medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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.

Characterization of Israeli COVID-19 Outbreak Drivers and Forecasting Using a Versatile Web App

Wayne M. Getz*[,1,2,3], Richard Salter*[,3,4], Nir Horvitz[1], Ludovica Luisa Vissat[1]

[1]Dept. ESPM, UC Berkeley, CA 94720-3114, USA
[2]School of Mathematical Sciences, University of KwaZulu-Natal, Durban 4000, South Africa
[3]Numerus, 850 Iron Point Rd., Folsom, CA 95630, USA
[4]Computer Science Dept., Oberlin College, Oberlin, Ohio, OH 44074, USA

*Equal co-first authors
Corresponding author: Wayne M. Getz, wgetz@berkeley.edu

Preprint: August 9, 2020

Summary

Background No versatile web app exists that allows epidemiologists and managers around the world to fully analyze the impacts of COVID-19 mitigation. The NMB-DASA web app presented here fills this gap.

Methods Our web app uses a model that explicitly identifies a contact class of individuals, symptomatic and asymptomatic classes and a parallel set of response class, subject to lower contact pathogen contact rates. The user inputs a CSV file containing incidence and mortality time series. A default set of parameters is available that can be overwritten through input or online entry, and a subset of these can be fitted to the model using an MLE algorithm. The end of model-fitting and forecasting intervals are specifiable and changes to parameters allows counterfactual and forecasted scenarios to be explored.

Findings We illustrate the app in the context of the current COVID-19 outbreak in Israel, which can be divided into four distinct phases: an initial outbreak; a social distancing, a social relaxation, and a second wave mitigation phase. Our projections beyond the relaxation phase indicate that an 85% drop in social relaxation rates are needed just to stabilize the current incidence rate and that at least a 95% drop is needed to quell the outbreak.

Interpretation Our analysis uses only incidence and mortality rates. In the hands of policy makers and health officers, we believe our web app provides an invaluable tool for evaluating the impacts of different outbreak mitigation policies and measures.

Funding This research was funded by NSF Grant 2032264.

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/10.1101/2020.08.06.20155804; this version posted August 11, 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.

Introduction

The COVID-19 outbreak that in February, 2020, threatened to overwhelm the healthcare system of Wuhan, China, was brought under control in early March through a combination of social distancing, contact tracing and quarantine measures [1–4]. Since then, these measures have been used by municipalities, counties, states and countries around the world to bring COVID-19 outbreaks around the world under full or partial control. Many of the partially controlled cases have subsequently experienced stronger breakouts than before, as control measures—which we refer to as drivers—have been prematurely relaxed leading to a second wave of outbreaks, many larger than the first.

As of late July, 2020, the pandemic state of COVID-19 was closing in on twenty million recorded cases, two-thirds of a million record deaths, and surging or resurging outbreaks. In the US itself, more than twenty states had experienced outbreaks of more then fifty thousand recorded cases. Thus, healthcare officials and civic leaders at various administrative levels worldwide are in dire need of quantitative tools to help formulate and implement policies designed to flatten and, ultimately extinguish, COVID-19 outbreak curves. Such tools exist only in the hands of computational epidemiological modeling and data analysis groups. These groups are insufficient to meet the policy design and implication needs of most communities around the world. Without such tools, governors, mayors, and other civic leaders responsible for implementing healthcare policy are flying blind when making critical decisions whether or not to open schools and various business sectors of communities representing different levels of risk with regard to transmission of the SARS-CoV-2 pathogen.

Here we present a Web based analytical tool, COVID-19 NMB-DASA (Numerus Model Builder Data and Simulation Analysis), that can be used to address questions regarding the impact of social distancing, social relaxation, changes in surveillance, implementation of contact tracing with quarantining, patient isolation, and vaccination (when widely available) on incidence and mortality rates. NMB-DASA does not require the user to have a mathematical or epidemiological modeling background or an understanding of the computational procedures needed to carry out deterministic and stochastic simulations for model parameter estimation or epidemic forecasting. It only requires users to provide a comma separated values file (CSV: a standard used by all common spreadsheet applications and data management programs) that contains the incidence and mortality time series of the particular outbreak to be analysed. The user is also able to use a set of default parameters that come with our COVID-19 NMB-DASA, or the user can modify these by either entering new values in data panels or manipulating sliders on web pages.

We illustrate the basic aspects of how NMB-DASA can be used to better understand dynamic aspects of outbreaks, extract effects that implicit drivers, such as social distancing and social relaxation, have played in producing observed incidence and mortality patterns. We also demonstrate how NMB-DASA can be used to forecast the impacts of changes to drivers that can then guide policy makers on the degree to which social distancing measures need to be implemented or hospitals need to gear up [5]. We do this in the context of data available for the Israeli COVID-19 outbreak. This outbreak is particularly suitable for investigating the impacts of social distancing in breaking an epidemic under partial control, with subsequent social relaxing leading to an outbreak rebound that is larger than the initial outbreak spike.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

Methods

Model

The epidemiological simulation and forecasting engine of our website is based on an extended SEIR (Susceptible, Exposed, Infectious, Recovered) formulation called SCLAIV (Fig. 1), which expands the E disease class into a contact class (C) and a latent class (L) [6]. Individuals in C, having made contact with the pathogen responsible for the disease in questions (COVID-19 in this paper), can either thwart a potential pathogen infection by returning to S or succumb to pathogen infection by moving onto L. SCLAIV also adds an asymptomatic class (A) to the infectious component of an SEIR process, where individuals in this class are assumed to be less infectious than in the class I and can either move to I or directly to the naturally immunized (i.e., naturally vaccinated) class V. A SCLAIV+D formulation also explicitly separates out the removed class R in the SEIR process into V and the class D of individuals that have died.

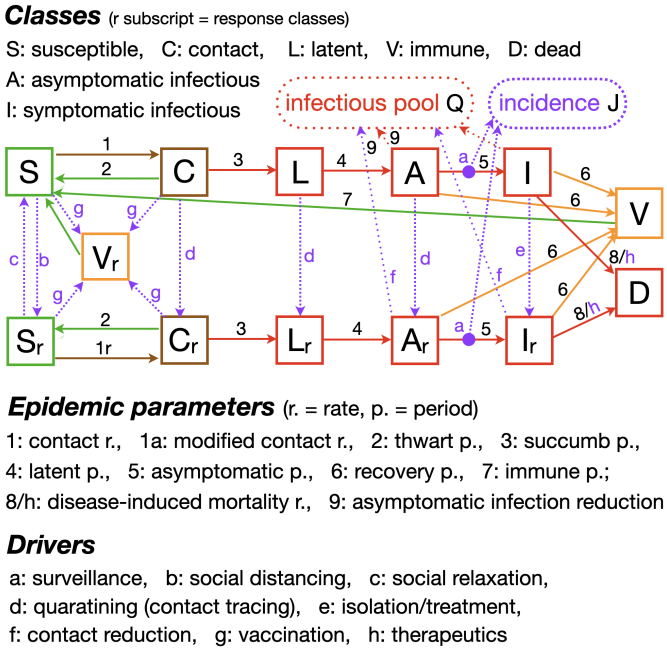

Figure 1: A flow diagram of the SCLAIV+D+response model with its 8 flow parameters identified by Roman numbers 1-8 plus an asymptotic infection parameter 9. The 8 drivers identified by lower case Roman letters a-h are either zero (apart from surveillance), a positive constant or have the form of a switching function, as described in the text. We have generalized the disease-induced natural mortality flow rate 8 from being a constant [6] to include the possibility of driving (h) the rate down to account for improvements over time in treatment and therapeutics.

Basic epidemiological, whether SEIR or SCLAIV+D, need to be modified to include the driving process used to "flatten the incidence and mortality curves" during outbreaks. A discrete time deterministic/stochastic formulation of our SCLAIV+D+response model [6], at the computational core of our web app, includes 7 response drivers (Fig. 1), each of which can either be specified as a constant or time varying rate—the latter, specifically, as a five parameter switching function (e.g., see specific cases in Fig. 8. These driving rates are responsible for transferring individuals among

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

basic SCLAIV and SCLAIV-response classes and monitoring the state of the outbreak. Specifically, these drivers are a social distancing rate, which involves a possible dynamic change in a contact rate reduction parameter, a social relaxation rate, a quarantining rate linked to contact tracing, case isolation/treatment and vaccination rates, and the level of surveillance (i.e., proportion of cases detected) implement to monitor the state of the outbreak.

Model parameters and data

Many, but not all, of the basic epidemiological parameters used to model COVID-19 and other infectious disease outbreaks can be obtained independently from data on the case histories individuals infected with SARS-CoV-2 [3, 7–9]. These parameters include the latent, asymptomatic, recovery (i.e., infectious), and immune periods (the latter can only be estimated from future data, depending on the mean period of acquired immunity that will be realized over time).

The disease induced mortality rate can also be obtained directly from epidemiological data and the average mortality rate over any period of time can be estimated from the number cases observed and the ultimate fate of those cases: recovery or death. During the course of an outbreak, in the context of COVID-19 however, the time between an individual being infected with SARS-Cov-2 and a definitive outcome could be several weeks. So this time lag has to be taken into account when estimating mortality [9]. Further, during the course of an outbreak, provided the healthcare system is not overwhelmed we might expect mortality rates to drop as hospitals learn how to take care of the more serious cases admitted for treatment and care.

Parameters that are difficult to explicitly identify from case-history data are absolute contact rates: how to define a viable or effective contact in terms of proximity, duration, and background environmental conditions (e.g., humidity, temperature, air movement, fomite characteristics [10] and so on). Also difficult to obtain is the state of the population at the time that the first few cases are detected. The first cases are the first set of individuals identified in disease class I at the nominal start of modeling the outbreak, which for convenience we denote as time $t = 0$. This point in time, of course, we anchor to some suitable calendar day. The number of individuals assigned the value $I(0)$ (italic fonts are use to refer to the variables that represent the number of individuals in roman font named classes—so $I(t)$ is the number of individuals in class I at time t) depends on the level of surveillance in effect at time t. In a well-developed health care system, depending on the proportion of asymptomatic carriers and the severity of the disease, half the cases may easily go undetected in the initial outbreak, though surveillance should improve during the course of the outbreak. In a poorly-developed healthcare system, the proportion of cases detected during the initial stage of an outbreak may be several fold lower.

Of course, only symptomatic cases will initially be detected: asymptomatic will require the implementation of some sort of testing regime. Thus, at the start of an epidemic, the number of individuals $C(0)$, $L(0)$ and $A(0)$ is unknown and also needs to be estimated. The latter two can be expressed in terms of the estimated $I(0)$ if the initial growth rate of the epidemic can be reasonably guessed at from past experience with similar epidemics, and thus may computed from the estimate of $I(0)$ [6] rather than having to be estimated as well. The initial value for $C(0)$, however, also needs to be estimate since its relationship to $I(0)$ depends on the various other parameters, such as the thwart and succumb periods, one of which also need to be estimated, while the other can be set to 1: viz., as a first cut we recommend setting the thwart period to 1, if the units of t are days, and relaxing this constraint later if needs be [6]. The initial value V_0 for the number of individuals

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

assumed to be immune to infection by SARS-Cov-2 at the start of the outbreak was assumed to be 0 because this virus is thought to be novel for humans. Of course, it may emerge that some cross immunity exists for individuals that have previously been exposed to SARS-Cov-1 or other corona viruses [11], and then either some suitable non-zero value for $V(0)$ or we can think of the population at risk, as discussed below, at some appropriately smaller value than the total population size.

Initially, two of the drivers—surveillance and patient isolation rates—can be set at positive constant values, based on a crude expectation of the current level of efficiency of the healthcare infrastructure. For a developed country, we may want to set initial surveillance levels at, say, 0.5 (50%) and hospitalization levels at, say, 0.35, or some other values that are more reflective of the severity or mildness of the disease. For countries with poor health infrastructure, or diseases milder than COVID-19 values of 0.1 for both this levels may be more suitable. These guesses can be changed later, as the forensic analysis of the structure of the outbreak proceeds, and the sensitivity of projected outcomes to these initial guesses are evaluated.

Another parameter that remains somewhat uncertain at the start of an outbreak is the size N of the population at risk. In a isolated, spatially homogeneous population (e.g. an island where all individuals regularly encounter others from different parts of the island—that is, the population is well mixed at, say, a weekly scale), N should be the size of the population. In large metropolitan areas where the boundaries separating the well-mixed portion of the population from outlying areas we may want to set the value of N to some nominal value N_{nom} that is, say, 1 million or even 10 million when the population size justifies this choice, since the initial outbreak dynamics in the SCLAIV model is insensitive to the value of N_{nom} [12]. As the outbreak progresses and the total number of infections approaches, say 10% of N_{nom}, then our choice for this value becomes significant [6] since it determines the point at which the exponential phase of the basic outbreak (i.e., the outbreak before social distancing and other mitigative measures are applied) starts to flatten and, ultimately, crest and turn down. At this stage of the analysis, it may be useful to replace N_{nom} with an "effective population size" N_{eff}, which itself is estimated in the process of fitting model output to incidence data [12].

Fitting the model to incidence data

NMB-DASA uses a maximum-likelihood estimation (MLE) method to fit selected parameters to the incidence and simulteneously or separately to the mortality data. It also provides for parameter values obtained from Bayesian Markov chain Monte Carlo (MCMC) methods [13, 14] generated elsewhere to be used in forecasting simulations. We plan to include MCMC directly in future versions of NMB-DASA

The following step-wise heuristic procedure finds excellent MLE fits to the Israeli COVID-19 case study considered in the next section.

Stepwise procedure for running and fitting the SCLAIV model

1. *Load Data.* Load the incidence and mortality data using a CSV (comma separated values) file in which the first row is an incidence data time series and the second row is a mortality time series of the same length as the first 2. Load the default set of epidemic and driver parameters. If any of the default parameters require modification, these can be overwritten by directly entering data into the appropriate window and then saved by selecting the `SAVE SETTINGS` button for direct loading next time around.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

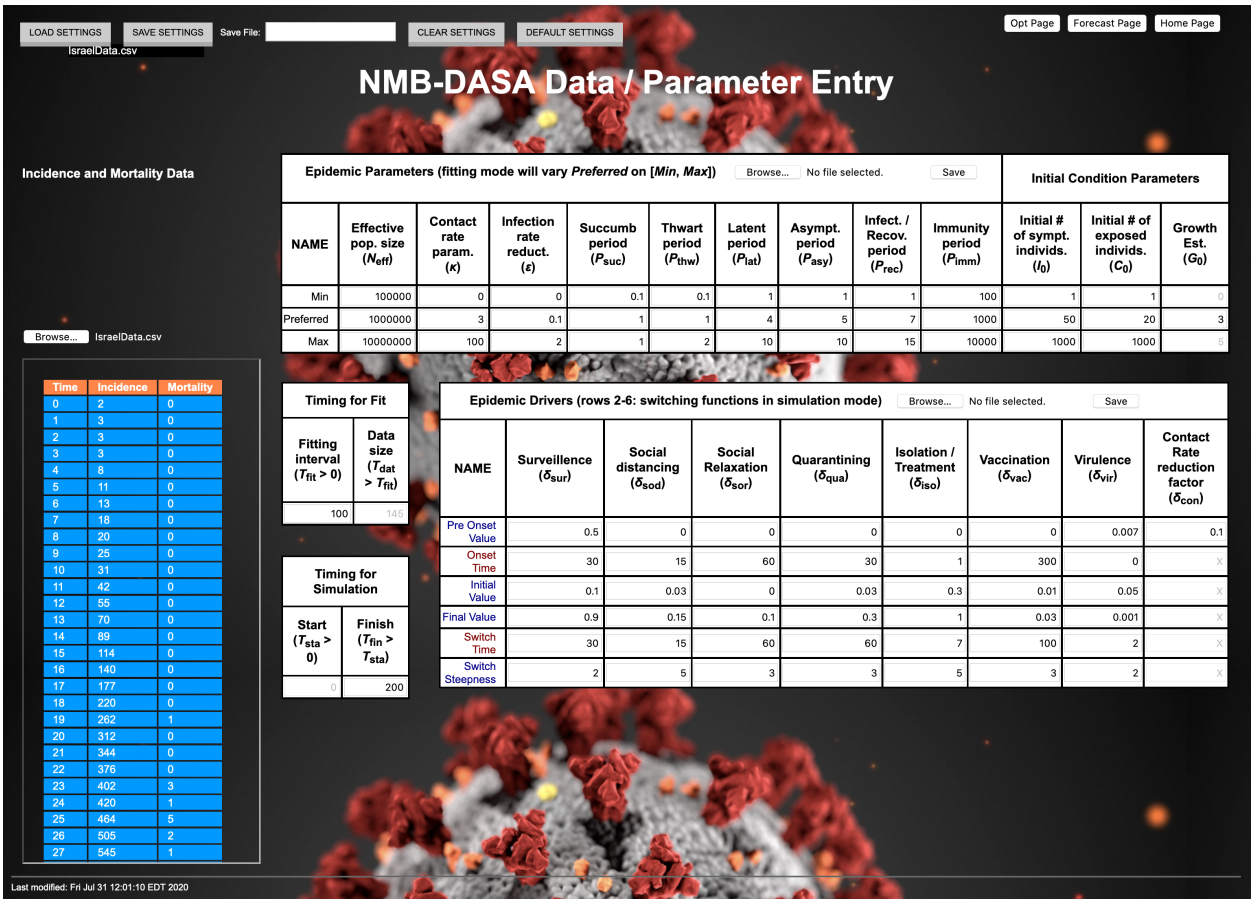

Figure 2: Incidence and Mortality time series are loaded by selecting a CSV file (our `IsraelData.csv` file is listed in the `Choose File` window just above the data in the blue cells on the left). These data along with `Epidemic Parameters` and `Epidemic Drivers` values can be loaded separately or default values can be loaded (which are the values shown here), and values can then be modified by hand and saved under some suitable name. Although the various types of data can be saved separately, they can all be saved together as a single .dat file using the `SAVE SETTINGS` button at the top left and loaded later (after selecting `CLEAR SETTINGS` using the `LOAD SETTINGS` button. The `Fitting interval` value has been set at 100, and the `Data size` value 145 is automatically computed from the length of `Incidence` and `Mortality` time series. Once all the parameter values have been set, the `Opt Page` button on the right can be selected to take the user to the NMB-DASA optimization page. A forecasting simulation `Finish` time can be inserted, 200 as shown here, or selected later on the `Forecast Page.`

2. *Fit Initial Outbreak Phase.* Fit those basic SCLAIV epidemic parameters that cannot be independently estimated by checking them off on the optimization page (purple panel in Fig. 3), setting `Wt:incidence/deaths` to 1, `Fit Range` to some desired value, such as 14 or 30, and then pressing the `Opt` button (pink panel in Fig. 3). These parameters are likely to be the contact rate (κ in [6]), the succumb period (P_{suc} denoted by π_{suc} in [6]), the initial value (C_0) for the contact class, and the initial value (I_0) for the symptomatic infectious class. Note, since the ratio $P_{\text{suc}}/P_{\text{thw}}$ is more salient that either parameter on its own, we normalize P_{thw} by setting it to 1 (see [6]).

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

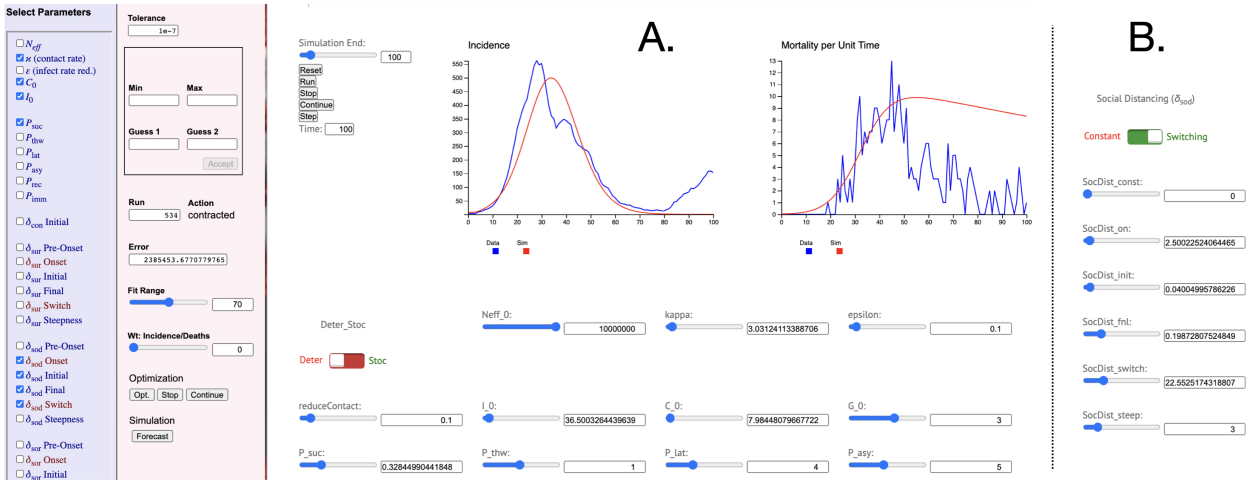

Figure 3: **A** Upper left corner of the optimization page of the NMB-DASA web app. Leftmost purple panel is where variables to be included in the optimization are checked. Next pink panel is where the details of the optimization procedure are entered: in this case the `Fit Range` is 70 and `Wt:Incidence/Deaths` is zero, implying that only the incidence curve is being fit. The mortality rate (i.e., virulence) for this fit has been set to 0.001, which provides a good fit to the outbreak phase but does not account for improvements in treatment that likely begin in early April (allowing for the estimated 16.5 delay between entering class I and death for those dying from COVID-19 [15]). **B.** The switching function parameter value windows on the the optimization page are arrayed along the bottom half of the page (not shown here). The social distancing driver switching parameters have been checked in the purple panel on the left, so here we show a cutout of the social distancing windows that contain the MLE values obtained during the optimization. In the pink panel we see the MLE took 534 iterations to converge, and the absolute log-likelihood error value is provided as well in the `Error` window.

After running optimization, the MLE values obtained can be saved using the `SAVE SETTINGS` button. If desired, an MLE value for the mortality rate can be obtained by unchecking all the variables use in the previous fitting, checking `vir_const`, resetting `Wt:incidence/deaths` to 0: and then pressing the `Opt` button again. We note that fitting incidence and mortality are not entirely independent processes, but they are are nearly so while the proportion dying is just a few percent and the proportion of susceptibles remains greater than 90%. By selecting `Wt:incidence/deaths` in between 0 and 1 we can fit both simultaneously, but are then faced with obtaining different solutions for different values of the `Wt:incidence/deaths` setting on $(0, 1)$.

3. *Fit Social Distancing Phase.* When using the model to project the outbreak pattern beyond the initial fit phase, a point will likely be reached where the model generated incidence time series is clearly growing more rapidly than the empirical time series. If the reverse occurs, then the *Fit Initial Outbreak Phase* step described above should be repeated using a longer `Fit Range` value. The reason why over estimation begins to occur is either because the effects of social distancing measure are beginning to take affect, or the effective population size is considerably smaller than initially anticipated. If the latter case is suspected then N_{nom} should be replaced by an N_{eff} parameter whose size is estimated as part of the MLE process described in the previous step and discussed in [12]. In the former case, however, the MLE parameters estimated in the *Fit Initial Phase* step should be fixed by unchecking all variables fitted in the

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

Fit Initial Phase optimization procedure, and the social relaxation driver variables checked to carry out a new MLE fitting procedure with the `Fit Range` parameter now set beyond its *Fit Initial Outbreak Phase* by a couple to several weeks, using the data as a rough guide to where this phases seem to be occurring.

4. *Fit Social Relaxation Phase.* The model can now be used to project the outbreak pattern beyond the *Fit Social Distancing Phase*. If the outbreak has been successfully quelled then the fitting task is complete. In many cases, however, the outbreak may either not be dampened sufficiently fast (as in the case of the COVID-19 in England [6]) or may begin to exhibit a second growth phase. In either case, the MLE values obtained in the *Fit Initial Outbreak Phase* and *Fit Social Distancing Phase* can be saved and a third round of MLE fitting executed, in this case we may choose to check only the social distancing driver parameters and optimize over a `Fit Range` that goes several weeks beyond the *Fit Social Distancing Phase*.
5. Scenario Projection and Counterfactual Assessments. We have considerable freedom to run counterfactuals and forecasts, depending on what values we give to the `Fitting Interval` and `Finish Timing for simulation` variables (Fig. 2), as illustrated in the Israel COVID-19 results reported next.

Results: COVID-19 in Israel

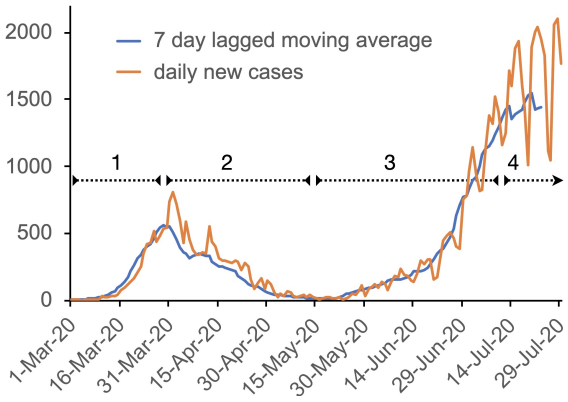

Figure 4: The daily incidence (new cases) time series (points connected by light brown lines) is plotted here from March 1, 2020 to July, 29, 2020. The blue curve is a lagged 7-day moving average [6]. The different phases of the outbreak analysed, indicated by broken range bars, are: 1. *Initial outbreak phase*, 2. *Social distancing phase*, 3. *Social relaxation phase*, 4. *Second wave mitigation phase.*

Background

The first infection of COVID-19 in Israel was confirmed on 21 February 2020 [15, 16]. Only 5 more cases were recorded in February, but starting March 1, a steady stream of cases began to be recorded. On March 9, a 14-day home isolation was imposed on all individuals arriving from abroad and on March 11 and then 15, gatherings were limited to a maximum of 100 people and then to 10 people. A national state of emergency was declared in the Israel and March 19 with the first recorded COVID-19 death of an Israeli citizen occurring on March 20 [15, 16].

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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 plot of the COVID-19 incidence data in Israel reveals a rising *initial outbreak phase* in March, followed by a steady drop during a *social distancing phase* to an average of fewer than 20 new cases per day during the second and third weeks of May (Fig. 4). During the last week of May, incidence began to rise steadily again during a *social relaxation phase*. A *second wave mitigation phase* set in around mid July as the second wave was flattened in response to measures used to bring the outbreak under control [17].

Fitting the first three phases of the outbreak

We begin by first using MLE to fit our basic free set of SCLAIV parameters (i.e., values for the contact rate, succumb period and initial number of individuals in classes C and I, which were not set *a priori* from literature data [6]) to the first 14 days of the *initial outbreak* data in March 2020. The fit was extremely tight, as shown in Fig. 5A, along with the four MLE values (Fig. 5B) obtained from the optimization procedure. A subsequent fit to the first month of data indicated that the effects of social distancing were already retarding the outbreak in the second half of March (Fig. 5C), yielding a new set of four MLE values (Fig. 5D) applicable to fitting the model over the first 31 days. The fit, however, was now quite poor, so we decided to added a social distancing component to our fit of the *initial outbreak* phase.

Our fit of the first 31 days with the addition of social distancing switching function yielded the fit depicted in Fig. 6A, where, though the fit only applies to the first 31 days, we simulated the subsequent model prediction out to day 137. If we try to fit all phases at once using the 13 free parameters in Fig. 6D, poorer fits may be obtained with the optimization procedure finding local rather than global or close-to-global optima.

Thus far, we have set the disease-induced mortality rate (i.e., virulence parameter) to 0.001. The fit obtained thereby appears reasonable for the initial phase of the outbreak, but not beyond the third week of April (left panel in Fig. 7). In May and June the mortality rates are considerably lower than this level, dropping from around 8-12 individuals per day in mid April to 0-2 individuals per day in late May. Thus we decided to fit a switching curve to the mortality data, by fixing all the values used to obtain the fit depicted in Fig. 6D, setting `Wt: Incidence/Deaths` to 1, and allowing the parameters `Vir_init`, `Vir_fnl`, `Vir_switch`, and `Vir_steep` to vary in an MLE fit of the model to the mortality data. The fit obtained is depicted in Fig. 7 (right panel). We note that the value for `Vir_onset` was a priori set at 0, since disease-induced mortality is always in effect (Fig. 8).

Stochastic projections

The forecasting page of the web app is designed to permit the flexibility of working with two data sets and two time windows: the first is to simulate the model in deterministic mode on an interval $[0, t_{\text{fit}}]$ and second to allow one (in deterministic mode) or repeated runs (in stochastic mode) on the interval $[t_{\text{fit}}, t_{\text{fin}}]$ (Fig. 9). The latter can be either executed by incorporating demographic stochasticity alone, or it can include a file of parameter values obtained from an MCMC fit of the model to data.

In Fig. 9, we illustrate a stochastic forecast from day 130 until day 274, which is the last day of November, 2020. The plotted standard deviations are based on 100 runs that use the MLE values obtained from fitting the model to the first 137 data points, as depicted in Fig. 6D, but reducing the `SocRel_fnl` from 0.1 up to day 130 (Fig. 6E with fit up day 137) to 0.015 beyond day 130 (see

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

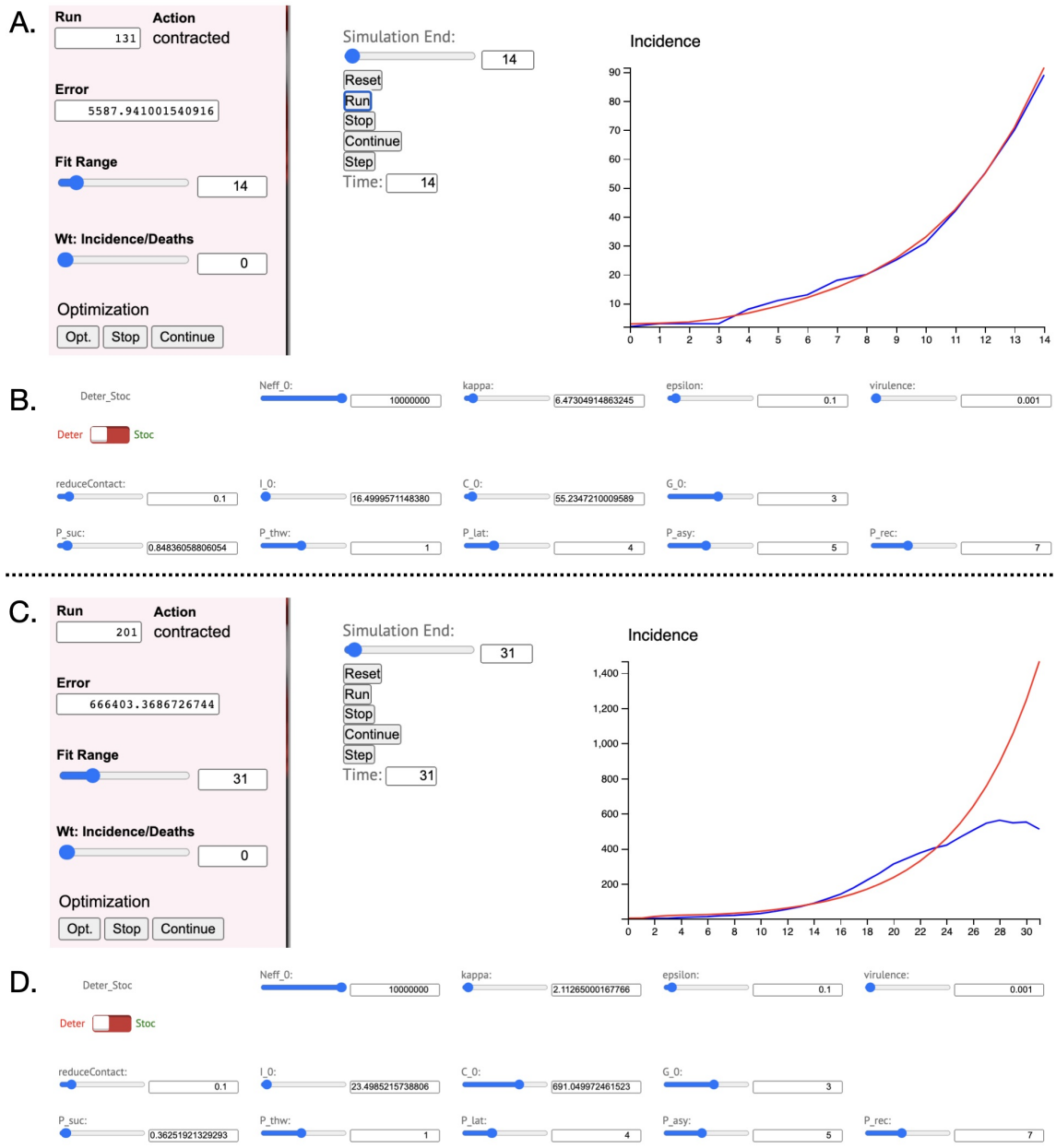

Figure 5: **A**. An MLE fit of the basic SCLAIV model to the first 14 days of data. **B**. MLE values obtained from the fit in **A. C**. An MLE fit of the basic SCLAIV model to the first 31 days of data. **D**. MLE values obtained from the fit in **C.**

the SOF for the same simulation using MCMC generated parameter fits and note that the $\pm$ 2 standard deviation band is considerably broader now). This change represents an 85% reduction in the final social relaxation rate. We see that this level of reduction steadies the outbreak at around 800 to 100 cases per day for the months of August to November. Other simulations (as depicted in our supplementary material) indicate that at least a 95% reduction in the final social relaxation

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

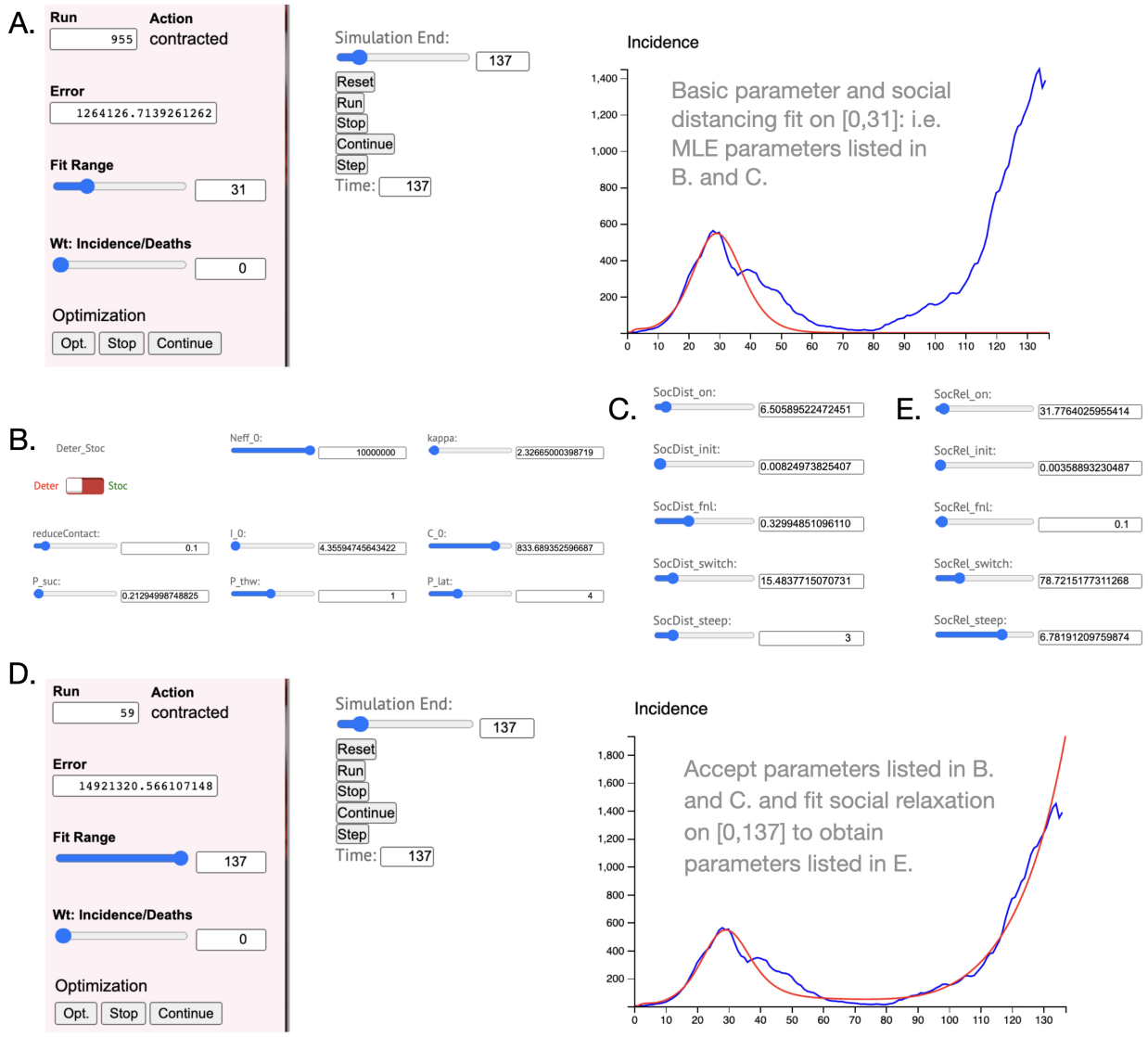

Figure 6: **A**. An MLE fit of the basic SCLAIV model to the first 31 days of incidence data, as in Fig. 5D, except now social distancing parameters are estimated as well, apart from `SocDist_steep` fixed at 3. The model is simulated out to day 137, rather than stopping at day 31 (`Fit Range` value). **B** & **C**. MLE values obtained from the fit in **A. D**. An MLE fit of the basic SCLAIV model to the first 137 days of data, with all parameters fixed at the values used to obtain the fit depicted in **A**, apart from the Social relaxation driver which has now been fit to the data. Note the value `SocRel_fnl` is not fixed at 0.1, but the fit is constrained by having set the lower bound on `SocRel_fnl` to 0.1. **E**. The MLE values obtained from the fit in **D.**

rate is needed to quell the outbreak during this period, while anything less then an 80% reduction is completely ineffective in containing the outbreak.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

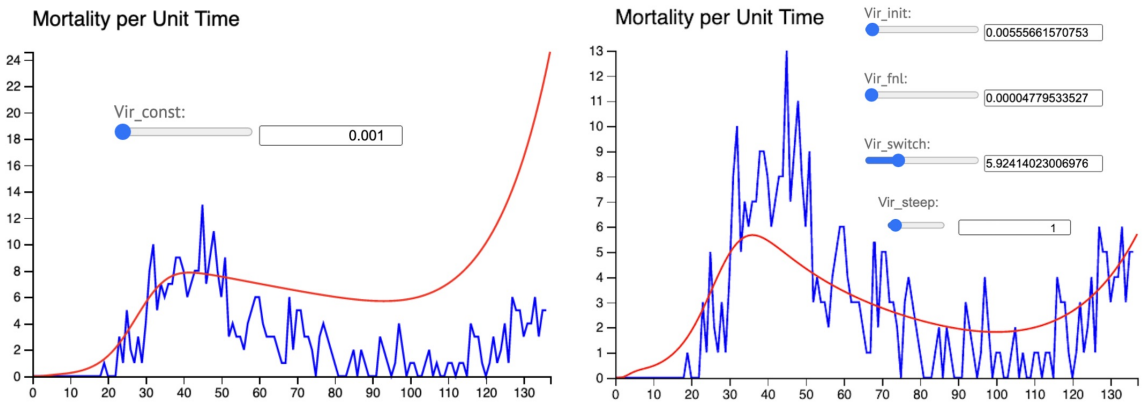

Figure 7: In fitting the parameters in Fig. 6D, the disease induced mortality had been set to `Vir_const`=0.001. In the left panel, we see that this value provides a reasonable fit until around day 50 but over estimates after day 50. If the all the parameters values used to fit Fig. 6D are fixed, except for letting four of the five virulence driver parameters vary (`Vir_on` is fixed at 0), then a better fit, as depicted in the right panel, is obtained, with four MLE switching values shown (note: `Vir_steep`=1 is a boundary constraint solution).

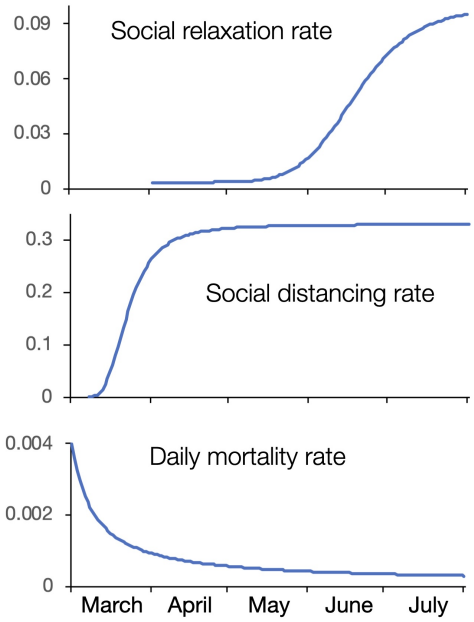

Figure 8: Social relaxation rate (applied only to class S), social distancing rate (applied only to class S_r, and disease-induced natural mortality rate (applied only classes I and S_r, as depicted in Fig. 1) switching curves have been fitted to the outbreak data, as illustrated in Fig. 6, using a progressive approach, as described in the text. The onset values for these three curves are on days 31 (April 1, 2020), 6 (March 7,2020), and 0 (March 1, 2020) respectively. The switching points (half way between initial and final values) are on days 78 (May 18), 15 (March 16) and 8 (March 9) respectively. The range of initial to final values are indicated by the values that scale the vertical axes of the three curves in question.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

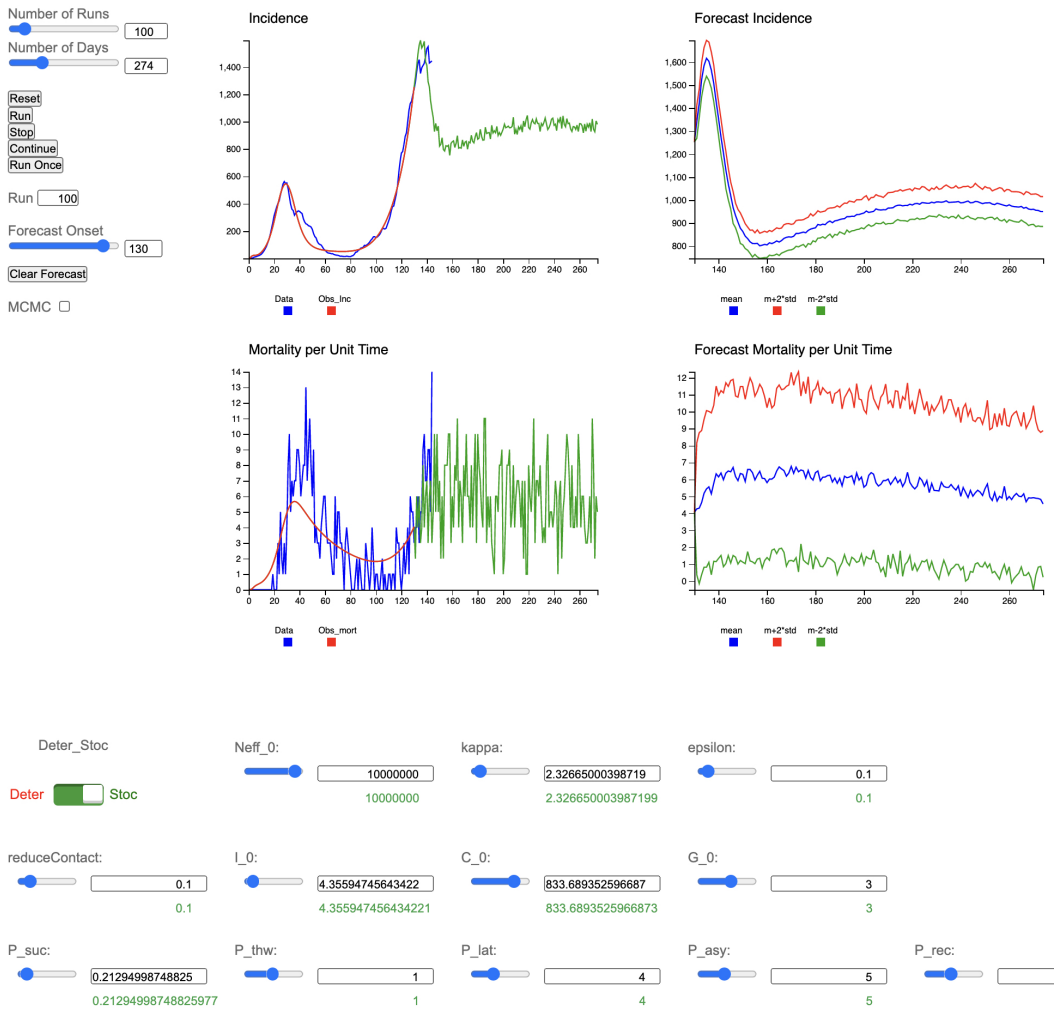

Figure 9: A partial view of the *Forecasting page* show the `Number of Runs` to be performed, `The number of Days` for the total simulations and the `Forecast Onset` day. The first part of the model projections of incidence and mortality per unit time are in red in the two left most graphics panels, along with the data plotted in blue. The green curves beyond the red are single stochastic simulations (since the `Deter_Stoc` switch is flipped to the `Stoc`). The two right side panels represent the mean plus/minus 2 standard deviations generated from 100 runs, as specified in the `Number of Runs` window.

Discussion

Currently, web sites available for exploring COVID-19 policy implications are rather limited. The most flexible of these is the COVID-19 Simulator, developed by a consortium that includes the Massachusetts General Hospital, Harvard Medical School, Georgia Tech, and Boston Medical Center. The COVID-19 Simulator provides "... a tool designed to help policymakers decide how to respond to the novel coronavirus." The tool can be used to explore the impact of social-distancing interventions by selecting one of four intensity levels applied at a selected time. The simulator does not allow the user to upload his or her own data and limits analyses to US National and State levels.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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 site claims that "The information presented can help policymakers understand consequences such as the rate of new cases, potential strain on the healthcare system, and projected deaths."

Our NMB-DASA is much more comprehensive in terms of allowing users to input their own data, fit model parameters to data using maximum-likelihood methods, undertake scenario and forecasting studies that include many more drivers than are available to users of the COVID-19 Simulator site, as well as extract the time varying forms of those driving (Fig. 8). Additionally, the Israeli COVID-19 case study presented here provides a clear exemplar on how our NMB-DASA Web App can be used to undertake a forensic analysis of the phases and drivers of a outbreak in any region of the world for which incidence time series are available, as well as mortality time series for those studies evaluating the impact of new therapeutics on reducing mortality. If other healthcare and social media infrastructural data are obtainable that can be used to independently specify or estimate surveillance [18, 19], social distancing and relaxation [20], quarantining [4] and treatment advances, then further analyses can be undertaken to verify results and check predicted proportions in the immune class using serology data [21], as discussed elsewhere [6].

Although our exemplar indicates that NMB-DASA can be used to perform analyses of disease outbreaks caused by directly transmissible pathogens for which incidence data alone are available, much deeper analyses can be undertaken. First, spatial structure can be taken into account [12, 22–24]. In, Israel, for example, the top 20 settlements in terms of cases per person are 16 ultra orthodox and 4 Arab communities, which are relatively isolated from one another in socio-geographic terms. Without including spatial structure, however, good fits may still be obtained (also see [12]). Second, links between model output and data may be sought that pertain to individual and societal factors influencing transmission. This can be done by making statistical links between model parameters putative factors [25, 26]. This would require the dynamic analyses of the types demonstrated here augmented with statistical models that linked dynamic measures (such as new cases growth rates over initial and later phases of the outbreak) to key factors of interest.

COVID-19 is not an anomaly, but one event in series of expected future events that have a past history in SARS, MERS, Ebola, not to mention avian influenza and other highly contagious emerging infectious diseases of zoonotic origin [27–29]. Our current experience with COVID-19 will help facilitate our response to these future outbreaks and potential pandemics. The availability of computational tools, such as NMB-DASA, will enable policy makers and healthcare administrators to be caught less flat-footed at the start of the next outbreak than has been the case for COVID-19 pandemic in terms of evaluating the effects of mitigating epidemic drivers that might be deployed to manage local outbreaks.

Authors Contributions

Wayne M. Getz, conceived the study, formulated the model and wrote the manuscript. Richard Salter developed the web site and the Numerus Modeling Builder software platform. Richard Salter and Ludovica Luisa Vissat were responsible for the MCMC simulations. Nir Horvitz scrapped data from the web. All authors read and edited the final version of the paper.

Supplementary Material

A supplementary file contain additional figures referred to in the text is provided. Three training videos are available with links provided by selecting the instructional material button at the NMB-DASA Home page. Direct links to these videos are: Training video 1, Introduction to the Numerus Data and Simulation

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

Analyses (DASA) Website for Covid-19; Training video 2, Parameter Estimation Through Model Fitting; Traning video 3, Deterministic and Stochastic Forecasting

References

[1] Zhang J, Litvinova M, Liang Y, Wang Y, Wang W, Zhao S, et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. Science. 2020;.

[2] Wilder-Smith A, Freedman DO. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. Journal of travel medicine. 2020;27(2):taaa020.

[3] Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. The Lancet Infectious Diseases. 2020;.

[4] Tang B, Xia F, Tang S, Bragazzi NL, Li Q, Sun X, et al. The effectiveness of quarantine and isolation determine the trend of the COVID-19 epidemics in the final phase of the current outbreak in China. International Journal of Infectious Diseases. 2020;.

[5] Weissman GE, Crane-Droesch A, Chivers C, Luong T, Hanish A, Levy MZ, et al. Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic. Annals of internal medicine. 2020;.

[6] Getz WM, Luisa-Vissat L, Salter R. A Contact-Explicit Covid-19 Epidemic and Response Assessment Model. medRxiv. 2020;.

[7] Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. Annals of internal medicine. 2020;172(9):577–582.

[8] Linton NM, Kobayashi T, Yang Y, Hayashi K, Akhmetzhanov AR, Jung Sm, et al. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. Journal of clinical medicine. 2020;9(2):538.

[9] Bar-On YM, Sender R, Flamholz AI, Phillips R, Milo R. A quantitative compendium of COVID-19 epidemiology. arXiv preprint arXiv:200601283. 2020;.

[10] Van den Driessche P, Watmough J. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. Mathematical biosciences. 2002;180(1):29–48.

[11] Sette A, Crotty S. Pre-existing immunity to SARS-CoV-2: the knowns and unknowns. Nature Reviews Immunology. 2020;p. 1–2.

[12] Getz WM, Salter R, Muellerklein O, Yoon HS, Tallam K. Modeling epidemics: A primer and Numerus Model Builder implementation. Epidemics. 2018;25:9–19.

[13] Hamra G, MacLehose R, Richardson D. Markov chain Monte Carlo: an introduction for epidemiologists. International journal of epidemiology. 2013;42(2):627–634.

[14] Roberts GO, Rosenthal JS, et al. General state space Markov chains and MCMC algorithms. Probability Surveys. 2004;1:20–71.

[15] Last M. The First Wave of COVID-19 in Israel-Initial Analysis of Publicly Available Data. medRxiv. 2020;.

medRxiv preprint doi: https://doi.org/10.1101/2020.08.06.20155804; this version posted August 11, 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.

[16] Rossman H, Keshet A, Shilo S, Gavrieli A, Bauman T, Cohen O, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. Nature Medicine. 2020;26(5):634–638.

[17] Bodas M, Peleg K. Self-Isolation Compliance In The COVID-19 Era Influenced By Compensation: Findings From A Recent Survey In Israel: Public attitudes toward the COVID-19 outbreak and self-isolation: a cross sectional study of the adult population of Israel. Health Affairs. 2020;39(6):936–941.

[18] Yue M, Clapham HE, Cook AR. Estimating the Size of a COVID-19 Epidemic from Surveillance Systems. Epidemiology. 2020;31(4):567–569.

[19] Silverman JD, Hupert N, Washburne AD. Using influenza surveillance networks to estimate state-specific prevalence of SARS-CoV-2 in the United States. Science translational medicine. 2020;12(554).

[20] Yasaka TM, Lehrich BM, Sahyouni R. Peer-to-Peer contact tracing: development of a privacy-preserving smartphone app. JMIR mHealth and uHealth. 2020;8(4):e18936.

[21] Bastos ML, Tavaziva G, Abidi SK, Campbell JR, Haraoui LP, Johnston JC, et al. Diagnostic accuracy of serological tests for covid-19: systematic review and meta-analysis. bmj. 2020;370.

[22] Huang R, Liu M, Ding Y. Spatial-temporal distribution of COVID-19 in China and its prediction: A data-driven modeling analysis. The Journal of Infection in Developing Countries. 2020;14(03):246–253.

[23] Kramer AM, Pulliam JT, Alexander LW, Park AW, Rohani P, Drake JM. Spatial spread of the West Africa Ebola epidemic. Open Science. 2016;3(8):160294.

[24] Getz WM, Salter R, Mgbara W. Adequacy of SEIR models when epidemics have spatial structure: Ebola in Sierra Leone. Philosophical Transactions of the Royal Society B. 2019;374(1775):20180282.

[25] Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. Nature. 2020;p. 1–11.

[26] Lakshmi Priyadarsini S, Suresh M. Factors influencing the epidemiological characteristics of pandemic COVID 19: A TISM approach. International Journal of Healthcare Management. 2020;13(2):89–98.

[27] Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. Nature. 2008;451(7181):990–993.

[28] Evans TS, Shi Z, Boots M, Liu W, Olival KJ, Xiao X, et al. Synergistic China–US ecological research is essential for global emerging infectious disease preparedness. EcoHealth. 2020;p. 1–14.

[29] Reperant LA, Osterhaus AD. AIDS, Avian flu, SARS, MERS, Ebola, Zika... what next? Vaccine. 2017;35(35):4470–4474.