medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A dynamic model for Covid-19 in Brazil.

Rubens Lichtenthäler Filho [a] and Daniel Gomes Lichtenthäler [bcd]

[a]Departamento de Física Nuclear, Instituto de Física, Universidade de São Paulo, 05508-090, São Paulo, Brazil.

[b]Hospital Israelita Albert Einstein, São Paulo, Brazil.

[c]Hospital das Clínicas da Faculdade de Medicina, Universidade de São Paulo, São Paulo, Brazil.

[d]Faculdade de Medicina, Universidade Santo Amaro, São Paulo, Brazil.

April 2020

1 Abstract

A dynamic model for the current coronavirus outbreak is presented. The most important parameters are identified which determine the number of cases progression. Results of a numerical simulation are compared with existing data of the number of infections in Sao Paulo and Brazil. On the basis of these results measures are proposed to control the epidemics and to flat the infection curve. A simple three steps procedure is proposed to predict the time evolution of the epidemics and a criteria to resume 'normal' activities after quarantine is proposed.

2 Introduction

The multiplication process of the number of cases in an epidemics is primarily determined by the infection rate parameter (R_0). This dimensionless quantity is defined as the average number of new infections caused by each infected person. In the case of SARS-COV2 outbreak in Sao Paulo this number is estimated to be between 2-6, if no social distancing measures are taken and the epidemics is left to follow its natural course. The infection rate (R_0) is the parameter that controls the number of new infections in a certain moment given the number of infected people in the days before. The assumption that the new number of cases is completely determined by the product between R_0 and the current

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

number of infections implies directly that the epidemics curve will be an exponential whenever $R_0 > 1$. As R_0 increases, the steepness of the exponential also increases. The infection rate parameter R_0 can be reduced by adopting measures of social distancing and quarantine [1, 2]. A decrease in the social interaction rate and hygiene measures have the potential to strongly decrease R_0 reducing the steepness of the exponential curve and flattening the infection curve. For $R_0 \leq 1$ the infection curve becomes flat and the epidemics is controlled.

Another important parameter in the dynamics of the epidemic is the fraction of the population which became immune to the virus. Its complementary quantity is fraction of the population susceptible to the virus (T). T is given as the ratio between the number of susceptible people and the total population and decreases in time as more and more people are being healed reducing its transmission probability. The fraction of susceptible people should decrease in time, assuming that all the healed population became immune although recent findings indicate that the latter assumption may not be rigorously valid [3].

Other important parameters are the incubation time and the healing time. These parameters basically determine all the dynamics of the epidemics.

In the next sections we will present details of the model and compare the results of the numerical simulation with epidemiological data.

3 Numerical Model and results

The algorithm calculates the number of new infections in a certain day, given the number of infections in the previous days. The total number of infections in a day n is obtained by the equation below:

$$a(n) = T_{n-1} * \sum_{1}^{n-1} \frac{p(i) * R_0 * a_{new}(i)}{\sum_1^{n-1} p(i)} + a(n-1). \tag{1}$$

Where $a_{new}(i)$ stands for the new infections in each day i. Note that the number of new cases in day n, given by the summation in the right side of Eq.1, is obtained from the contribution of all previous cases since the beginning of the epidemics (day 1) up to day n-1, weighted by a probability $p(i)$ which is a function of time. This probability was taken as a Gaussian distribution centered in a distance backwards from day n equal to the incubation time (τ_{inc}) with a given width (w). Every infected person will contribute to a certain number of new infections with a probability that depends on how far backwards it is from day n. This probability increases from day 1 up to a maximum in the incubation time day, decreasing down to day n-1. The Gaussian is normalized to 1 by the factor in the denominator.

The factor (T_n) multiplies the summation in the right side of Eq.1 and is an important one. It is defined as $T = \frac{N_{pop} - N_{inf}}{N_{pop}}$, where N_{pop} is the total population and N_{inf} is the current total number infections. As N_{inf} is a function of time, T_n will also be a function of time starting from $T = 1$ ($N_{inf} = 0$) going down to $T < 1$ as the total number of infections N_{inf} increases.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The number of new cases in day n has to be added to the total number of cases until day n-1 to obtain the cumulative number of infections.

In Figure 1 we plot the cumulative number of current infections given by Eq.1 versus the epidemic time. The epidemic time is measured in days, starting in the day of the first infection. The infection curve initially rises exponentially decreasing its inclination as it approaches the plateau. In the plateau the number of new infections falls exponentially to zero as $T_n \to 0$ and the epidemics stops.

In Figure 1 we present the results of the simulation for three different infection rate parameters (solid lines) $R_0 = 3.5, 1.9, 1.4$ compared to the reported number of infections in Brazil (blue circles). The other parameters of these calculation are: incubation time τ_{inc} =4.4 days, width of Gaussian distribution w =2.2 days and a total population of 200 million people. The reported number of cases in Brazil (blue circles) have been multiplied by a constant factor 7.14 which stems from the estimated undocumented factor of 86% as reported in [1]. With this correction factor our mortality rate is around 1%.

The black squares are the São Paulo state data compared to the black curve for $R_0 = 1.3$ and 10 million people.

The curves are plotted in logarithm scale. This is an important remark since, by taking the log of an exponential, any terms multiplying the exponential become constant and are washed out when the derivative is taken. As R_0 is related to the derivative (inclination) of the curves, this means that the determination of R_0 is independent of constant multiplicative factors in the exponential. As a consequence, even if the data are underestimated as they are, that would not affect R_0, as long as the underestimating factor stays constant in time.

In Figure 1 we clearly see 3 regimes in the time evolution. Until 23/03/2020 data clearly are following the orange curve corresponding to $R_0 = 3.5$. From 23/03 up to 08/04/2020 it follows $R_0 = 1.9$ (red curve) and after that until 08/05/2020 it follows $R_0 = 1.4$. The doubling time is indicated in the figure. The relation between R_0 and the time to double the number of cases (doubling time τ_d) in the ascending part of the curve can be estimated as:

$$\tau_d \approx \frac{1}{\log_2 R_0} * \tau_{inc} \tag{2}$$

where τ_{inc} is the incubation time.

The correspondence between the epidemic time (x-axis) and the real time is not trivial. This relation is of great interest since it would allow to predict how far we are from the peak. In principle there is no way to determine this parameter other than by a measurement of the real total number of cases in a certain time. This is virtually impossible, since it would require testing all the population. The determination of the true number of infections requires a precise measurement of the undocumented factor which is not easy since it depends on particular characteristics of the testing methodology in each particular country. On the other hand, random surveys of the type performed in elections could, in principle, give a good idea of this parameter. Random surveys performed in selected samples of the population allow to determine the percentage of the

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

population which has already been in contact with the virus, providing and estimation the T_n factor which, by its turn, would give an idea of the position in the epidemic time. Unfortunately such surveys are not available in Brazil yet.

In Figure 2 we compare the number of new cases reported every day in Brazil to the simulation results for R0=1.4 and the same parameters as quoted in Figure 1 (black curve). The calculation have been re-normalized by an arbitrary factor and shifted (x-axis) to be compared to the data. The data goes up May,8. Changes in R_0 due to relaxation in the social distance measures can modify this prediction however the all curves have been calculated with the same parameter but normalized by different constants and shifted in time by different amounts. This corresponds to assume different total population in the sample. We see that all curves are consistent with the data but give different predictions for the position of the peak indicating that it is trivial to predict when the maximum of infection with be attained [5]. Different total populations can modify the peak position. Another important factor that has to be taken into account is the delay in the reporting time. there are indication the, in Brazil, the reported data can be delayed by up to 14 days implying that the real position in the curve should be shifted to the right by this amount of time.

In Figure 3 we show the result of two calculations. The red curve is the total number of infections and the black curve is the infection curve for $R_0 = 1.4$ and population of 10 million. For $R_0 = 1.4$ 51% of the population is contaminated [6]. The vertical dashed line shows the position of the peak its correspondence on the total number of infections curve. We see that at the peak of the epidemics about 3.3 million cases (red curve) in a total of 5.1 million are reported. It means that, in the peak of the new cases contamination curve, about 65% of the total number of infections is reached, corresponding to 33% of the total population. The latter corresponds to the herd immunity factor ($R_0 \times T_n \leq 1$) as quoted in [6]. This will be discussed in more details in the next subsection.

After the peak, the number of healing becomes larger than the number of infections and the total number of current cases decreases. As the healing time is of 14 days for COVID-19, after that time the healing curve will get to its peak. At this position, about 88% of the infections took place and the contamination will now decay exponentially to zero. We propose this as a criteria to start relaxing quarantine and social distance measures.

3.1 Herd immunity

Herd immunity can be understood as the percentage of the population that needs to be immune in order to slow down the spread of the virus. It can be defined as the value of T_n below which the product $R_0 \times T_n^{herd} \leq 1$. T_n is a decreasing function of time as discussed above and it is given as the fraction of the population susceptible to infection. Then, for a given R_0, as T_n get to a value below which $T_n \leq \frac{1}{R_0}$ the the effective infection rate (RT_n) turns out to be smaller than one and the epidemic becomes controlled. For $R_0 = 3.5$ one gets $T_n^{herd} = 0.28$, meaning that $1 - T^{herd} = 72\%$ of population is immune. The herd immunity depends on R_0. For $R_0 = 1.4$ the fraction of immunity drops

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

down to 28%. For $R_0 = 2$, $T^{herd} = 50\%$. This is an important parameters since it gives an idea of the magnitude of the immune population sufficient to stop the exponential epidemic spread (see Figure 4.

Assuming that present São Paulo/Brazil's R_0 ranges between $1.4 - 2.0$, one would expected $28 - 50\%$ required immunity to stop epidemics, as long as R_0 stays constant in this range.

In the light of these findings we propose a simple three steps procedure to estimate the herd immunity directly from epidemiological data:

- Determine the infection doubling time from epidemiological data.
- Apply the simple formula [1] that relates R_0 with the infection doubling time to obtain $R_0 = 2^{\frac{\tau_{inc}}{\tau_d}}$ where τ_{inc} and τ_d are incubation and doubling time respectively.
- Once R_0 is known, the herd immunity can be obtained as $1 - 1/R_0$.

4 Conclusions.

We develop a model and a computer code to calculate the dynamics of Covid-19. The model is a simple one but contains all the important ingredients, to be know, the infection rate parameter R_0, incubation time, healing time and the immunity factor (T).

Calculations have been performed for different R_0 factors and comparison with epidemiological data shows that it is possible to determine quite precisely the value of the infection rate R_0 even if the data are sub-notified as long as the sub-notification factor is constant. The result of our calculations show there have been 3 moments in this epidemics in Brazil so far, with different R_0 factors $R_0 = 3.5; 1.9$ and 1.4. In Sao Paulo the infection curve is following a little smaller $R_0 = 1.3$ value. The observed reduction of R_0 was probably induced by the quarantine and social distancing measures and eventually lock-down adopted by State governments. It became clear that the social distancing politics is essential to reduce the total number of cases and to control the epidemics. The total number of infections drops from 97% down to 51% as the infection rate falls from $R_0 = 3.5$ to for $R_0 = 1.4$. We show that, at the peak of the contamination curve, about 64% of the total contamination has been reached and 14 days after the peak about 88% of the total contamination took place suggesting the healing time as a possible criteria to start relaxing social distance measures.

Conducting random surveys to determine the percentage of infected people in the population is a critical measure to determine the position in the epidemic curve, which would by its turn, provide as estimation of the distance to the peak. The fraction of infected people (1-T_n) in the population is an important parameter to estimate how far we are from the herd immunity condition (for a given R_0) and to guide the implementation of relaxation in the social distancing

[1] this formula is the inverse of equation 2.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

measures. This percentage can be obtained from serological tests performed in random selected samples of the population. Otherwise, if there is no control of the percentage of infected people, it turns out very difficult to make reliable predictions of how far we are from controlling the epidemics.

In order to implement these procedures we have to keep in mind that all the official reports provide a delayed picture of the epidemics and the delay time depends on several factors as incubation time, time to hospitalization, reporting time. An estimation of the reporting delay time is highly desirable in order to give a reliable estimation of the time position in the epidemic curve.

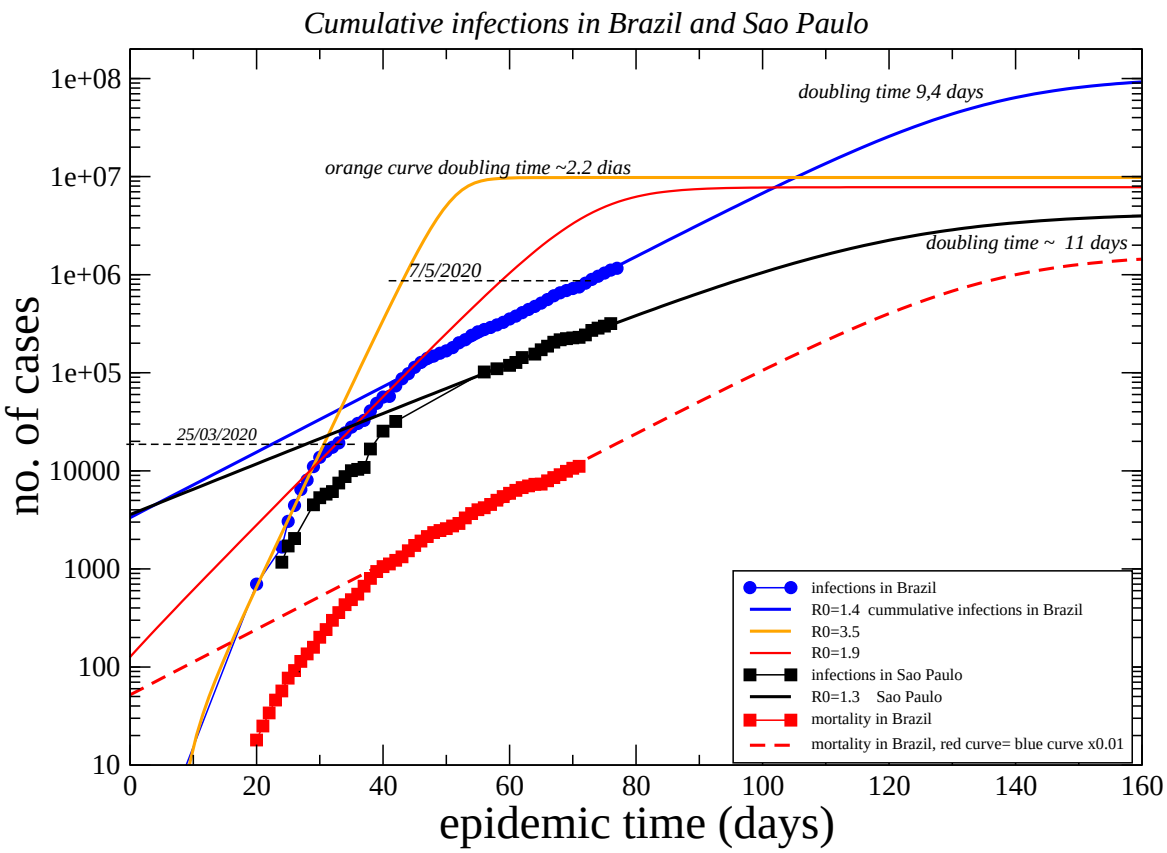


Figure 1: Contamination curves (solid lines). Number of cases in Brasil (blue circles).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

[1] Ruiyun Li, Sen Pei, Bin Chen, Yimeng Song, Tao Zhang, Wan Yang, Jeffrey Shaman.
Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)
Science 10.1126/science.abb3221 (2020).

[2] Seth Flaxman *et al.*, Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 countries.
Imperial College COVID-19 Response Team, doi:10.25561/77731

[3] OMS report.

[4] Wölfel, R., Corman V.M., Guggemos, W., *et al.*.
Virological assesment of hospitalized patients with COVID-19
Nature (2020) doi:10.1038/s41586-020-2196-x

[5] Predictive Monitoring of COVID-19
Data-driven Innovationn Lab, Singapore University of Technology and Grugs (2020).

[6] S. Cobey, *Science*10.1126/science.abb5659 (2020).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

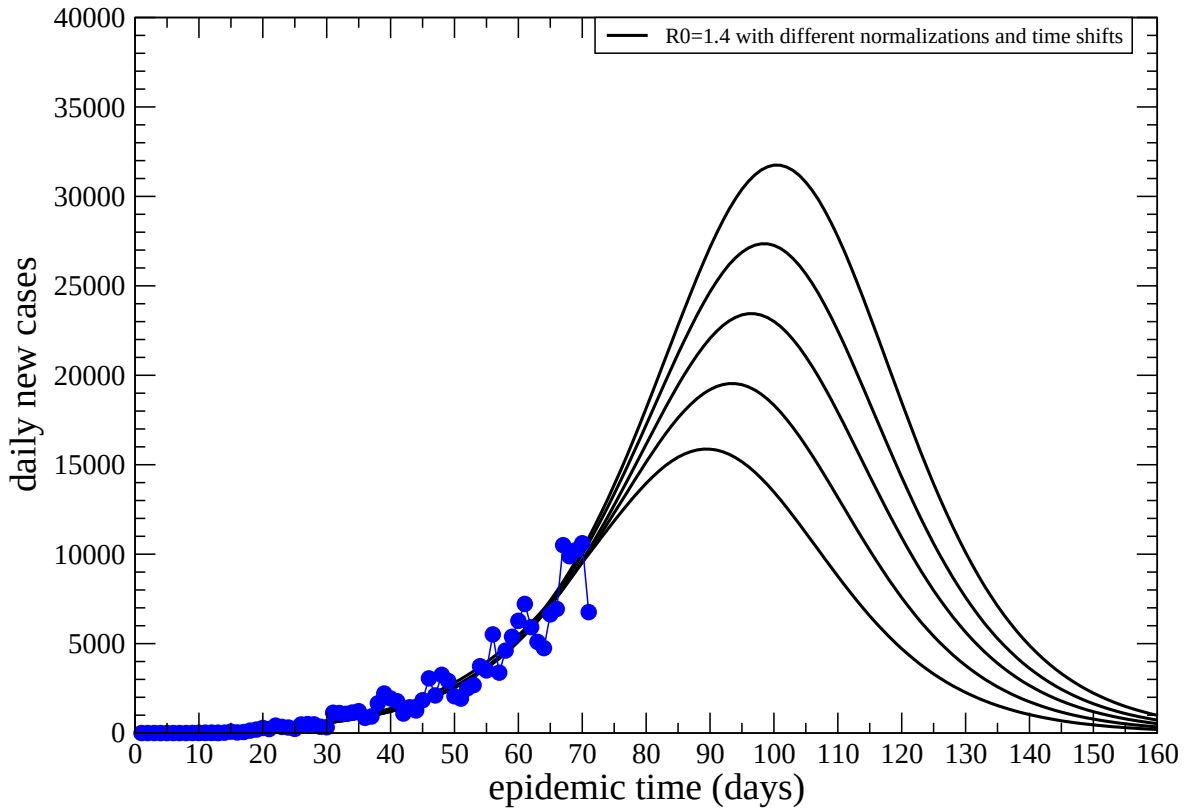


Figure 2: Daily number of new cases in Brazil (black dots), solid line is the result of the simulation with the same parameters as in Figure 1 (blue curve) with different normalization and time shift factors.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

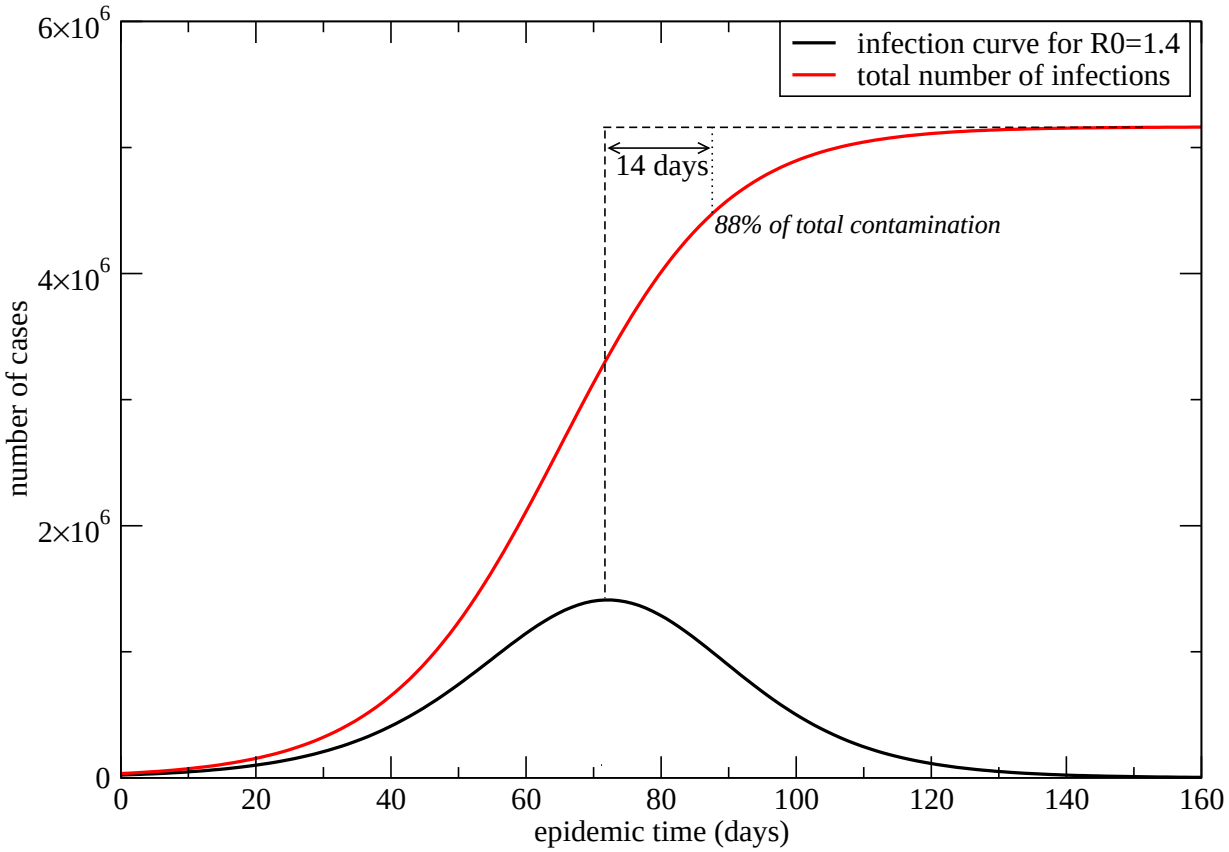
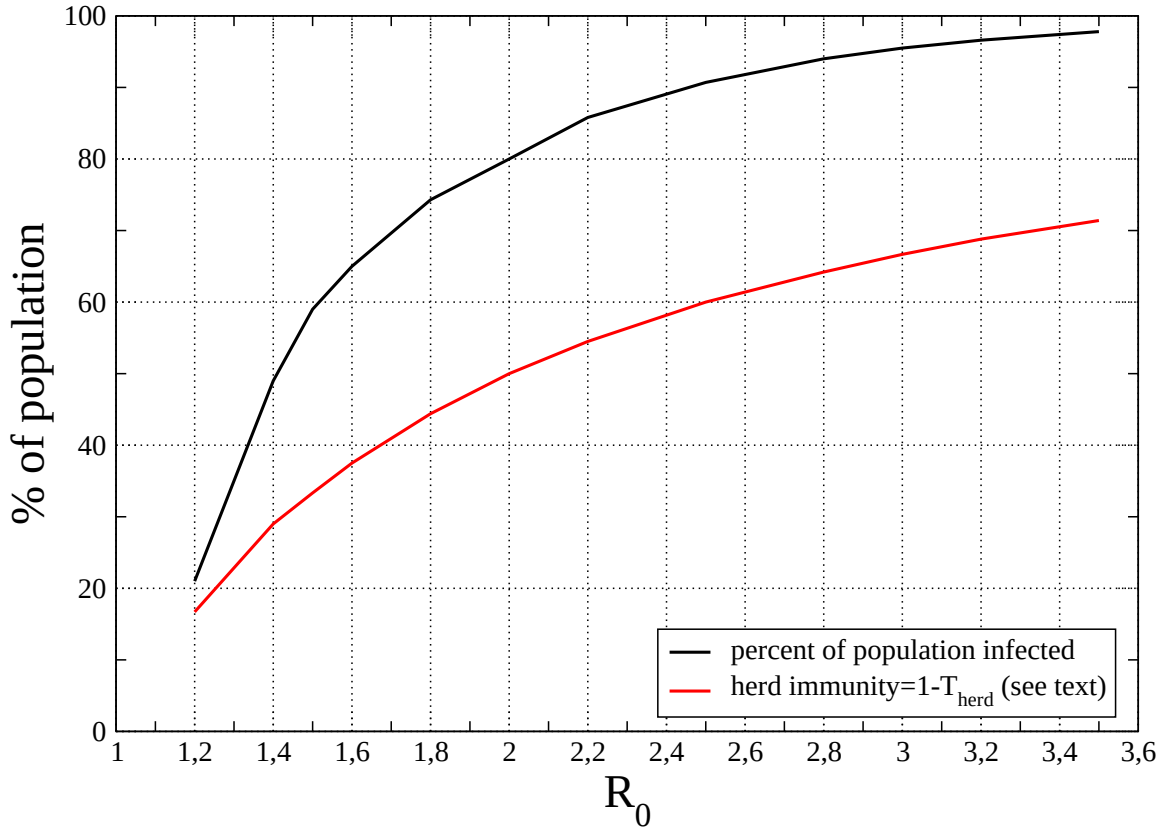


Figure 3: Integrated number of infections (red line) and total number x time. Calculation performed for $R_0 = 1.4$ and 10 million population.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.10.20097550; this version posted May 14, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



Percent of population contamined and herd immunity as a function of R_0.

Figure 4: