

Why, when and how will the Corona epidemic end?

Trivial Notes

Avishai Dekel^{1*}

¹*Racah Institute of Physics, The Hebrew University, Jerusalem 91904 Israel*

17 March 2020

ABSTRACT

Trying to teach myself very elementary epidemiology, I analyzed a simple model where the evolution all depends on the “basic reproduction number” R_0 , the average number of people infected by an infectious. For COVID-19 it seems that it starts with $R_0 \sim 2-3$ in the exponential growth phase, a doubling time of $\sim 3-4$ days, consistent with the current data. If nothing is done to reduce R_0 , about 80% of the world population is expected to be eventually infected ($\sim 1\%$ dead). The peak, to be reached in $\sim 3-4$ months, is expected to involve $\sim 15\%$ of the population, namely $\sim 1-2\%$ in critical conditions that will require respiratory machines. This cannot be managed by any medical system. R_0 has to be reduced to unity (a doubling time of ∞) in order to shallow the peak to the medical-system capacity and reduce the final infected fraction to significantly below unity. Another possibility is to reduce R_0 to 1.1, which will postpone the peak by a few years, allowing a vaccine to be developed, but the Coronavirus will be with us for long term. R_0 can be reduced by obvious measures of (a) reducing the risk given a contact, e.g., by washing hands, masks, etc., and (b) reducing the number of contacts with infectious by isolation, avoiding flights, etc. The ultimate solution would be a full quarantine for a month, starting now.

Key words: Corona — COVID-19

1 INTRODUCTION

Very early in the current Coronavirus epidemic, I wanted to obtain a simple understanding of how an epidemic evolves, why and when it is supposed to end, what will be the fraction of sick people at the peak and when it will happen, and what fraction of the population is expected to be infected (and die) overall. This may help interpret the data poured on us without relying on “experts” in the media, and evaluate the prospects of the epidemic.

The following seems to be the simplest, straightforward model, which makes sense to me as a physicist. In one version or another it is supposed to be common wisdom among epidemiologists. Not being an expert myself, I enjoyed learning the very basics from David Earn, a professor of mathematical epidemiology in McMaster University (and my former astrophysics postdoc at HU).

In this simplest form, I expect the model to be valid as an approximation for a self-interacting closed community (a town, a country, but maybe even the whole

humanity), assuming that the basic reproduction number is constant at least for a given period. It should be generalized for the case of a reproduction number that is varied by force. To understand the early stages where most of the infections in a given community still occur by import (as it seemed to be in Israel during February), the model has to include interaction between clusters. However, my epidemiologist friends tell me that simple models like the one described below turned out to be crudely valid in quite general circumstances of other epidemics.

2 THE SIR MODEL

2.1 Parameters and equations

Assume that the population is divided to three groups, Susceptibles, Infectious and Removed (cured or dead), with fractions S , I and R respectively, such that

$$S + I + R = 1. \quad (1)$$

Assume that there is one way evolution from S to I to R , with no return from R to I .

Let γ be the **removal rate** from I to R , namely γ^{-1}

* E-mail: dekel@huji.ac.il

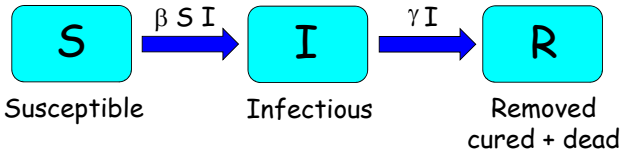


Figure 1. The simple SIR model. Three sub-populations with fractions S , I and R of the total populations. One way evolution, from S to I at a rate βSI and from I to R at a rate γI . γ^{-1} is the duration of being infectious, and can serve as the time unit. The parameter $R_0 = \beta/\gamma$, which determines the fate of the epidemic, is the average number of people from S that are infected by each person from I .

is the average contagious period, assumed to be the time between infection and removal ($\gamma^{-1} \sim 1 - 2$ weeks). Set this here to be the time unit, namely $\gamma = 1$.

Let β be the **transmission rate**, the average number of contacts between S and I that lead to a new I , per unit time, per I , per S .

The model is specified by the three differential equations (a dot refers to derivative with respect to time), describing the evolution of each group,

$$\dot{S} = -\beta SI, \quad (2)$$

$$\dot{I} = \beta SI - \gamma I, \quad (3)$$

$$\dot{R} = \gamma I. \quad (4)$$

The most important parameter is the **basic reproduction number**, defined as the ratio of rates,

$$R_0 = \beta/\gamma. \quad (5)$$

This is the average **number of secondary cases caused by a primary case**. It is the ratio between the duration of being infectious γ^{-1} to the average time between infectious contacts (per I per S) β^{-1} . We assume below that R_0 is constant in time, but the hope is that it can be reduced intentionally by preventive actions.

2.2 Numerical solution

Figure 2 shows a numerical solution of the equations with $R_0 = 2$ (as currently assumed for COVID-19), assuming an initial $I_i = 10^{-5}$ (e.g., in March 6, 2020, about 100,000 infectious out of a population of 10 billion). The infectious fraction (I) grows exponentially, reaches a peak, and then drops exponentially, similar to a Gaussian bell shape. The total infected $Z = R + I$ fraction also grows exponentially, and eventually saturates at the final size Z_∞ . The model predicts that, with $R_0 = 2$ remaining constant, the final size is $Z_\infty \simeq 0.8$, namely about 80% of the population will eventually be infected. The peak will involve $\sim 15\%$ of the population sick at that time, and half the population infected by that time (eq. (16)). This peak is to be reached in 80 – 160 days (since the time when $I_i = 10^{-5}$, for $\gamma^{-1} = 1 - 2$ weeks respectively).

The final size and the peak level and time will depend on how far would R_0 be reduced. In order to avoid

having most of the world population infected, and in order to avoid a catastrophe of the medical services at the peak (as is now happening in Italy), R_0 has to be reduced by a factor of order two, namely to unity or less. Figure 3 shows the evolution for $R_0 = 2, 1.3, 1.1$. With $R_0 = 1.1$, the total infected is less than 20%, and the peak, after 9% of the population have been infected (0.5% infectious at the peak time) is postponed to 500 – 1,000 days. This is still a high fraction of the population, so $R_0 = 1.1$ is not good enough, but it gives a chance for a vaccine to be developed.

2.3 Exponential growth

Assume that at a given initial time $I = I_i \ll 1$. Approximate at that time $R_i = 0$ (a different assumption will not make a qualitative difference). Then $S_i = 1 - I_i \simeq 1$, so by integrating eq. (3) we get

$$I(t) \simeq I_i e^{(R_0 - 1)\gamma t}. \quad (6)$$

If $R_0 > 1$ this is an exponential growth, while for $R_0 < 1$ this is an exponential decay. Using $S \sim 1$ and eq. (6), and integrating eq. (4), we obtain for the total being infected, $T = R + I$,

$$T(t) \simeq \frac{I_i}{R_0 - 1} [R_0 e^{(R_0 - 1)\gamma t} - 1]. \quad (7)$$

For R_0 significantly larger than unity and slightly after $t = 0$, the second term is negligible, so we have an exponential growth. This is a straight line in the semi-log growth curve

$$\log T(t) \simeq 0.43 (R_0 - 1) \gamma t + \log \left(\frac{I_i R_0}{R_0 - 1} \right), \quad (8)$$

where the 0.43 factor stands for $\log e$ in the transition from natural to decimal log.

The value of R_0 can be determined empirically by comparing the model slope

$$\text{slope} = 0.43 (R_0 - 1) \gamma \quad (9)$$

with the best-fit experimental slope of $\log T(t)$. If both $R(t)$ and $I(t)$ are known empirically, then R_0 can also be read from the offset

$$\log T - \log I \simeq \log [R_0 / (R_0 - 1)]. \quad (10)$$

If the empirical doubling time τ_2 is known, then from eq. (9)

$$R_0 \simeq 1 + 0.7\gamma^{-1}/\tau_2. \quad (11)$$

If $\gamma^{-1} \sim 10$ days, and $DT \sim 3.5$ days as indicated empirically from Fig. 4, then $R_0 \sim 3$, namely still a severe uncontrolled exponential growth in most countries.

Figure 4 shows the empirical recent evolution of Covid-19 in several countries, some after the peak (China, South Korea), and the rest during the early uncontrolled growth phase. As predicted by the model in the growth phase, we see an exponential growth in each country, with different starting points but rather similar slopes. Interestingly, the behavior is similar worldwide.

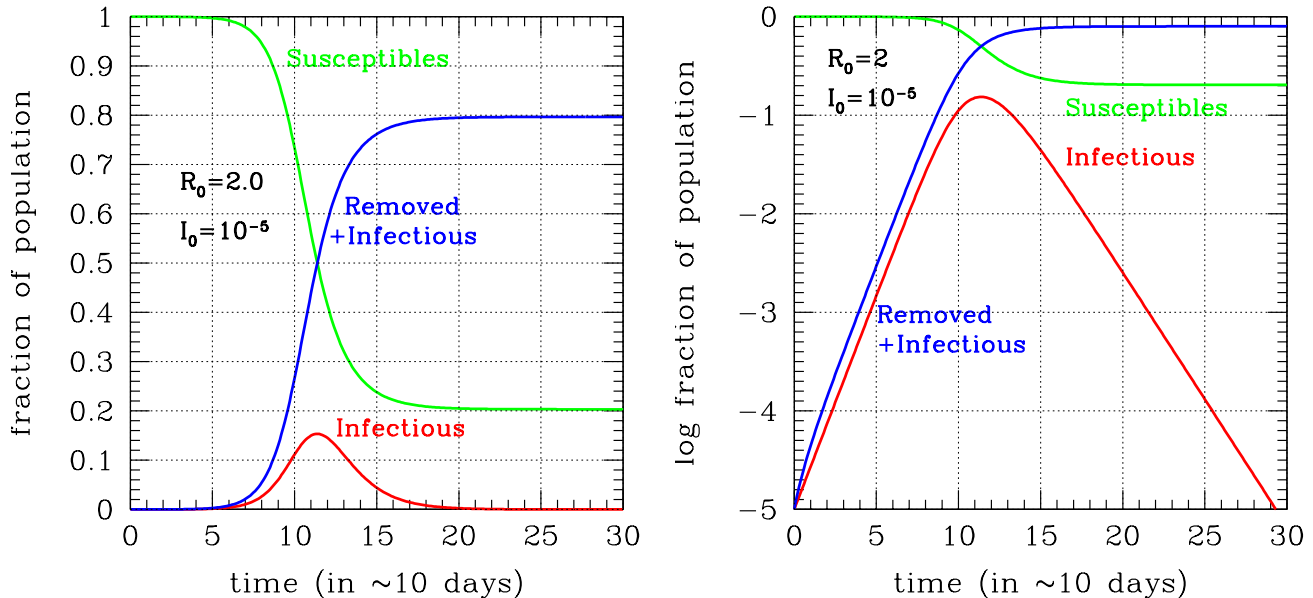


Figure 2. Evolution of an epidemic according to the SIR model with a basic reproduction number $R_0 = 2$, as currently estimated for COVID-19. Shown is the fraction of the population in the three groups: Susceptibles, Infectious, and Removed (plus Infectious). The time is in units of the duration of the infectious phase, set to $\gamma^{-1} = 1$ (assumed to be about 1-2 weeks for COVID-19). The peak time depends on the assumed initial fraction of infectious, I_i (marked I_0), here taken to be 10^{-5} , corresponding to 100,000 Infectious in a population of 10 billion (as for COVID-19 in early March 2020). We learn that 80% of the population is expected to be infected, with a peak of 15% to be reached at $t \simeq 11.5 \gamma^{-1}$, which is $\sim 80 - 160$ days. This is unless R_0 is reduced toward unity.

If the duration of the infectious period is $\gamma^{-1} \sim 10$ days, and the doubling time is ~ 3.5 days, then $R_0 \sim 3$.

The other extreme limiting case is when any contact between S and I is avoided, namely $R_0 = \beta = 0$, then from eq. (3) we obtain an exponential decay,

$$I(t) = I_i e^{-\gamma t}. \quad (12)$$

Thus, a total confinement of the whole population (since the infectious are not all known) for a period $\gtrsim \gamma^{-1}$ is the ultimate way to stop an epidemic, though not too practical in most cases.

2.4 A solution curve in phase plane

A solution curve of the equations in the phase plane, connecting S and I, is

$$I + S - (I_i + S_i) = R_0^{-1} \ln(S/S_i). \quad (13)$$

Can check it by taking a time derivative and comparing to eqs. (2) and (3). For a new epidemic $S_i \simeq 1$ and $I_i \simeq 0$.

2.5 The peak

From eqs. (3) and (5)

$$\dot{I} = (R_0 S - 1) \gamma I. \quad (14)$$

If $R_0 \leq 1$, then $\dot{I} \leq 0$, so this is not an epidemic. If $R_0 > 1$, this is an epidemic. It starts slowing down

when I is at maximum, namely $\dot{I} = 0$. From eq. (3), this peak of I happens at

$$S_m = \frac{\gamma}{\beta} = R_0^{-1}. \quad (15)$$

Using the solution curve of eq. (13), inserting S_m from eq. (15) (as well as $S_i = 1$ and $I_i = 0$), one obtains the fraction of the population that is infectious at the peak

$$I_m = 1 - R_0^{-1} (1 + \ln R_0). \quad (16)$$

This is ~ 0.15 for $R_0 = 2$, namely a significant fraction of the population is sick during the peak. If the peak of I is too high, it could lead to a collapse of the medical system, thus increasing the death rate among the infected. By reducing R_0 one can lower the peak to the level that could be handled by of the medical system. With $R_0 = 1$ the peak is reduced to zero (above I_i).

Figure 5 shows the peak height I_m and peak time t_m as functions of R_0 . If R_0 is maintained at 2 – 3, the peak level is 15 – 30% of the population, guaranteeing a collapse of the medical system. When R_0 is reduced toward unity, the peak height is reduced, but it may become manageable only when R_0 is really reaching unity or less. Along the "flattening of the curve", the reduction of R_0 is associated with a delay in the peak time, pushing it to a few years once R_0 is below 1.1. This means that if R_0 is not reduced to below unity, the epidemic will stay with us for a very long term.

Assume that the medical system can provide respiratory treatment to 10^{-4} of the population simul-

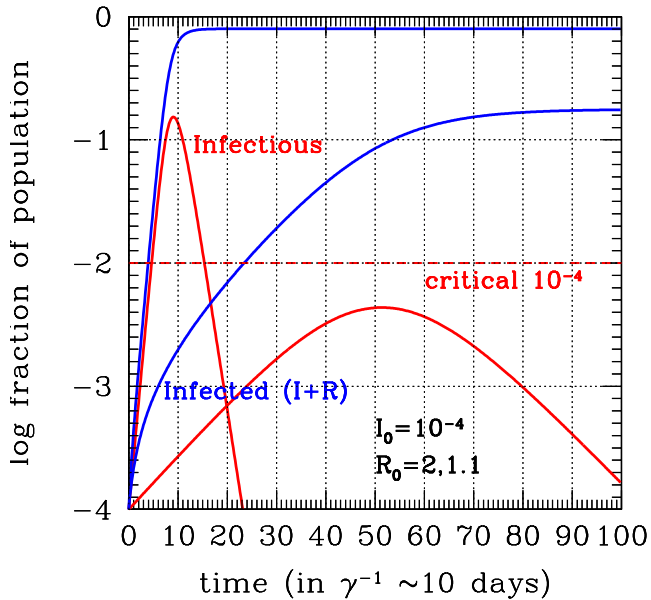


Figure 3. Evolution of log fraction as a function of time (in $\gamma^{-1} \sim 10$ days). $I_0 = 10^{-4}$ (1,000 as of mid March in Israel, assuming $\times 5$ undetected). Shown are the cases $R_0 = 2$ (high curves) and $R_0 = 1.1$ (low curves). A peak capacity of respiratory machines of 10^{-4} of the population (1,000 in Israel), and assuming optimistically that 1% of I (including a-symptomatic) need them, correspond to the critical line for medical-system collapse at $I \sim 10^{-2}$ (100,000 in Israel). With $R_0 = 2$, the critical line will be crossed in ~ 40 days, and will be surpassed by more than a factor 10 at the peak. R_0 has to be reduced to 1.1 for the peak to get below the critical line. Then the peak is delayed by more than a year, giving a chance for a vaccine to be developed.

taneously (e.g. 1,000 people out of 9 million in Israel). With 10% critical cases out of the infected, this means $I_m = 0.001$ infectious at the peak, which requires $R_0 \simeq 1.05$, and will take $t_m \sim 100\gamma^{-1} \sim 2 - 3$ years to get to. We better get used to live with Covid-19...

This is actually the case for a common flu, for which R_0 remained about 1.5 – 2 in recent years. This means that about 10% of the population is being infected every year. However, the death rate once infected by a flu is 0.001 compared to ~ 0.03 for COVID-19.

2.6 The final size

Using eq. (13), one can deduce the **final size** of the epidemic at $t \rightarrow \infty$, $Z_\infty = 1 - S_\infty$,

$$Z_\infty = -R_0^{-1} \ln S_\infty = -R_0^{-1} \ln(1 - Z_\infty). \quad (17)$$

This yields the **final-size formula** (Kermack and McKendrick 1927)

$$\boxed{Z_\infty = 1 - e^{-R_0 Z_\infty}}. \quad (18)$$

Again, with R_0 significantly larger than unity, the solution is with $Z_\infty \sim 1$, namely most of the population infected. As R_0 approaches unity the final size approaches $\ll 1$. The total fraction of deaths is a fraction of Z_∞ , say on average $0.03 Z_\infty$, and higher for youngsters older than 70...

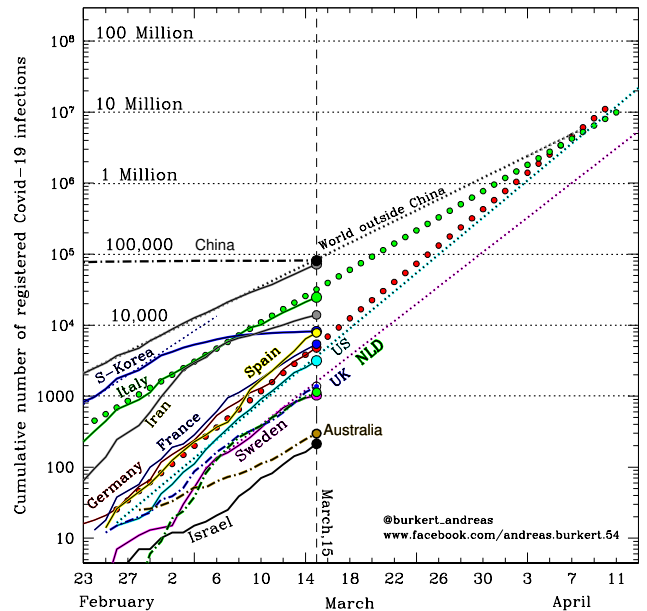


Figure 4. Empirical early evolution of Covid-19 (put together by Andi Burkert). As predicted by the model, we see an exponential growth (straight line in this semi-log plot) in each country, with different starting points but similar slopes. Interestingly, the behavior is similar worldwide. If the duration of the infectious period is $\gamma^{-1} \sim 10$ days, and the doubling time is ~ 3.5 days, then $R_0 \sim 3$.

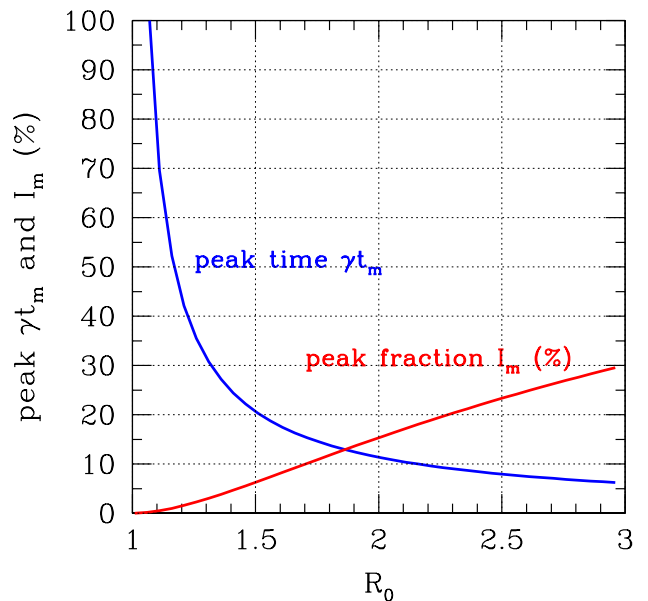


Figure 5. The peak height I_m (eq. (16)) and peak time t_m (numerical solution) as functions of R_0 . With $R_0 \sim 2-3$ the peak will involve a significant fraction of the population and lead to collapse of the medical system. For a sizeable reduction that will make it manageable, $R_0 = 1$ has to become unity or less. In parallel, a reduction of R_0 is associated with a delay in the peak time, implying a long-term epidemic as long as R_0 remains slightly above unity.

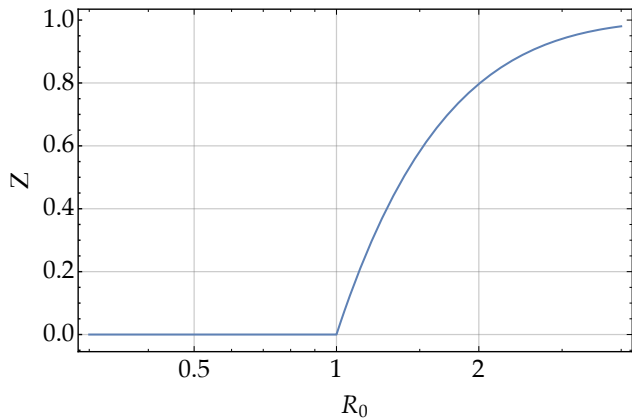


Figure 6. The final size of infected Z_∞ as a function of the basic reproduction number R_0 . This is a solution of eq. (18) or eq. (19). A value of R_0 of unity or less is required to avoid the infection of a large fraction of the population. Reducing R_0 to a value that is still above unity is not good enough for reducing Z_∞ to $\ll 1$, though it helps in flattening the curve, reducing and delaying the peak of I .

BTW, if Z_∞ is known, one can invert the final-size relation to explicitly estimate post-hoc what R_0 was,

$$R_0 = -Z_\infty^{-1} \ln(1 - Z_\infty). \quad (19)$$

Figure 6 shows the final size as a function of R_0 . Indeed, R_0 has to become very very close to unity or less for the final size to be much smaller than the population size.

2.7 Varying R_0

One can mimic the effect of reducing R_0 by connecting solutions where R_0 is constant for finite periods. Figure 7 shows a case where R_0 is reduced from 2.5 to unity within $5.5\gamma^{-1} \sim 55$ days, such that reaching the critical hospital capacity is avoided.

3 CONCLUSION

The first thing I learned, soon after Covid-19 started, is that it is going to be a disaster unless strict measures are taken to reduce R_0 to unity or below. It seems that now it became the common wisdom.

There are two obvious ways to reduce R_0 . One is to reduce the chances to be infected given a contact. This can be achieved by personal measures of washing hands frequently, not touching face, using masks, etc. The other is to reduce the numbers of contacts, first by isolation though closing borders and avoiding flights, and then all the way to a full quarantine. If this is taken very seriously, forcing severe confinements and ignoring the economic costs, it seems that reducing R_0 to unity may be doable. It seems that doing it drastically (a full quarantine) will shorten the epidemic and may eventually be less damaging to the economy.

One did not need the above model for coming up

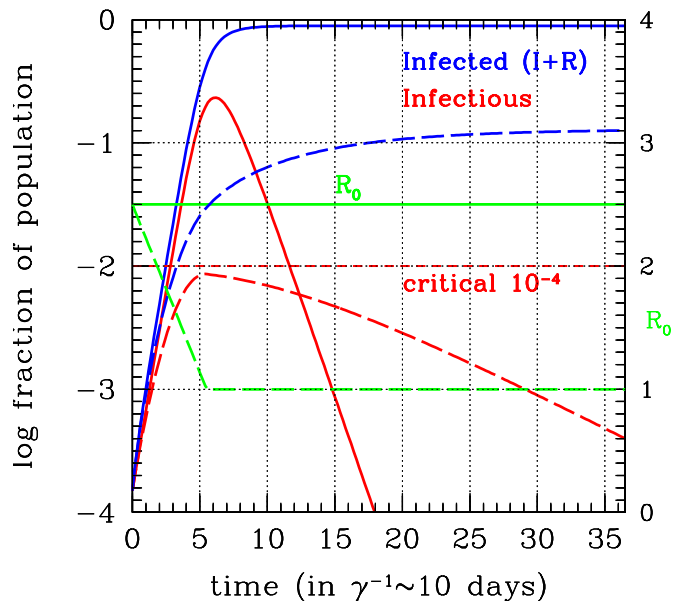


Figure 7. Compared to the standard case with $R_0 = 2.5$ (solid), here is a second case where R_0 starts at 2.5, and is forced to decrease to 1.0 at $t = 5.5\gamma^{-1} = 55$ days. With this kind of reduction, the infectious population turns around before reaching the critical hospital capacity, marked by the dashed line. It is the infectious fraction (including non-symptomatic), assuming that 1% of them need respiration. With respiratory machines available for 10^{-4} of the total population ($\sim 1,000$ in Israel), this is $I = 10^{-2}$ of the population infectious at the same time (100,000 in Israel).

with these obvious preventive steps. However, besides satisfying my curiosity about how an epidemic works, the severe prospects that emerge from the model with $R_0 \sim 2$ made me take this epidemic much more seriously than I would have otherwise, at a very early stage. Going to wash my hands now, and put myself in a voluntary home quarantine...

APPENDIX A: INTERACTION BETWEEN COMMUNITIES

The SIR model could be applied to multiple communities that interact with each other, such as countries with open borders (a generalization proposed by my colleagues Tvi and Zoe Piran). Each community is divided to three groups S_i , I_i and R_i , such that

$$S_i + I_i + R_i = 1. \quad (A1)$$

Each community has its own transmission rate β_{ii} , and each pair of communities has a cross transmission rate β_{ij} . The equations are

$$\dot{S}_i = -S_i \sum_j \beta_{ij} I_j, \quad (A2)$$

$$\dot{I}_i = S_i \sum_j \beta_{ij} I_j - \gamma I_i, \quad (A3)$$

$$\dot{R}_i = \gamma I_i. \quad (\text{A4})$$

Between countries the cross transmission rate is typically lower than in each country. One may consider alternatively subpopulations within a given community where the cross transmission rate is larger when involving a specific subpopulation, e.g., a large transmission rate for people who have frequent contacts with many others, such as bus drivers or supermarket cashiers.