

Supplementary material
Appendix: Model Documentation

Simulation-based Estimation of the Spread of COVID-19 in Iran

Navid Ghaffarzadegan (navidg@vt.edu) & Hazhir Rahmandad (hazhir@mit.edu)

The model, described in this paper, is the first and the simplest version of a series of models with different units of analysis, that the authors have been developing to examine the spread of the disease in different regions. As more information becomes available, and we learn more about the disease and infected cases in Iran, we continue the process of modification and calibration of the model.

A compartmental representation of patient flow

The model uses a country-level system of differential equations to capture the dynamics of contagion. Figure A1 shows the basic chain of population groups. State variables are shown as boxes, with flows between them represented explicitly. The infected population is first asymptomatic and later becomes symptomatic. The inflow to the asymptomatic population mostly comes from the susceptible population. However, as it is the case in new diseases, there are often first cases coming from other regions that already have a high prevalence. These first cases are critical for starting the epidemic, but once several seeds of infection are planted, later importation of cases is largely inconsequential to the overall trajectory (Chinazzi, Davis et al. 2020). The symptomatic infected population will follow two different paths of recovery or death. In practice different symptomatic subpopulations have different risks both in severity of illness and mortality, but to keep things simple we use and estimate general population averages.

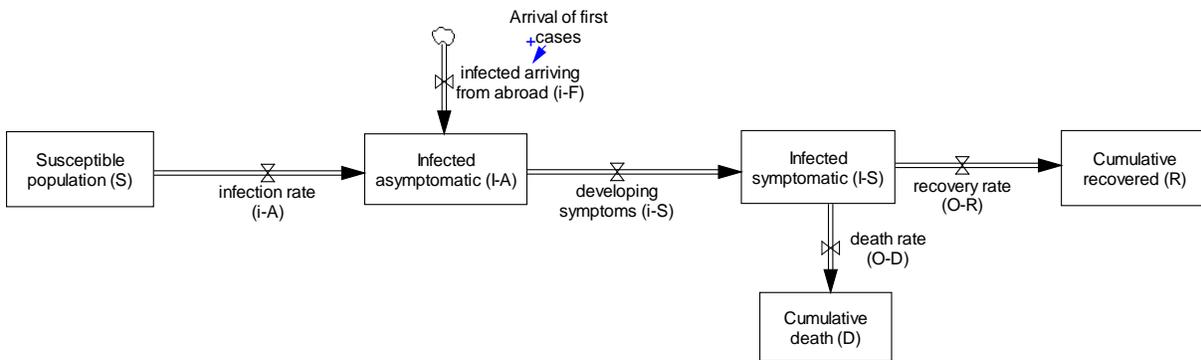


Figure A1-A compartmental representation of the flow patients

The following equations represent the Figure in a mathematical form:

$$S = \int_0^t (-i_A) dt + S_0$$

$$I_A = \int_0^t (i_A + i_F - i_S) dt + I_{A,0}$$

$$I_S = \int_0^t (i_S - O_R - O_D) dt + I_{S,0}$$

$$R = \int_0^t O_R dt + RE_0$$

$$D = \int_0^t O_D dt + D_0$$

where, S : Susceptible Population [person]; i_A : Infection rate [person/day]; S_0 : Initial value for S [person]; I_A : Infected Asymptomatic [person]; i_F : Infected arriving from abroad at time t_0 [person/day]; i_S : Developing symptoms [person/day]; $I_{A,0}$: Initial value for I_A [person]; I_S : Infected symptomatic [person]; O_R : Recovery rate [person/day]; O_D : Death rate [person/day]; $I_{S,0}$: Initial value for I_S ; R : Cumulative recovered [person]; RE_0 : Initial value for R [person]; D : Cumulative death [person]; D_0 : Initial value for D .

As the number of infected cases increase, the probability of infection increases given the increased number of contacts between infected and susceptible populations. Some of these contacts lead to new cases of infection. Figure A2 represents these contagion feedback loops (R1 and R2), indicating the reinforcing nature of the feedback that creates the early exponential growth in the number of cases. Infections may be due to a fraction of asymptomatic cases who can infect others (Bai, Yao et al. 2020), and the symptomatic population. The number of contacts of susceptible is proportional to contact rate and number of susceptible population. Death and recovery rates are estimated using fractional death rate. A simplifying assumption is that on average the time to death and time to recovery are similar.

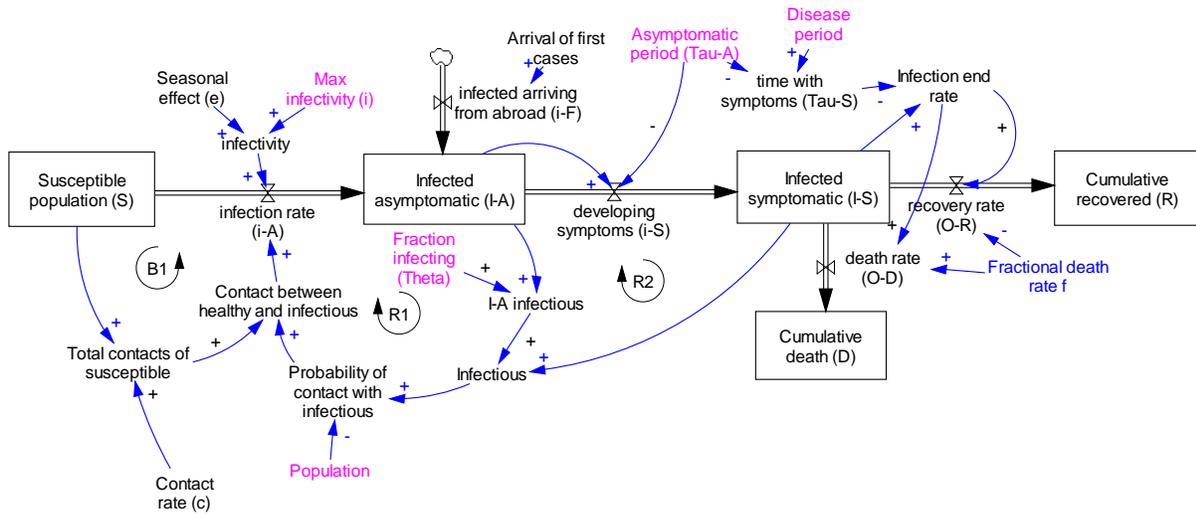


Figure A2-Basic dynamics of contagion, depletion, recovery and death

Infection rate (i_A) in the diagram of Figure A2 is formulated as following:

$$i_A = cS \left(\frac{\theta I_A + I_S}{S_0} \right) ie$$

where c : contact rate [1/day]; θ : fraction of I_A who are infectious; i : infectivity [Dimensionless]; e : seasonal effect [Dimensionless].

We set e at 1 (no seasonal effect) for historical calibration, and later for scenario analysis, check different functions that give lower values during summer, qualitatively consistent with influenza

infection statistics. The parameter θ captures the relative infectivity of asymptomatic patients compared to those with symptoms. The equation has two degrees of freedom, and since the actual value of c and i are not independently knowable, we assume $i = 0.02$ and vary c to fit the model output against data (details below). In case one needs to explore effects of quarantine, in the equation above I_S should be replaced by $(1 - q)I_S$, q representing the fraction in quarantine. Alternatively such quarantine effects could be combined into a single fractional scaler that down regulates contact rate in response to perceived risk, which is the approach we take in quantifying various behavioral responses including reduced social contacts, better hygiene, and self isolation. Explicit policy measures incorporating quarantine would benefit from additional structure.

Furthermore,

$$i_S = I_A/\tau_A$$

$$O_R = (1 - f)I_S/\tau_S$$

$$O_D = fI_S/\tau_S$$

where the disease duration is assumed to be total of τ_A (asymptomatic period) plus τ_S (time with symptoms before recovery or death); O_R : Recovery rate [person/day]; O_D : Death rate [person/day]; f : Fractional death rate [Dimensionless].

Based on epidemiological reports on the disease (Novel Coronavirus Pneumonia Emergency Response Epidemiology 2020, Wu and McGoogan 2020) we assume the following parameters: We assume the asymptomatic period lasts for 4 days (confidence intervals in prior estimates is 2 to 7 days) and the time with symptoms to last for 10 days. Given detailed case data we do not change these parameters to fit our model to data. Fractional death rate (f) is estimated through model calibration given the wide range of uncertainty in prior reports of this parameter.

Endogenous representation of public response to the outbreak

Contact rate can change over time in response to the progression of the disease and consequently change infection rate. The impact of perceived risk on contact rate captures not only the endogenous changes in social interactions, gatherings, self-isolation of suspected cases, and hygiene, but also government mandated closure of events, schools, and businesses to name a few. For example in Iran schools and sport and cultural events were closed down after the seriousness of the epidemic was recognized in late February. Moreover people started to reduce their discretionary interactions in public spaces, cut down on social and family gatherings, and followed hand washing and mask wearing recommendations more closely. We assume public risk perception depends on reported cases of death.

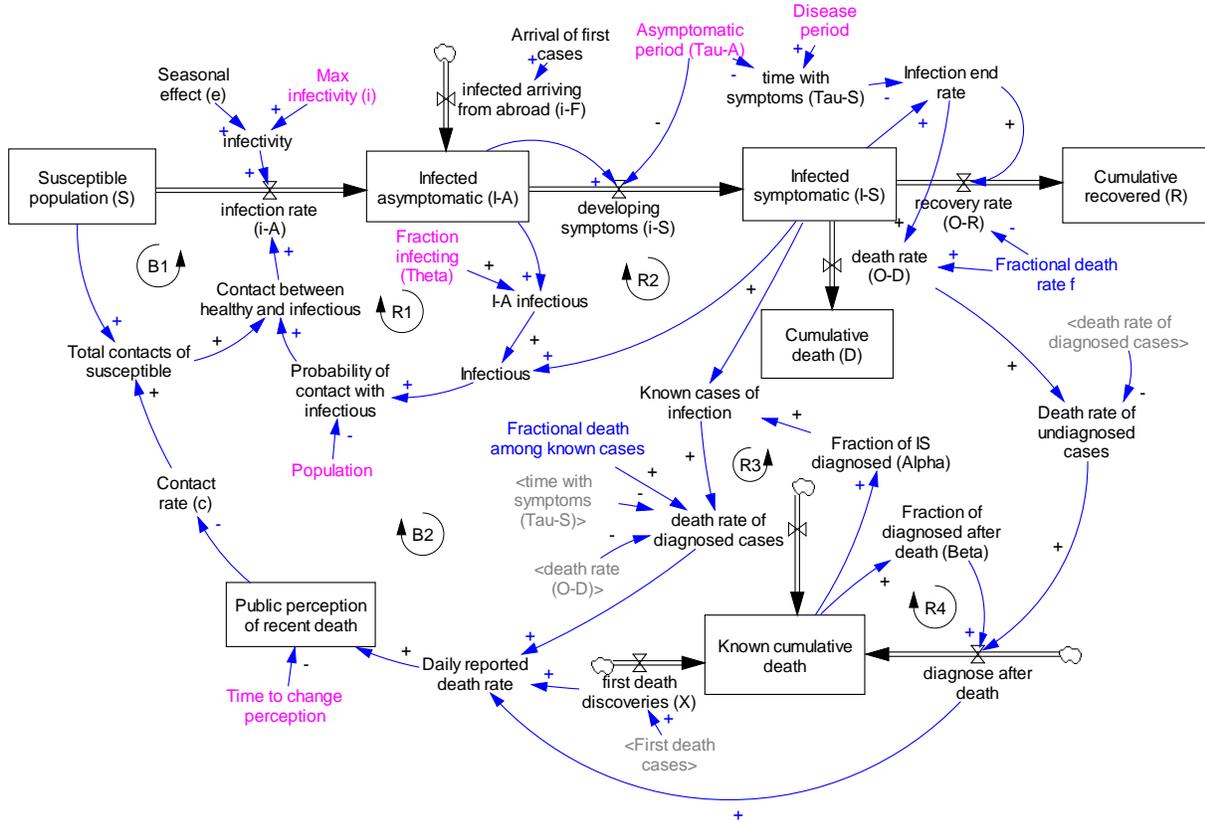


Figure A3-The main structure of the model.

With a more transparent government and one that has invested quickly in testing infrastructure, a higher fraction of known cases and deaths are reported to the public. However, finding infected cases is also challenging given that in the case of coronavirus, many cases remain mild and unknown, yet potentially infectious. Furthermore, especially early on, many cases were misdiagnosed as healthcare providers were unaware of the existence of the virus in the country. As the first cases are revealed and reported, a higher fraction of cases are found (Loops R3 and R4).

There are three death-rate-related variables in this model: actual death rate (death per day, O_D), official reports of death rate ($\overline{O_D}$), and public perception of death rate ($\overline{\overline{O_D}}$). In our chain of causality, O_D influences $\overline{O_D}$ (see below for details). Then $\overline{O_D}$ influences $\overline{\overline{O_D}}$, which in turn influences average contact rate (c). $\overline{\overline{O_D}}$ is represented as a $\tau = 7$ days first-order lagged variable of $\overline{O_D}$.

Average contact rate is inversely related to public perception of death rate. Mathematically, $c = c(\overline{\overline{O_D}})$, $\frac{\partial c}{\partial \overline{\overline{O_D}}} < 0$. We bound c between 0 and c_{max} . We use the following sigmoid function to represent change in contact rate:

$$c = \rho c_{max} \left(1 - \frac{1}{1 + e^{s(\sigma_D^* - \overline{\overline{O_D}})}} \right)$$

The parameter ρ represents the impact of other policy measures beyond the perception of death that may impact contact rate (e.g. various government interventions) and is set to 1 for base run simulations but experimented with for policy analysis. The free parameters, O_D^* , s , and c_{max} , are estimated.

Relations between official report and actual death

We assume only symptomatic cases of patients can be diagnosed. Let α be the fraction of diagnosed cases, so that αI_S is the number of diagnosed infected. More severe cases are more likely to be diagnosed, while milder cases are more likely to be missed, so mortality rate is likely to be higher among diagnosed individuals. We therefore use a different mortality parameter of f' for this population than the undiagnosed population of patients. Thus death rate of diagnosed individuals is $f' \alpha I_S / \tau_S$, with the constraint $f' \alpha \leq f$. Missed cases of death will be $(f - f' \alpha)(I_S / \tau_S)$, and we assume a β fraction of them will be diagnosed after death. Including first cases of death discovery diagnosed early on which trigger the start of government and public responses (X):

$$\begin{aligned} \overline{O}_D &= X + f' \alpha (I_S / \tau_S) + \beta (f - f' \alpha) (I_S / \tau_S), & f' \alpha &\leq f \\ \overline{D} &= \int_0^t \overline{O}_D dt + \overline{D}_0, & \overline{D}_0 &= 0 \end{aligned}$$

where \overline{O}_D is reported cases of death per day, and \overline{D} is the cumulative reported cases of death. X only includes the first few cases at t_1 , the time first death cases are reported. This factor is equal to 2 cases per day on February 19th and for the rest of the simulation it is zero. For the fractions, α and β , we can either assume constant values or give more degrees of freedom to explore a wide range of values and functions that change over time. We take a more conservative approach and let the fractions change as more cases of death are reported, and the country starts feeling the state of emergency. Specifically, we set $\alpha = \text{Min}(a\overline{D}, 1) \cdot \alpha_{max}$ and $\beta = \text{Min}(b\overline{D}, 1) \cdot \beta_{max}$, and let the model estimate α_{max} , β_{max} , a , and b through calibration (explained below). A limitation of this simple formulation is that, after α and β reach their maximum values, \overline{O}_D can increase rapidly (with O_D), potentially outpacing the existing testing infrastructure. In practice when the testing infrastructure is a major bottleneck, official death data may fall increasingly below the true death. On the other hand the model may estimate small values for α_{max} , β_{max} which would implicitly capture these bottlenecks in testing.

For reported cases of recovered, we use the following equations:

$$\begin{aligned} \overline{O}_R &= (1 - f') (\alpha I_S / \tau_S), & \overline{O}_R &\leq O_R \\ \overline{R} &= \int_0^t \overline{O}_R dt + \overline{R}_0, & \overline{R}_0 &= 0 \end{aligned}$$

where \overline{O}_R is reported cases of recovery per day, O_R is the actual recovery rate, and \overline{R} is the cumulative reported cases of recovery.

For reported cumulative cases of infection (I_{cum}) we have:

$$\begin{aligned} I &= I_A + I_S \\ I_{cum} &= I + D + R \end{aligned}$$

$$\overline{I_{cum}} = \alpha I_s + \overline{D} + \overline{R}$$

where I is total infected at any time period. Note that this formulation assumes the reporting of recovered cases happens with the same rate (α) as the reporting of infections and death. While Iran's data is supportive of this assumption, exploration of statistics from other countries suggests many do not track the recovered cases and thus the sum of recovered and dead may not equal the total infections more generally.

Unofficial data

Besides the official data, we use a few unofficial data points including three observations about number of Iranian passengers diagnosed with COVID-19 upon arrival in international airports (\overline{P}), and four unofficial estimations from BBC and IranInetrnational news sources about total cases of death from COVID 19 (\overline{D}_u) early in the epidemic.

In our model, we estimate \overline{P} and \overline{D}_u as following:

$$\begin{aligned}\overline{P} &= \mu(t + 1 - t_1)I_{cum} \\ \overline{D}_u &= \int_{t_1}^t \gamma O_D dt + \overline{D}_{u,0}, \quad \overline{D}_{u,0} = 0\end{aligned}$$

Where \overline{P} is cumulative number of diagnosed passengers from t_1 to t , and μ is the ratio of number of passengers going from Iran abroad in early days of the outbreak. The term $(t + 1 - t_1)$ represents days since the first cases of infection were diagnosed. The assumption is that the daily number of passengers is fixed and death is minimum, thus this estimation only works during the first few days of the outbreak. In addition, it is assumed that passenger cases are reliable sources and infection in airport or during the flight is not the source of identified cases in foreign airports. For the other source, \overline{D}_u , which is the unofficial reports for death cases, we assume γ represents the fraction of actual deaths identified by these sources. The parameter γ is likely below one, representing the underestimation inherent in any count of deaths. The parameter γ will be estimated through model calibration.

Model calibration

We initialize the model setting all initial values of stocks of patients at zero and the susceptible population at total population of Iran; $I_{A,0} = I_{S,0} = R_0 = D_0$, and $S_0 = 8.1$ million people. We start with no observed cases of recovery or death; $\overline{R}_0 = \overline{D}_0 = 0$. The model will start as new infected cases are injected to the model through i_F , at time t_0 , to be estimated through model calibration. Although we are not aware of number of foreign patients entering Iran, since t_0 is also unknown, we have two degrees of freedom, and one can be assumed and the other estimated through model calibration. We set i_F equal to 100 patients per day at t_0 , and 0 throughout the rest of the simulation. We run the model from $t = 0$, December 31, 2019, and our time unit is a day. t_1 , the time first cases are diagnosed, is February 19, 2020 ($t_1 = 50$).

We have time series data for \overline{D} , \overline{R} , $\overline{I_{cum}}$, and a few data points for \overline{P} and \overline{D}_u which can especially help during the early phases of simulation. Our method is mainly based on forming a likelihood function for observing the actual time series data conditional on model parameters. We then conduct a Markov Chain Monte Carlo (MCMC) simulation to estimate the joint posterior

distribution of the model parameters subject to observed data. We define a likelihood function for change over time (net-flow) of \bar{D} , \bar{R} , \bar{I}_{cum} assuming they are count events drawn from model-predicted rates (Poisson distribution). We use a similar Poisson distribution assumption for \bar{P} and \bar{D}_u as well, since they both fit well into a count measure framework. The MCMC method searches over the feasible ranges for eleven uncertain parameters in our model. These include: five data-related parameters of α_{max} , β_{max} , a , b , and γ ; three public-behavior-related parameters of maximum contact rate (c_{max}), death threshold (O_D^*), behavior sensitivity to reported death (s); and three disease-related parameters of fractional death rate (f and f') and time for the injection of first cases of infection into the population (t_0). Prior experience and this exercise point to rather tight confidence intervals coming from MCMC methods directly applied to large nonlinear models. The problem could be addressed by the use of filtering (e.g. extended Kalman filtering or particle filter) methods (Ionides, Breto et al. 2006, Li, Pei et al. 2020) or heuristics that scale the likelihood function in MCMC method to be consistent with model's predictive performance. In light of the computational costs of the former and time sensitivity of the research topic, we chose the latter option. To be conservative, we therefore downgrade our confidence in the data, scaling the likelihood function by a factor of 0.1, which allows the algorithm to explore the parameter space much more freely. This approach is a heuristic which we use to avoid putting too much confidence in our estimation results; however, without strong theoretical reasons to support the specific scaling factor, we caution against a strong interpretation of these results. With this caveat in mind, the MCMC method offers 90% confidence intervals for joint-distribution of the parameters. The rest of the model parameters are specified based on current literature. Table A1 lists parameter values and their sources.

Table A1: Model parameter values. Standard deviations for estimated parameters are reported in parentheses.

Parameters	Estimated value	Source
α_{max} : Maximum cap on percentage of patients tested	0.025 (0.009)	Model calibration
β_{max} : Maximum cap on percentage of undiagnosed death, diagnosed after death.	0.256 (0.177)	Model calibration
a : The slope of change in test of infected.	0.022(0.002)	Model calibration
b : The slope of change in test of death cases.	0.005 (0.003)	Model calibration
γ : Fraction of true death captured in unofficial death reports	0.418 (0.152)	Model calibration
s : Behavior sensitivity to reported death	0.028 (0.006)	Model calibration
f : Fractional death rate	0.037 (0.004)	Model calibration
f' : Fractional death rate among diagnosed	0.141 (0.013)	Model calibration
c_{max} : Maximum cap on contact rate	22.1 (3.5)	Model calibration
t_0 : Injection time of first 100 cases	10.116 (3.177)	Model calibration
O_D^* : Death threshold	22.8 (10.9)	Model calibration
τ_A : asymptomatic period	4 days	Estimated based on different sources such as (Li, Pei et al. 2020)
τ_S : Time with symptoms	10 days	Estimated based on different sources such as (Novel Coronavirus Pneumonia Emergency Response Epidemiology 2020, Wu and McGoogan 2020)
i : Infectivity base value	0.02	Infection rate is multiplied by contact rate so we can assume a value for i and estimate c_{max} without loss of generality
θ : fraction of IA who are infective	0.25	Estimated based on last day of a 4-day incubation period.

τ : time to change public perception	7 days	Assumption
μ : the ratio of daily international passengers to the total population during first few days	7.4e-4 (1/day)	Based on the assumption of about 6000 travels per day.
ρ : Contact multiplier from extra policy actions	1	Can change as a policy parameter between 0 and 1; 0 is maximum pressure.

Limitations of our model

This study was conducted under rapidly changing conditions to inform emerging understanding of an epidemic to help policy makers and interested individuals. As such we had to make significant tradeoffs and simplifying assumptions, which are listed here to underscore the uncertainty in the following analysis. Where in doubt we erred towards under-counting the actual cases to avoid promoting undue fear, and thus our results may be closer to lower bounds for the magnitude of the epidemic in Iran.

In setting our model boundary we focused on capturing some of the behavioral feedbacks that we think have a stronger impact on the outcomes (e.g. human reactions to the changes in the system, delays, and feedback loops) and left out detail complexities (such as age and risk factor distributions), spatial complexities (such as diffusion patterns in different cities), and network complexities (such as effects of social network structures and the role of most connected nodes). Ignoring heterogeneities in population groups and interactions may limit confidence in our results though, based on prior comparative work, we think those limits are unlikely to change the qualitative findings (Rahmandad and Sterman 2008).

A few other simplifying assumptions are noteworthy. Regarding healthcare resources, as the burden of disease increases it is unclear to what extent the healthcare system can manage the flow of patients, and how that would affect survival of infected cases. Currently our model assumes those issues will not change the dynamics in a major way, an admittedly simple assumption one should question in a more detailed model. Furthermore, healthcare facilities can themselves result in new cases of infection by infecting individuals seeking diagnostic tests. This iatrogenic transmission is ignored in our model. We also do not consider the fact that the burden of the disease affects healthcare workforce productivity. The high workload, stress, and risk of infection and death among healthcare staff is considerable and that can affect healthcare system performance as well as infections and mortality among this sector.

Other major assumptions are that, first, the model is scoped only around Iran -- we ignore potential effects of global spread of the disease on Iran. Second, we assumed that the recovered population are immune during our simulation time period of four months. While not established for COVID-19, this is usually the case for infectious diseases where the infected individual recovers though immune response of the body. Third, we ignore mutations in the virus which may change its contagiousness and case fatality rate. Finally, and potentially most significantly, we assumed rather rapid and strong behavioral responses to the epidemic are feasible which could bring down contact rate enough to contain the epidemic. While our calibration estimates the slope of such responses, we still do not have enough data to know how far those responses can go. We also do not consider normalization of the disease (and thus loosening of the response) or the reintroduction of new cases from other countries in future. Overall these assumptions err on the side of being conservative about the epidemic's growth to avoid unnecessary fear. As such we may offer an optimistic view of the epidemic, but

one that is still much worse than the official figures. With these caveats in mind, the next section summarizes our results. Our model, in Vensim DSS, is available for further improvement and testing.