infection curve and the number of tests performed daily by Italy.



(a) Change in dI/dt as well as the undernotification as time progresses.

(b) Change on the notified number of infections and undernotification as time days passed.



(c) Relation between the undernotification and the number of tests performed per day per 1000 habitants.

(d) Relation between the undernotification and the total number of tests performed per 1000 habitants.

Figure 4. Relations between the number of infections, the rate of change of the infection curve and the testing data. 4a presents a correlation of 0.58, and a hypothesis test eliminated the null hypothesis that the correlation between the data set is not significant ($p < 0.05$). 4b represents the variation of undocumented cases with respect to the number of registered infections, the correlation between data is -0.85 and the F-Test rejected the null hypothesis ($p < 0.05$) indicating a significant correlation. 4c shows the change in undernotification with respect to the number of tests performed per day per 1000 habitants, presenting a correlation of -0.92 ($p < 0.05$), also rejecting the null hypothesis. The final image 4d presents a correlation of -0.99, however the F-Test did not reject the null hypothesis.

From the hypothesis test and the correlation between the data, the rate of change dI/dt seems to affect the increase in undernotification, that is, as dI/dt increases, so does the undernotification, however, the lower correlation of these quantities when compared to others suggests that the rate of change dI/dt plays a role not as heavy as the testing data. Similarly, the increase in tests per day decreases the undernotification. The number increase in registered cases also plays a role in decreasing the amount of undocumented infections, however this quantity is highly correlated with the number of tests (0.94), rejecting the null hypothesis by the F-Test. The last quantity was the total number of tests, which is highly correlated to the undernotification, but might not be a significant correlation. The testing rate per day is the main variable when decreasing the fraction of undocumented infections, although we see that naturally the undernotified cases tend to decrease as dI/dt comes to lower values, that is, as the pandemic crisis approaches the end.

Conclusion

The model yielded realistic results and allowed a profound analysis on the relation between undocumented infections and other data. One considerable limitation considerable is the undernotification in hospitalization numbers. We assume here that hospitalized patients are a priority for testing; therefore, the fraction of undocumented hospitalizations should be lower, enabling a more precise reconstruction of the infection curve.

Another limitation is the possible variation of international parameters to the specific case of Italy, for example, IFR might be lower or higher for Italy than for France due to differences in both populations.

Our analysis finds a direct relation between the number of tests performed per day and the tracking of COVID-19 cases. Increasing the testing rate on the early days of the infection is crucial for maintaining control of the outbreak. Lastly, with 3-4% of the population infected, a second infection wave is highly probable if social contact levels increase. Other studies already suggested a great amount of undocumented infections on Italy on the early stages of the outbreak, 72% (61-79) until February 29th²⁷, our method estimates between 74-90%, considering the margin of error of the infection curve. Social distancing policies should hold for as long as it is needed until an efficient vaccine is available, since herd immunity is not a considerable reality after the first wave of infection.

References

1. Zhang, S. *et al.* Estimation of the reproductive number of novel coronavirus (covid-19) and the probable outbreak size on the diamond princess cruise ship: A data-driven analysis. *Int. J. Infect. Dis.* **93**, 201–204 (2020).
2. Zhao, S. *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. *Int. J. Infect. Dis.* **92**, 214–217 (2020).
3. Marra, M. A. *et al.* The genome sequence of the sars-associated coronavirus. *Science* **300**, 1399–1404 (2003).
4. Lu, R. *et al.* Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565–574 (2020).
5. Li, Q. *et al.* Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New Engl. J. Medicine* (2020).
6. Organization, W. H. *et al.* Coronavirus disease 2019 (covid-19): situation report, 67. (2020).
7. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C. & Di Napoli, R. Features, evaluation and treatment coronavirus (covid-19). In *StatPearls [Internet]* (StatPearls Publishing, 2020).
8. Surveillances, V. The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (covid-19)—china, 2020. *China CDC Wkly.* **2**, 113–122 (2020).
9. Wu, Z. & McGoogan, J. M. Characteristics of and important lessons from the coronavirus disease 2019 (covid-19) outbreak in china: summary of a report of 72 314 cases from the chinese center for disease control and prevention. *Jama* (2020).
10. Flaxman, S. *et al.* Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on covid-19 in 11 european countries. (2020).
11. Salje, H. *et al.* Estimating the burden of sars-cov-2 in france. *medRxiv* (2020).
12. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-ncov outbreak originating in wuhan, china: a modelling study. *The Lancet* **395**, 689–697 (2020).
13. Walker, P. G. *et al.* The global impact of covid-19 and strategies for mitigation and suppression. *WHO Collab. Centre for Infect. Dis. Model. MRC Centre for Glob. Infect. Dis. Analysis, Abdul Latif Jameel Inst. for Dis. Emerg. Anal. Imp. Coll. Lond.* (2020).

14. Liu, W. *et al.* Two-year prospective study of the humoral immune response of patients with severe acute respiratory syndrome. *The J. infectious diseases* **193**, 792–795 (2006).
15. He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of covid-19. *Nat. medicine* 1–4 (2020).
16. Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (covid-19) cases on board the diamond princess cruise ship, yokohama, japan, 2020. *Eurosurveillance* **25**, 2000180 (2020).
17. Day, M. Covid-19: identifying and isolating asymptomatic people helped eliminate virus in italian village. *Bmj* **368**, m1165 (2020).
18. Al-Tawfiq, J. A. Asymptomatic coronavirus infection: Mers-cov and sars-cov-2 (covid-19). *Travel. Med Infect Dis* **27**, 1–2 (2020).
19. Nishiura, H. *et al.* Estimation of the asymptomatic ratio of novel coronavirus infections (covid-19). *medRxiv* (2020).
20. Backer, J. A., Klinkenberg, D. & Wallinga, J. Incubation period of 2019 novel coronavirus (2019-ncov) infections among travellers from wuhan, china, 20–28 january 2020. *Eurosurveillance* **25** (2020).
21. Linton, N. M. *et al.* Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J. Clin. Medicine* **9**, 538 (2020).
22. Guan, W.-j. *et al.* Clinical characteristics of coronavirus disease 2019 in china. *New Engl. J. Medicine* (2020).
23. Lauer, S. A. *et al.* The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals internal medicine* (2020).
24. Ruan, Q., Yang, K., Wang, W., Jiang, L. & Song, J. Clinical predictors of mortality due to covid-19 based on an analysis of data of 150 patients from wuhan, china. *Intensive care medicine* 1–3 (2020).
25. Luigi Palmieri, P. B. A. B. S. B. E. B., Xanthi Andrianou. Characteristics of sars-cov-2 patients dying in italy. report based on available data on may 7th , 2020. (2020).
26. Moghadas, S. M. *et al.* Projecting hospital utilization during the covid-19 outbreaks in the united states. *Proc. Natl. Acad. Sci.* **117**, 9122–9126 (2020).
27. Tuite, A. R., Ng, V., Rees, E. & Fisman, D. Estimation of covid-19 outbreak size in italy. *The Lancet infectious diseases* (2020).