

The big unknown: The asymptomatic spread of COVID-19

Roumen Anguelov, Jacek Banasiak, Chelsea Bright,
Jean Lubuma, Rachid Ouifki

Department of Mathematics and Applied Mathematics
University of Pretoria

1 Introduction

In the current COVID-19 pandemic we have a rather overwhelming situation of an enormous amount of data produced worldwide everyday and no reliable way of making sense of it or making any motivated predictions. To a large extent, this situation is created by the fact that data are passively collected, mostly in terms of number of cases, severity, recovered and deaths, as recorded in clinics and hospitals. As this information is publicly available, it is distributed worldwide on various official and private social media platforms, possibly making little contribution to the understanding the disease and its epidemiological characteristics, and preventing any reliable predictions of what lies ahead.

Here we would like to draw attention to the role of asymptomatic or mildly symptomatic infections on the epidemic dynamics. For COVID-19 the severe symptomatic percentage of infections requiring hospitalization increases with age from 0.1% for those under 5 years old to 27.3% for the over 80 age group. *Caeteris paribus*, the number of infectious needing hospitalization depends on the age structure of the population. South Africa's population is relatively young, e.g. 82.9% of the population is under 50. Using the detailed age structure given in [4, page 10] and the percentages of severe symptomatic cases given in [2, Table 1], the average percentage of COVID-19 infectious people needing hospitalization is 4.02%. This could mean, when the infection follows locally/internally determined dynamics, i.e. it is no longer fueled by imported cases, a relatively small fraction of the infectious will be hospital cases and counted as such. Possibly, a larger number will be seen at clinics and outpatient rooms, but it is quite likely that a very large fraction of infections in the population will remain undetected. In [2], it is noted that the data from China and repatriating flights suggests that 40% - 50% of infections were not detected as "cases". This means a large fraction of the population will experience the infection asymptotically or with mild symptoms and will not seek medical attention. Hence, the counting of confirmed cases, if not interpreted appropriately, may present quite a distorted epidemiological assessment.

2 On the relative size of asymptomatic COVID-19 spread

It was suggested in [1] that, via testing of random samples of the population, one can determine the prevalence of the infection and possibly of immunity of the general population not treated by the health system. Such study was carried out in the municipality of Gangelt in Germany, where a random sample of the population was tested for the virus and for antibodies. In [8], an intermediate result is given from the study. Approximately 500 people from a total sample of 1000 were tested. An existing immunity of approximately 14 percent was recorded and 2 individuals were found infected at the time of testing. A total of approximately 15 percent were recorded to have been infected. A mortality rate of 0.37 percent was calculated from the total infections in Gangelt. This contrasts the mortality rate of 1.98 percent in Germany calculated by Johns Hopkins University, 5 times higher than the mortality rate in Gangelt. The mortality rate based on the total population in Gangelt is currently 0.06 percent. The lower mortality rate in Gangelt is indicative of the fact that the study in Gangelt considers all infected people in the sample, including those with asymptomatic and mild symptoms. Hence, the study suggests a lower lethality of the virus than previously thought. That is, testing only/mostly those with medium to severe symptoms may give a distorted view on the spread of the disease and the mortality rate. The authors of [8] also state that it is possible to achieve herd immunity as the virus does not lie dormant in the body after recovery, immunity is estimated to last 6-18 months, and the epidemic dies out when 60-70 percent are immune. It is also suggested that lower initial viral load may result in less severe symptoms and, at the same time, development of immunity.

A study in Iceland, reported in [9, 10], suggests that almost all infections are either asymptomatic or mildly symptomatic, with about 50 percent of positive cases asymptomatic. This is significant, since, as of 11 April 2020, 10 percent of Iceland's population has been tested, the highest percentage in comparison to any other country. Current data show that, as of 10 April 2020, the fatality rate in Iceland is 0.41 percent [7], close to that of Gangelt. In addition, since mid-March, the frequency of the virus among those without co-morbidities or symptoms is either decreasing or stable, suggesting increased prevalence of immunity.

Publications highlighting the significant role of asymptomatic cases in the COVID-19 epidemiology include a recent study, published in the British Medical Journal [5], stating that 130 of 166 new infections (78 percent) in China, identified on 1 April 2020, were asymptomatic. This is further supported by a study in [6], where it is stated that blanket testing in a isolated village in Italy, with a population of approximately 3000, recorded a drop, within 10 days, of 90 percent of people with symptoms due to isolating those who were symptomatic and asymptomatic.

3 Mathematical model and associated observable variables

The main focus of health authorities, government and media is the current burden of the disease represented as the number of so called "confirmed cases", these typically being symptomatic patients seeking medical assistance. This is rightfully so, as the present need of medical assistance as well as reducing immediate future demand are issues requiring urgent attention and action. However, as the review in the previous section suggests, the asymptomatic cases and unreported cases involving mild symptoms constitute a significant, if not dominant part, of infections. Considering the relatively younger population of South Africa, the asymptomatic cases are likely to be a larger fraction of all infections than in Europe or USA. Hence, these are likely to have a strong and decisive impact on the long term epidemiological dynamics of COVID-19 and eventual results in terms of the loss of life and economic burden.

In general, the impact of this, yet invisible, COVID-19 epidemiological component can be expected to result in various changes in the confirmed cases, which might be difficult to explain via the confirmed cases count only. For example, quarantine measures based only on symptomatic cases are not likely to produce the expected results. Further, asymptomatic cases may contribute strongly to building herd immunity, leading to unexpected (but desired) decline in cases.

The proposed model is of the well known SEIR type, the main difference being that the infectious are structured according to the severity of symptoms. This model is also used in [3], but with additional compartments related to intervention. Here our focus is on the epidemiological dynamics of the infection. The flow chart is given in Figure 1.

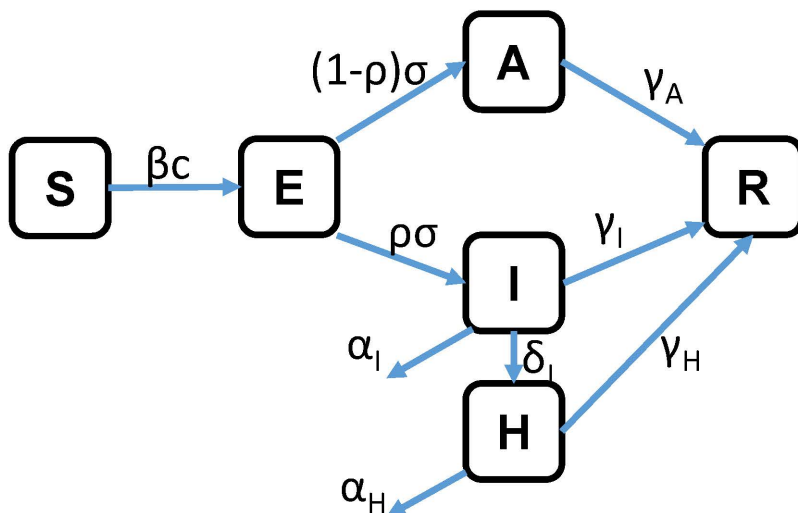


Figure 1: Compartmental flow chart for COVID-19

The susceptibles (S), due to infections with the force determined by the standard incidence with coefficient βc , move to the compartment of the exposed (E). As an

exposed individual becomes infectious (waiting time $\frac{1}{\sigma}$), he/she moves either to compartment A (asymptomatic or mild symptoms), or to the compartment I of those with medium to severe symptoms who are likely to seek medical attention. Some of the latter (transfer rate δ_I) will require hospitalization (compartment H). From compartments I and H the two other exits are to the recovered compartment R (rates γ_I and γ_H , resp.) or death (rates α_I , α_H , resp.). The flow chart is implemented as the system of equations (1)–(6).

$$\frac{dS}{dt} = -\beta c(I + A)S - \lambda(t)S, \quad (1)$$

$$\frac{dE}{dt} = \beta c(I + A)S + \lambda(t)S - \sigma E, \quad (2)$$

$$\frac{dA}{dt} = (1 - \rho)\sigma E - \gamma_A A, \quad (3)$$

$$\frac{dI}{dt} = \rho\sigma E - (\delta_I + \alpha_I + \gamma_I)I, \quad (4)$$

$$\frac{dH}{dt} = \delta_I I - (\alpha_H + \gamma_H)H, \quad (5)$$

$$\frac{dR_A}{dt} = \gamma_A A, \quad \frac{dR_{IH}}{dt} = \gamma_I I + \gamma_H H, \quad \frac{dD}{dt} = \alpha_I I + \alpha_H H. \quad (6)$$

The compartments represent fractions/percentage of the population so that the force of infection given in the right hand side of equation (1) represents standard incidence, where β represents the probability of infection at contact and c is the number of contacts per person. The product βc sometimes is referred to as the number of sufficient contacts per person, where sufficient contact means contact in which transmission of the infection occurs. The recent interventions of government, like lockdowns and sanitary measures, aim to reduce precisely this parameter. Hence, in principle, it may be used for testing such interventions. The parameter λ is a function of time used to account for infections brought in by people who travelled abroad. It is mostly relevant at the beginning of the infection, before the borders are strictly controlled. We take λ to be a smooth and monotone decreasing function of the time t , such that

$$\lambda(t) = \begin{cases} 2.18 \times 10^{-6} & \text{for } t \in [0, 8] \\ 0 & \text{for } t \geq 9 \end{cases}$$

The parameter ρ , which reflects that ratio of split of the output of E between compartments A and I , is the main interest of this paper.

The three equations (6) can be decoupled from the system and, indeed, it is useful to do so in the theoretical analysis or for computing the solution. However, the variables R_A , R_{IH} and D are convenient for representing the total number of cases in each category. More specifically, the total number of symptomatic cases, past and present, is given by $I + H + R_{IH} + D$. Similarly the total number of all cases, past and present, is $A + I + H + R + D$, where $R = R_A + R_{IH}$.

Typical graphs of the outcome from the system (1)–(6) are presented in Figure 2. The values of most parameters are given in Table 1.

The parameter of most critical importance is the coefficient βc . Many of the interventions, like improved hygiene, social distancing and quarantining, can be

γ_A	γ_I	γ_H	δ_I	α_I	α_H	σ	ρ
1/14	0.9/14	0.1	0.09/14	0.005/14	0.2/7	1	0.4

Table 1: Parameter values

modelled via their impact on the value of βc . For the simulation in Figure 2 we use $\beta c = 0.17$. Then, the basic reproduction ratio is

$$\mathcal{R}_0 = \beta c \left(\frac{\rho}{\delta_I + \alpha_I + \gamma_I} + \frac{1 - \rho}{\gamma_A} \right) = 2.3848,$$

and satisfies $2 < \mathcal{R}_0 < 3$ as estimated by many health agencies so far [11, 12].

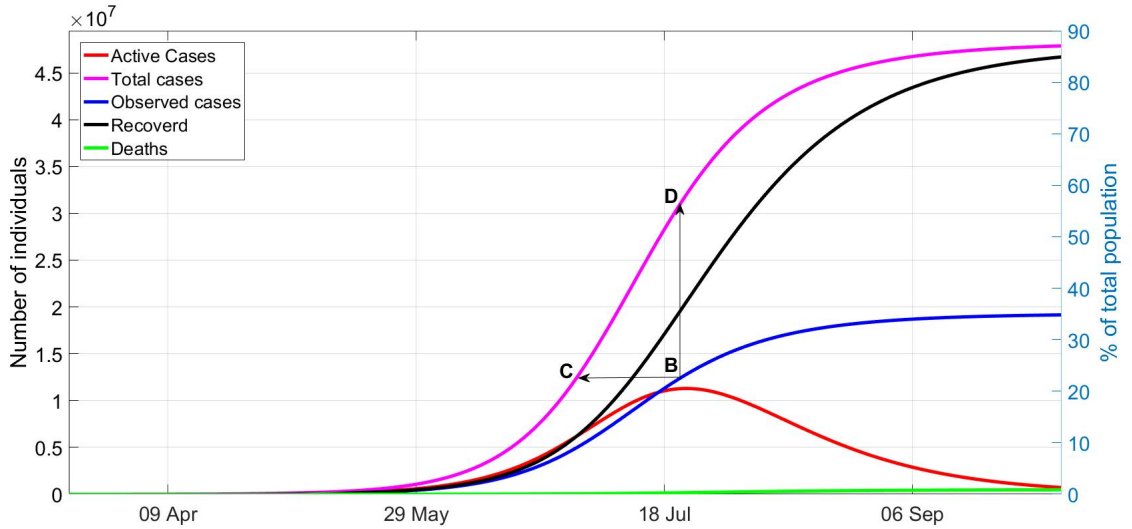


Figure 2: Typical progression of the infection, without interventions.

As common in mathematical modelling, not all variables of the model are observable. In fact, typically only a function (observation operator) of the model variables is observable. In this model, the observable variables are daily recruitment (new cases) into compartment I as well as the current values of I , H , R_{IH} and D . The data on the daily recruitment rate into I tend to oscillate significantly due to many random factors, e.g. due to various reasons a case can be counted a day early or a day late. Hence, it is more appropriate to consider the cumulative distribution of the recruitment rate $\rho\sigma E$ into the I compartment. It is easy to see that

$$\int_0^t \rho\sigma E(\theta) d\theta = I(t) + H(t) + R_{IH}(t) + D(t),$$

that is, this cumulative distribution is exactly the total number of symptomatic cases, past and present. We consider this number to be approximately represented by the total confirmed cases as reported daily. Hence, we consider this sum as observable variable and will adjust the model to fit the data available for it in the next section. This variable is represented by the blue line on Figure 2. Similarly,

the total number of infections, symptomatic and asymptomatic, that is $A(t) + I(t) + H(t) + R(t) + D(t)$, is cumulative distribution of the total infective recruitment σE . It is represented on Figure 2 by the magenta line. Clearly the total number of infections is well above the observed cases. The distance between the two is determined by the parameter ρ . In order to illustrate the significance of the distance between the two curves, we note that when the graph of active infective cases $A + I + H$ (the red line) picks up, the total confirmed cases are at point B , that is at about 22% of the population. If these were all the cases, then one should deduce that we are at point C on the curve of total cases. In actual fact, the total cases at that time would be at point D , at about 55% of the population. The immune population R (the black curve) is at that time about 36% of the population and is the factor stopping the further increase of the active cases. Indeed, it is easy to calculate that at that time the susceptibles are at about 42%, so that $\mathcal{R}_0 \times S \approx 1$.

Figure 2 is not intended to provide accurate prediction. Many of the parameters are not precisely known, most notably the parameter ρ , and it does not reflect any interventions currently taking place. It is intended as a qualitative illustration of the role of the asymptomatic compartment in any predictions and on directing some research effort in estimating its size. One way is to try to establish the already existing immunity in the population by testing for antibodies. The research work in [8] has precisely this goal. Similar testing has been initiated in other countries as well. Due to the many specific factors, this data is likely to be country specific. Hence, tests need to be carried out in South Africa as well.

4 Modelling the COVID-19 spread in South Africa

The first case of COVID-19 in South Africa was confirmed on 5 March 2020. Until 20 March or so, the confirmed cases were dominated by individuals who travelled abroad. In view of the high prevalence of asymptomatic cases, it is highly likely that many more asymptomatic infective travellers entered South Africa. Since our model is based on deterministic differential equations, it can provide accurate representation of the infection spread only when the numbers are large. We initiate the model on 19 March, when there were 150 confirmed cases. We assume that at that time there were also 300 asymptomatic cases outside the health system records. Further, it is also assumed that there was influx of infective South Africans returning from abroad (as represented by $\lambda(t)$) until the lockdown came into effect. The influx from abroad is essential to explain the fast growth rate of confirmed cases until the lockdown. If this growth was a result only of local infection, this would imply a basic reproduction ratio far outside the indicated range. For the initial stage, with the given value of λ until the lockdown, we obtain a good fit with $\beta c = 0.2$ resulting in $\mathcal{R}_0 = 2.8$. A country-wide lockdown was implemented as from midnight of 26 March. This resulted in a significant drop in the increase of daily new confirmed cases. In fact, for the first two and half weeks or so, these remained almost constant. The data and the simulation are presented in Figure 3.

Until about 13 April the model is fitted to the data with $\beta c = 0.07$ resulting in $\mathcal{R}_0 \approx 1$, visible from the fact that the graph of confirmed cases (the blue line) is approximately linear. However, there is a visible exponential increase of confirmed

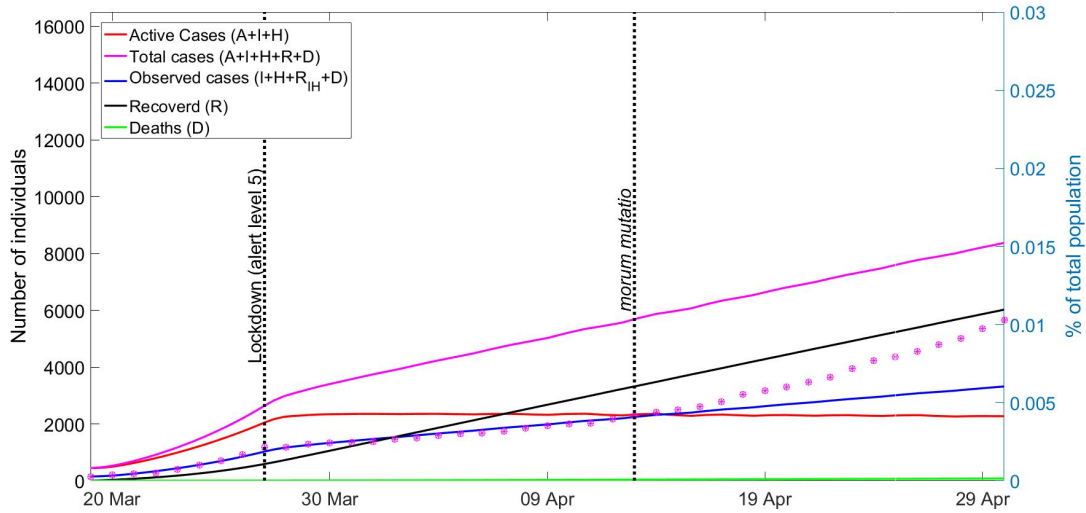


Figure 3: COVID in South Africa: Lockdown

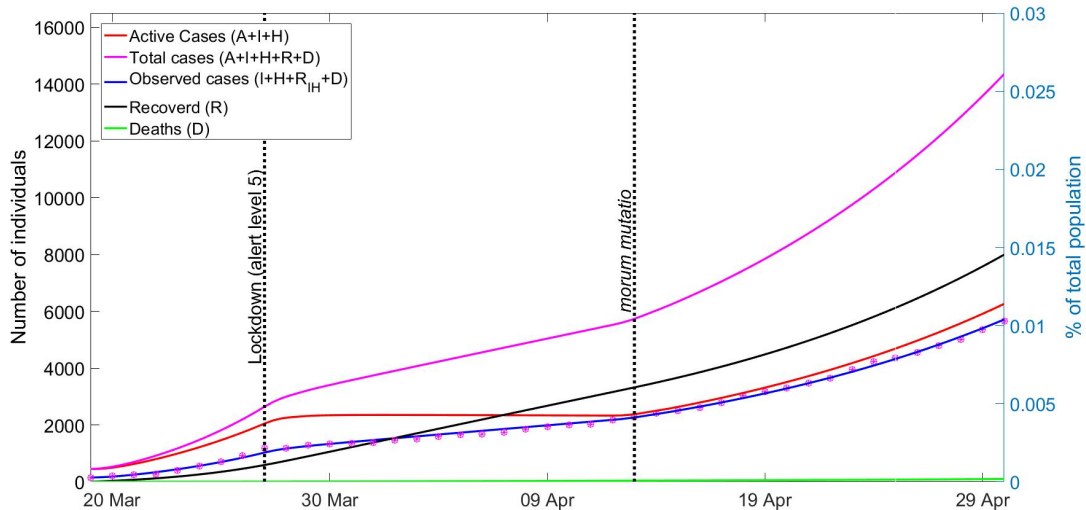


Figure 4: COVID in South Africa: *morum mutatio*

cases as from 13 April, deviating significantly from the near straight line for the first two and a half weeks. There could be reasons of different nature. It was widely reported in the UK media that, as the time under lockdown increases, the amount of pedestrian and motor traffic is also increasing. This lower level of compliance with the regulations is referred to as a lockdown fatigue. In South Africa, the regulations of the lockdown were amended a few times, which might also be a contributing factor. Another contributing factor could be black market activities related to the prohibition of sales of tobacco and alcohol. Addiction to either of the two could be strong driving factor of illegal social interactions. While the reasons are not clear, the fact of switch from approximately linear growth to exponential growth in the observable variable $I + H + R_{IH} + D$ about two and half weeks or so into the lockdown can be clearly seen in the available data. As we do not know the precise reason,

we will refer to the point of change of the lockdown efficiency as *morum mutatio* (behaviour change).

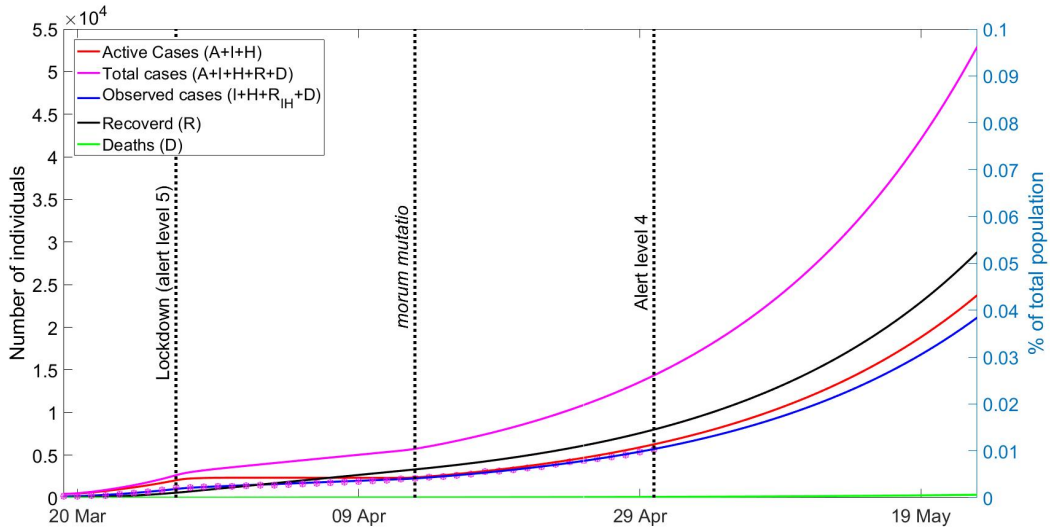


Figure 5: COVID in South Africa: Alert Level 4

On Figure 4 we present simulations, where the model is fitted to the period after 13 April with $\beta c = 0.137$ and associated basic reproduction number $\mathcal{R}_0 = 1.92$. We note that in this setting, due to asymptomatic cases, by the end of April 2020 the total number of cases could be over 14000 with 7500 or so who have already acquired immunity - a factor which will begin to play a significant role in the further dynamics.

The lockdown, also called Alert level 5, is changed as from 1 May to Alert level 4, which allows for some economic activity. It is expected that the basic reproduction ratio will not significantly increase as this would require a new lockdown. Figure 5 indicates that if the infection progresses with the same basic reproduction ratio, in 3 weeks the total cases are expected to cover nearly 0.1% of the population, producing immunity which surpasses the number of recorded/confirmed cases.

Quantitative predictions beyond the mentioned period are not likely to be accurate due to the many unknowns. On the one hand, the parameters, most notably ρ , are yet to be precisely determined. On the other hand, one cannot predict what further actions health authorities or governments may take, or how human behaviour and interactions may change. Nevertheless, in order to illustrate the significance of ρ , we run long term simulations for $\rho = 0.4$ (Figure 6), as in the simulations so far, and with $\rho = 0.2$ (Figure 7). In addition to the other graphs, in these figures we present the graph of the active symptomatic infections (dashed red line). One can observe that while the graphs of the other presented variables are more or less the same, the graphs of the active symptomatic infections ($I + H$) and the total symptomatic cases ($I + H + R_{IH} + D$) are quite different. The graph of active symptomatic cases peaks at 5.41% for $\rho = 0.4$ (Figure 6), while for $\rho = 0.2$ this peak is at 2.71% (Figure 7). Further, the saturation level of the total symptomatic cases ($I + H + R_{IH} + D$ - the blue line) is at about 30% for $\rho = 0.4$ and at about 16% for

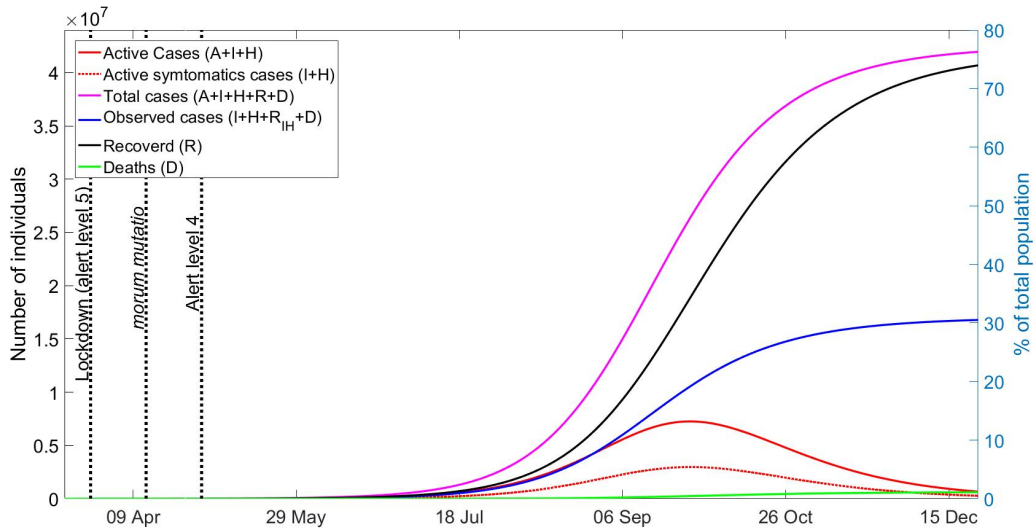


Figure 6: Long term dynamics with $\rho = 0.4$

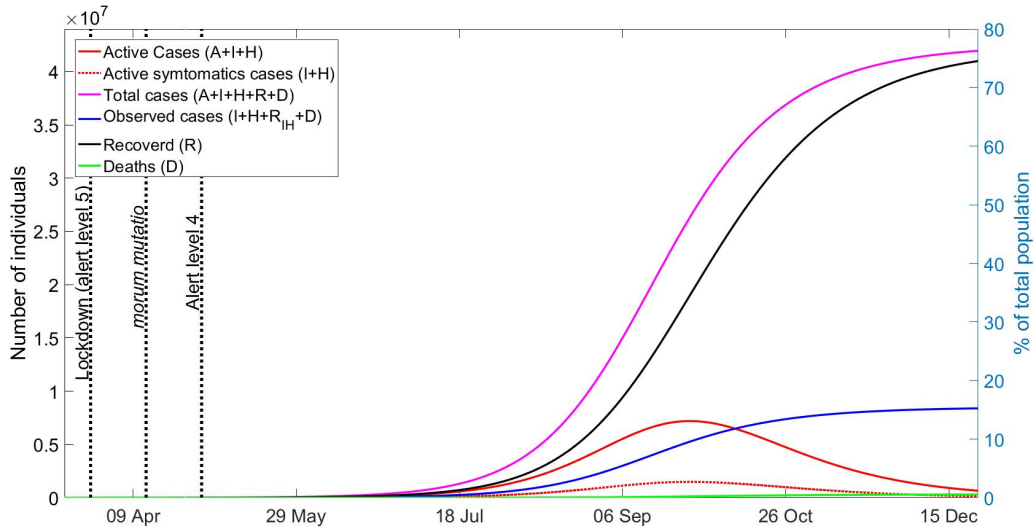


Figure 7: Long term dynamics with $\rho = 0.2$

$\rho = 0.2$.

The size of the symptomatic infections is a critical variable which needs to be below a certain level so that the health system can cope. We recall that a fraction of the symptomatic infective individuals would need hospitalization and a fraction of them will need critical care and possibly ventilators. Hence, knowing the ratio of symptomatic to asymptomatic infections is of crucial importance for determining the time and the size of the peak of active infective cases in any relevant setting and, therefore, inform an appropriate action.

5 Conclusion

In this paper we suggest that further COVID-19 epidemiological research does not just need more data, but it also needs data which goes beyond the case count captured by the health system. The importance of testing for the virus cannot be doubted as it is an important tool of reducing the spread through quarantine measures, thus reducing βc or equivalently \mathcal{R}_0 and flattening the curve of active infective cases. However, the long term dynamics is strongly impacted by the level of immunity acquired by the population. It is built through both symptomatic and asymptomatic infections. Since the latter ones, while likely to be the majority of the cases, are mostly unrecorded, they represent a significant unknown factor for the long term epidemiological dynamics.

The only relevant testing that we are aware of, is the testing for antibodies, under the assumption that the presence of antibodies can provide immunity for 6-18 months, as suggested by some authors. We fitted the model to data for South Africa on the total number of confirmed cases (the blue curve).

Part of our future research work is to focus on better estimation of the unknown parameter ρ . To this end, we will be following closely all research effort on testing for antibodies, as this would provide data which the curve of the recovered (the black curve) can be fitted to.

References

- [1] José Lourenço, Robert Paton, Mahan Ghafari, Moritz Kraemer, Craig Thompson, Peter Simmonds, Paul Klenerman, Sunetra Gupta, Fundamental principles of epidemic spread highlight the immediate need for large-scale serological surveys to assess the stage of the SARS-CoV-2 epidemic, preprint, <https://doi.org/10.1101/2020.03.24.20042291>
- [2] Neil M Ferguson, et al. (Imperial College COVID-19 Responce Team), Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand, <https://doi.org/10.25561/77482>.
- [3] Biao Tang, et al., Estimation of the Transmission Risk of the 2019-nCoV and Its Implication for Public Health Interventions, *Journal of clinical medicine* 9 (2020), 462, <https://doi.org/10.3390/jcm9020462>
- [4] Government Department: Statistics of South Africa, Mid-year population estimates, Statistical Release P0302, 2019, www.satassa.gov.za
- [5] Michael Day, Covid-19: four fifths of cases are asymptomatic, China figures indicate. *BMJ* 2020; 369:m1375. <https://doi.org/10.1136/bmj.m1375>
- [6] Michael Day, Covid-19: identifying and isolating asymptomatic people helped eliminate virus in Italian village. *BMJ* 2020;368:m1165. <https://doi.org/10.1136/bmj.m1165> pmid:32205334
- [7] "Iceland coronavirus tracker map". visalist.io.

- [8] H. Streeck et al. Vorläufiges Ergebnis und Schlussfolgerungen der COVID-19 Case-ClusterStudy (Gemeinde Gangelt)
- [9] Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Early Spread of SARS-Cov-2 in the Icelandic Population. medRxiv. 2020:2020.03.26.20044446.
- [10] John P. A. Ioannidis, Cathrine Axfors, Despina G. Contopoulos-Ioannidis, Population-level COVID-19 mortality risk for non-elderly individuals overall and for nonelderly individuals without underlying diseases in pandemic epicenters. medRxiv preprint, <https://doi.org/10.1101/2020.04.05.2005436>
- [11] Ying Liu, Albert A. Gayle, Annelies Wilder-Smith, Joacim Rocklöv, The reproductive number of COVID-19 is higher compared to SARS coronavirus, Journal of Travel Medicine, Volume 27, Issue 2, (2020) <https://doi.org/10.1093/jtm/taaa021>
- [12] World Health Organization, Coronavirus disease 2019 (COVID-19) Situation Report – 46. (2020) https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200306-sitrep-46-covid-19.pdf?sfvrsn=96b04adf_2