Generalized modeling of interventions across 77 countries

Richard K. Belew,[*,†] Cliff Kerr,[*,‡] Jasmina Panovska-Griffiths,[*,¶] and William Waites[*,§]

†*Univ. California – San Diego, La Jolla CA*

‡*Institute for Disease Modeling, Bellevue, WA*

¶*University College London*

§*London School of Hygiene and Tropical Medicine*

E-mail: rbelew@ucsd.edu; ckerr@idmod.org; j.panovska-griffiths@ucl.ac.uk; wwaites@tardis.ed.ac.uk

Draft: 16 Mar 2021

Abstract

COVID-19 continues to spread around the world and modeling plays an important role in informing policy[1]. An individual-based model called Covasim has recently been fit to data regarding confirmed cases and deaths experience in the United Kingdom during the first half of 2020, and then used it to evaluate alternative intervention strategies there[2]. We extend this methodology to consider data from 77 countries to retrospectively model interventions employed in the first months of 2020 in these countries. Because the age distribution of populations is a key feature of the COVID-19 pandemic and contacts among young people are often age-stratified and may play an especially important role, we focus here on school closure interventions.

A Covasim model was built for each country, with its population, age distribution, disease emergence and school closing dates specified. The model was then calibrated against data concerning COVID-19 diagnosed cases and related deaths. All countries' simulations were run with and without school interventions. Because assumptions regarding background testing rates being performed are critical to statistics regarding the number of diagnosed patients, three conditions were considered: the same *fixed* testing rated was assumed for all countries; *searching* for an estimated testing rate as part of model calibration; the actual *data* available. The goal of this *post hoc* analysis of historical data is to understand the limits of our modeling tools as we move forward to use them for predictive tasks. We consider the hypothesis: Does incorporating the additional knowledge of school interventions into a model makes it a better predictor of data concerning diagnoses and deaths?

The calibration of model parameters plays an important role in their evaluation, and an analysis of methods used to calibrate models is considered in detail. Experiments using Covariance Matrix Adaptation Evolution Strategy (CMA-ES) to search for model parameters that allowed a model to best fit available data, and contrasted with parameters found using the pyABC library to implement Approximate Bayesian Computation (ABC) techniques which also provided kernal density estimates (KDE).

In only one of the 77 countries considered did model parameters using testing rate data differ significantly from ones using search to calibrate this value. Incorporation of school intervention in models lead to significant differences in 22 cases, but surprisingly with only two having better model fit with intervention modeled. Beyond the KDE estimates it provides, in most cases pyABC proves to be an even more effective parameter search method than CMA-ES.

1 Introduction

COVID-19 continues to spread around the world and modeling plays an important role in informing policy[1]. Here we use a broad range of countries' experiences to identify robust

results common across them. Comparisons between countries are essential for the control of COVID-19[3]

> Although international comparisons are often disparaged because of different data quality and fears of the 'ecological fallacy', if done carefully they can play a major role in our learning what works best for controlling COVID-19.[4]
>
> ... the COVID-19 epidemic shows the need for epidemiology to go back to its roots—thinking about populations. Studying disease occurrence by person, place and time (often referred to as 'descriptive epidemiology') is usually taught in introductory courses, even if this approach is then paid little attention subsequently. COVID-19 is a striking example of how we can learn a great deal from comparing countries, states, regions, time trends and persons, despite of all the difficulties.[4]
>
> Predictive models for large countries, such as the US, are even more problematic because they aggregate heterogeneous sub-epidemics in local areas.... Models should also seek to use the best possible data for local predictions.[3]

Covasim is a stochastic agent-based simulator for performing COVID-19 analyses. It depends on two primary parameters, INITINFECT, the number of people initially infected at the start of the contagion, and a baseline value for β, the transmission rate from susceptible to infected individuals. A key feature of Covasim is its modeling of interaction network "layers" distinguishing various settings (work, school, transport, households, communities) which modulate β differently in these settings.

With travel restrictions as they were in the spring and summer of 2020, levels of migration across national borders were considerably smaller than that across state or provincial boundaries. Given the data available, only models at the level of individual countries and interventions ordered *nationally* are considered here.[1] Because the age distribution of pop-

[1]The absence of a national school closure policy in the United States is the reason it is not included in this analysis.

ulations is a key feature of the COVID-19 pandemic and contacts among young people are often age-stratified and may play an especially important role, we focus here on school closure interventions.

Table 1 lists the 77 countries included in this study with their ISO-3 code and name, together with key statistics:

- population in 2019, from ECDC
- the start date for their simulation (when the number of infection first went above 50)
- the start and end dates of the school intervention; the "effective" start date is different in those cases that the intervention began before the simulation's start date
- number of days closed by school interventions, and the fraction of total simulated days this represents

Note that the countries modeled varied in population from just over a million (EST) to IND with a population of 1.4 billion.

To give a preview of the experiments to be presented, consider the comparison of two runs presented in Figure 2. This shows six curves summarizing two simulation run for the country of Serbia (ISO3 code SRB). The two curves in the top left graph show the simulation's estimate (solid line) and data (boxes) of cummulative diagnoses during early 2020, the middle graph shows estimated vs. data on cummulative deaths, and the lowest graph shows the cummulative number of tests performed. These same curves are repeated for two different experimental conditions, in this case simulations with and without school interventions included in the model. Better models are those whose simulated estimates for diagnoses and deaths have the best match with available data (cf. Section 3.2). Section 2.2 contrasts models using different bases for testing, and Section 2.3 contrasts models with and without school closure interventions in the model.

On their face, because the right simulation's estimates of deaths is a closer match to data, these results make it seem as if incorporating school closures in the model is an improvement.

This would make us believe that modeling interventions is worthwhile, and that the model parameter values (β and INITINFECT) associated with this model should be considered more accurate. But is the difference between the two models *statistically significant*? Typically, global optimization of simulation parameters is performed to find values which minimize model mismatch[5]. In Section **??** we describe the optimization of Covasim parameters using Covariance matrix adaptation evolution strategy (CMA-ES). In Section 2.4 these results are also put in the probabilistically more sound context of Approximate Bayesian Computation (ABC).

As the database supporting infections, deaths and school interventions during the first months of 2020 recedes into the past, it becomes increasingly irrelevant to the pandemic situation now. The goal of this *post hoc* analysis over the broad set of historical data across many countries is to understand the limits of our modeling tools as we move forward to use them for analysis of newer data and predictive purposes.

2 Results

The first section below reports on "base model" experiments that do *not* include school intervention, which use *data* for the testing rate, and which has been calibrated using ABC methods. Experiments contrasting these models with ones *searching* for the testing rate are then presented in are in Section 2.2, ones including school interventions in Section 2.3, and ones calibrated using CMA-ES global optization in Section **??** An analysis of the statistical significance of these differences is considered in Section 2.4.

2.1 Base model

- use of ABC on noSchool, trate=data as base case
- Example country

- contrasted with trate=srch, school, CMAES

Covasim defines the value of a model's fit to be a weighted sum of the error between its predicted number of diagnoses and deaths related to COVID-19 and data for these values. Figure 3 shows the range of resulting model fits across 77 countries, with lower fitness scores assigned to simulations with a better match of predictions for the number of diagnoses and deaths. Most countries have fit values below 150, and the average fit across countries is 91.6, including the poor fit examples of UGA and SGP.

Figure 4 shows the result of two simulation for ETH and SGP; ETH fit is one of the best, and SGP is the worst. For each country, the top two plots capture model behavior relative to data for the number of diagnoses across the time interval considered. The bottom plot shows the number of tests given over time.

Figure 5 shows the values of the key INITINFECT and β parameters for all countries. Best KDE estimates for INITINFECT vary by three orders of magnitude, and those for β vary by a factor of three.

2.2 Using available data for testing rate

A central issue in epidemiological modeling concerns assumptions regarding the background testing rates on which diagnoses counts are based. Experiments over countries with data on testing rates (daily testing counts as a fraction of the country's population) via OWID were contrasted with average testing rates found via calibration. Figure 6 contrasts the testing rate found via calibration search vs. the data-based rate, with bubble size reflecting the difference in fitness between the two models[2].

The testing rates identified via calibration were generally much larger (20x) than the data-based values. Two countries, ARE and BHR, had much higher testing rates reported in the data than the others, with little difference reflected by their calibrated values. Only

[2]Because bubble sizes must be positive, blue bubbles are used to show improved fitness with calibrated testing rates, red bubbles with data-based testing rates.

the difference in testing rates for OMN proves signficantly different (cf. Section 2.5).

As a baseline, these results were also compared to ones using the same, constant testing rate (as a fraction of the population tested per day) of TESTING_RATE=3.3e-4, the average across all countries with data (shown as a red line in Figure 6). Simulations using this value for testing rate did not appreciably change model fit from ones using data-based values (data not shown).

Note that the difference between predicted testing rate and data regarding this value was *not* used by the calibration process in these experiments to guide the search for model parameters, although Covasim's Fit measure could be modified to do so.

2.3 Incorporating school closure interventions

The next set of experiments contrast runs with and without school closure interventions included in the model. As shown in Figure 7, most models' fits were not changed much by including school intervention. Since fit is measured using a scale where smaller values mean better fit, negative differences indicate models which were improved with interventions modeled, and positive differences indicate poorer fit when interventions are included. When model fit changed, the model *without* school intervention modeled was typically superior. Figure 7 also breaks out those countries for which the different in β values between the two models was found to be significantly different (cf. Table 1).

The example of SRB used in Figure 2 showed a slight improvement in fit (difference = -23.6). The dashed lines in the right-hand figures reflect the beginning and ending dates of the interventions. ISR is the only other country's model with improved fitness (difference= -35.3) and significant difference in β values. In contrast, the results for ITA shown in Figure 8shows incorporation of school interventions increased (harmed) the fit most (difference=196.0). The differences in β values was statistically significant for ITA and all others with fitness differences > 100.

2.4 Model calibration using ABC and CMA-ES

> Calibration is a complex and dark art and cannot be covered fully here; many books have been written about it and it continues to be an area of active research. A good review article about calibrating agent-based models like Covasim is available here[5]. Calibration is usually expressed as an optimization problem: specifically, find a vector of parameters θ that minimizes the mismatch between the data D and the model $M(\theta)$. Covasim Tutorial#7

Computational complexity theory "...help[s] us comprehend why so many practical problems seem to be resistant to efficient solution by computer" and warns that for "total search" problems such as this, there can be no guarantee of finding the actual optimum value[6].

Initial experiments calibrating Covasim with school intervention data were done using CMA-ES. The goal of providing a statistical foundation for model parameters found via calibration, however, subsequently lead to the use of ABC methods as described in Section 3.4. The ABC method is motivated there, along with experiments calibrating a simpler compartmental ODE model to build intuition about ABC search. The values found by ABC calibration are the ones described in all preceding experiments. Section 2.5 addresses the statistical signfiicance of the observed differences across experiments.

Here we begin by putting all the various experimental results into a common frame by considering their differing values for the calibrated parameters β and INITINFECT. Figure 9 focuses on SRB, the same example mentioned in the introduction and Section 2.3. The KDE for these parameters β and INITINFECT found by ABC search are represented by the highlighted red point. A key advantage of ABC methods is that they can provide kernel density estimates of the parameters they find that minimize model fit. For SMC-ABC this requires weighting samples according to their "importance" and then computing credible intervals of some confidence. Confidence=0.9 error bars on the β and INITINFECT of the KDE estimate are also displayed.

Three other points showing the parameter values β and INITINFECT found and their

respective fitnesses in their labels are also shown:

1. using search to establish the testing rate (vs. the base model's use of data);
2. incorporating school interventions (vs. not); and
3. using CMA-ES to discover parameter values (vs. ABC)

The first two experiments also provide KDE estimates, and the error bars for these are shown.

The pyABC library does not explicitly associate a fitness value with the KDE parameter estimate[3] and so the *minimum* fitness point found during ABC search is also shown, and labeled with its fitness. The fitness values shown in the other labels is similarly taken from the corresponding minimum fitness.

2.5 Statistical significance

The disjoint error bars in Figure 9 imply that in the case of SRB, the difference in β and initInfect parameters on effects on models with/out school intervention can be considered statistically significant at the 90% confidence level. Across the 77 countries considered, in 22 of them the differences in β satisfied this significance test.

Only two models, for and ISR and SRB, showed improvement in fit with school interventions of more than 10.

Most but not all of the significant differences observed were in β between the base experiment contrasted with school closures. In ARG the difference was also signficant with respect to the difference in initInfect, and in OMN the difference in β was signficant with respect to the model using calibration vs. data for testing rate.

It is reasonable to hypothesize that longer school closures will make it easier for models of them to demonstrate signficant differences. Figure 10 orders countries (left-to-right) based on the fraction of the modeled days had school closures in effect (cf. 1), with blue boxes

[3]Specifically: pyabc.visualization.credible.compute_quantile(vals, w, 0.5)

Table 1: Countries with significant differences in β

ISO3	KDE_fit	school_fit	diff	Beta-KDE
ISR	60.8	25.5	-35.3	0.0049
SRB	55.2	31.6	-23.6	0.0044
PSE	122.9	116.8	-6.1	0.0016
GRC	44.8	40.3	-4.5	0.0035
MYS	37.1	44.3	7.2	0.0030
THA	94.2	101.7	7.5	0.0029
PRT	63.9	72.8	9.0	0.0038
SVK	43.0	52.1	9.1	0.0034
FRA	108.7	126.1	17.5	0.0053
DEU	65.2	92.4	27.1	0.0056
NLD	79.7	117.3	37.6	0.0045
IDN	24.5	65.2	40.8	0.0047
PHL	26.5	77.5	51.0	0.0044
GBR	91.5	144.6	53.1	0.0058
NGA	24.7	84.8	60.1	0.0047
CAN	72.5	147.8	75.2	0.0049
IRN	91.2	173.1	82.0	0.0048
IND	18.1	122.6	104.6	0.0051
ARG	12.8	119.7	106.9	0.0055
MEX	39.6	164.9	125.3	0.0052
CHL	29.9	200.2	170.4	0.0051
ITA	104.1	300.1	196.0	0.0055

placed at 1.0 if they demonstrated a signficant difference in β, and zero otherwise. Also plotted is the cummulative fraction of countries that showed signficant differences at that closure level.[4]

Examples of four countries (highlighted in Figure 10) are shown below. ARG is one of 30 countries that had school closures during the entire period modeled, and one of the four demonstrating a signficant difference in β. DEU showed significant difference even though schools there were open 68.6% of the time. GBR schools were closed 52.7% of the time, and THA schools 37.3% of the time. As mentioned above, for ARG the difference in INITINFECT is also signficant, while in DEU the value of INITINFECT in models with/out school closure is nearly identical.

2.6 Difference with CMA-ES calibrated values

As discussed above, model calibration using ABC methods brings with it the virtue of an estimate of the the parameters found are the best KDE. However, other global optimization methods may be more successful at discovering well-fitting model parameters, without any claims regarding posterior estimates. The results of CMA-ES can be compared in terms of their discovery of better or worse model fits; this distribution is shown in Figure 12. In most cases, the model fit using KDE parameters was about the same as using CMA-ES values. On average, however the pyABC calibration methods found better fitting model parameters, sometimes significantly better.

The parameters of those models with better CMA-ES fit are worth comparing to the KDE parameters. Figure 13 shows the example of KAZ which with a model fitting slightly better (difference = -21.3) does so with values (marked with a large X in the figure) outside the confidence interval associated with the KDE parameters. . Only 11 differences of this "semi-significant" sort[5] are found across all countries' models, and these are listed in Table

[4]The many red points along the vertical axis correspond to the 30 countries closed during the entire period.

[5]Recall that no confidence intervals can be associated with the CMA-ES parameters.

2.

Table 2: CMA-ES superior fitness

ISO3	Fit diff	param	KDE value	CMA-ES value
ARE	-12.8	initInfect	8760	2761
BHR	-35.5	initInfect	2302	1001
CAN	-22.0	initInfect	32197	49833
KAZ	-21.3	beta	0.00517	0.00632
KAZ	-21.3	initInfect	4035	1000
LBY	-16.3	beta	0.00437	0.00562
MEX	-12.1	initInfect	36659	49917
PSE	-12.5	beta	0.00161	0.00432
PSE	-12.5	initInfect	2428	1000
SLV	-30.8	beta	0.00433	0.00527
TGO	-12.6	beta	0.00238	0.00397

3 Methods

3.1 Data sources

The ECDC was used as the primary organizing data source. ISO3 codes were used to identify and merge country data.[6]

- ECDC
- OWID testing[7]
- UNESCO interventions
- OxCGRT interventions
- Age distribution data from the Neher Lab as distributed as part of Covasim[7]

[6]Two countries missing from ECDC lists, TWN and HNG. 2019 population populations for these two was obtained from 'pop19': 23773876 }WorldOMeters

[7]cf. `covasim.data.country_age_data.py`

Only countries with populations greater than one million were considered, and the start date for each countries' simulation was picked to begin when data showed the number of infections went above 50.

3.1.1 School interventions

Preliminary experiments used a database of interventions developed by a database of international intervention specifics called Covid-19 Control Strategies List, developed by Amélie Desvars-Larrive and colleagues [Complexity Science Hub Vienna]. This data includes a fine-grained analysis of which school levels (kindergarten, primary, secondary, university) were ordered closed, and therefore supported fine-grained variations in the age distributions of the sort used effectively by Covasim's population generator. 37 countries were included in this data.

The experiments reported here used instead a *consensus* database of school closures coming from two distinct data sources, OxCGRT and UNESCO. While broadly consistent, their respective accounts as to just when individual countries were closed nationally was not identical. Figure 15 shows two countries' examples: BHR, a country with exactly overlapping accounts and DNK with more discordant ones. The graphs each show two plots of when school closures were in effect (a value of 1, vs. 0 when not in effect) as described by a source.

The specification of school closures used in these experiments was conservative: Only intervention periods that were identified by *both* sources, for periods of at least 14 days were considered. All countries' simulatons were begun on this date (between February 14 and May, 22, 2020; cf. Figure 1), interventions of some period (14-150 days) were applied, and all simulations were terminated at July 31, 2020.

3.1.2 Testing rate data

Testing data from OurWorldInData[7] was used. For the comparisons in Section 2.2, this daily testing data was converted to an average testing rate using the country's population

and averaging across the time period of the simulation.

3.2 Covasim

Covasim[8] is an open-source agent-based (a.k.a. individual-based) model that uses demographic information on age structure and population size to build realistic transmission networks within distinct social layers. It allows age-specific disease outcomes, intrahost viral dynamics, and the ability to incorporate many sorts of interventions affecting these model elements.

The basic model requires specification of two key parameters, the initial number of infected individuals in the population, and β, the transmission rate from susceptible to infected individuals. The global β parameter is proposed by the calibration optimizer, then enhanced/attenuated across individual network layers according to parameters based loosely on time-use surveys that track how many hours a week people spend in various settings (work, school, transport, etc), with households getting an additional multiplier for closeness of contact.

Better models are those whose simulated estimates for diagnoses and deaths are closer to available data. In experiments here we used Covasim's default Fit measure (normalized absolute difference) of mismatch. As data regarding deaths is believed to generally more accurate than that about positive diagnoses, Covasim uses a weighted sum of error with death rates weighted twice as heavily as diagnoses.

School closures are modeled as a simple on/off: school closings start/stop on a specified dates, and layer-specific values for β are enhanced or attenuated. Following[2], infection rate was set very low within the schools level, increased within the home level, and lowered significantly in work and community levels. The net values for β incorporating both default Covasim values and the result of school interventions are shown in Table 3.

Table 3: Per-level β values

β	Covasim	Intervention	Net with intervention
Home	3.00	1.29	3.87
School	0.60	0.02	0.01
Work	0.60	0.20	0.12
Community	0.30	0.20	0.06

3.3 Calibration using CMA-ES

Covariance matrix adaptation evolution strategy (CMA-ES)[9] is a library for stochastic optimization over continuous domains of non-linear, non-convex functions. The "ES" refers to "Evolutionary strategy," a type of evolutionary computation using the metaphors of populations and generations, and distinguished by adaptive rates of mutation across generations. The CMA library has a "tell-ask" interface that allows easy bundling of parallel executions of a population alternative values in each generation. The initial range of mutation σ_0 was set to 0.1, and the initial search range for all parameters was set to one quarter of the range between their upper and lower bounds.

CMA-ES was used as one calibration outer-loop around the basic Covasim simulation to search for key parameter values that allow the model to best fit against data. To make computational effort comparable to the approximately 8000 model evaluations used in experiments with pyABC, 67 generations of 120 individuals = 8040 trials total were allocated to find the search for parameter values causing the model to best fit the data. Optimization is always over two parameters, β and initInfect, the number of initial infections. Initial search bounds for the optimization were set:

- initInfect: best = 21000, lower bound =16000, upper bound=26000
- β: best = .005, lower bound = 0.001, upper bound=0.01

In the experiments described in Section 2.2, a third parameter corresponding to testing rate is also included. Bounds on its search were established using the data for the 77 countries available:

- testingRate: best=3.3e-4, lower bound = 2e-6, upper bound=3e-3

3.4 Calibration using ABC

Bayesian methods are distinguished by their inference of a posterior distribution for some set of parameters θ from the likelihood of observed data given these values $\pi(y|\theta)$ and estimation of the parameters prior to using any data $\pi(\theta)$:

$$\pi(\theta|y) \propto \pi(y|\theta)\,\pi(\theta)$$

Many issues arise (e.g., determining the normalization constant) in translating this probabilistic strategy into an empirical estimate, but a deeper issue concerns models with "hidden" variables not reflected in observable data. The basic Bayesian inference can be extended to consider hidden variables x using the joint likelihood of the data with these variables $\pi(y, x|\theta)$:

$$\pi(\theta, x|y) \propto \pi(y, x|\theta)\,\pi(\theta)$$

Simulations like Covasim can, when given input parameters θ, provide samples of data and hidden variable values from the joint likelihood function. Sampling methods like Markov Chain Monte Carlo (MCMC) can then provide numerical estimates for the posterior distribution. Approximate Bayesian Computation versions of MCMC (ABC-MCMC) proceed by iteratively selecting "particles" (simulations run with parameters θ_i) that produce predictions z_i that are required to be closer $||y - z_i|| < \epsilon$ to observed data, rejecting those that are not, and resampling[10]. Sequential Monte Carlo ABC (SMC-ABC) techniques specify a transition kernel for generating new parameters to test and a method for reducing ϵ_t over iterations[11].

pyABC provides a Python implementation of SMC-ABC supporting parallel execution of simulations on multicore hardware[12]. Simulations were given a uniform prior distribution over the same bounds used for CMA-ES optimization. Other parameters control resources

allocated to SMC-ABC search such as the maximum number of "generations" of search reducing ϵ_t, an absolute minimum on ϵ_t, etc.

Figure 16 gives a sense of the search performed by pyABC, in this case for SRB. 16.a shows the converging posterior estimates for β and initInfect over the nine generations of search, and 16.b shows the declining ϵ_t used to drive the search to better and better fitting models.

4 Conclusions

Experiments reported consider a standardized modeling approach applied across a broad range of countries. The countries were selected because they had available data on both nation-wide school closures and testing rates during the early, dynamic days of the pandemic in early 2020 when strategies varied widely and the impact of these interventions can be imagined to have large impact. It was expected that incorporating school interventions into models would allow them to better fit data concerning deaths and diagnoses, but generally that was not found to be the case; restricting attention to only models with statistically significant difference in parameters only SRB and ISR showed improved fitness. Similarly, it was expected than making use of data on testing rates within countries would lead to better models than assuming a constant testing rate or allowing calibration to set testing rates at levels that lead to better fitting models. While calibrated testing rates did differ considerably from data-based values, differences in the resulting models did not generally lead to better fit with deaths and diagnosis data and in only one case (OMN) lead to models with significantly different parameter values.

A comparison of global CMA-ES optimization techniques to Bayesian SMC-ABC calibration using the pyABC library showed broadly consistent results in the parameter values discovered by the two methods, but also that pyABC was generally even more effective at calibrating Covasim model parameters than CMA-ES.

4.1 Scientific sharing, publishing and open source models

> Early models relied on sparse, sometimes unreliable, data, and modelers did not anticipate the emergence of important new facts.... For use in an emergency, models developed through basic research need to be "operationalized"—that is, made robust for evaluating specific policy interventions. "Nowcasting" requires models that integrate incomplete, real-time data and emerging medical knowledge to provide situational awareness.... Models must also incorporate behavioral responses to policy interventions that may change the course of an epidemic.[1]

Traditional compartmental models represent the core of epidemiological prediction efforts, but become constrained as more fine-grained compartments are considered. Models like Covasim bring the expressive power of agent-based desciptions of behaviors over networks to allow evaluation of many varieties of intervention strategies and empirical testing against many data sources, but are more difficult to make statistical inferences about.

Well-engineered, open source codebases like Covasim and pyABC allow independent model components to be investigated and incorporated separately. Basic agent-based models can be related directly to traditional ODE compartmental model analogs. The publishing of full model implementation details, like those included with the Panovska-Griffiths publication[2] and code repository provide an excellent example. Careful experimentation building from agent-based models may extend the results into mathematically intractable regimes inaccessible to simpler ODE mdels. The COVID-19 pandemic, and others to come, demand fast-paced scientific sharing that is catalyzed by such interactions.

References

(1) Press, W. H.; Levin, R. C. Modeling, post COVID-19. *Science* **2020**, *370*, 1015–1015.

(2) Panovska-Griffiths, J.; Kerr, C. C.; Stuart, R. M.; Mistry, D.; Klein, D. J.; Viner, R. M.; Bonell, C. Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study. *The Lancet Child & Adolescent Health*

(3) Jewell, N. P.; Lewnard, J. A.; Jewell, B. L. Predictive Mathematical Models of the COVID-19 Pandemic: Underlying Principles and Value of Projections. *JAMA* **2020**, *323*, 1893–1894.

(4) Pearce, N.; Lawlor, D. A.; Brickley, E. B. Comparisons between countries are essential for the control of COVID-19. *International Journal of Epidemiology* **2020**, dyaa108.

(5) Hazelbag, C. M.; Dushoff, J.; Dominic, E. M.; Mthombothi, Z. E.; Delva, W. Calibration of individual-based models to epidemiological data: A systematic review. *PLOS Computational Biology* **2020**, *16*, 1–17.

(6) Papadimitriou, C. Algorithms, complexity, and the sciences. *Proceedings of the National Academy of Sciences* **2014**, *111*, 15881–15887.

(7) Max Roser, E. O.-O., Hannah Ritchie; Hasell, J. Coronavirus Pandemic (COVID-19). *Our World in Data* **2020**, https://ourworldindata.org/coronavirus.

(8) Kerr, C. C. et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. *medRxiv* **2020**,

(9) Hansen, N. The CMA Evolution Strategy: A Tutorial. 2016.

(10) McKinley, T. J.; Vernon, I.; Andrianakis, I.; McCreesh, N.; Oakley, J. E.; Nsubuga, R. N.; Goldstein, M.; White, R. G. Approximate Bayesian Computation and

Simulation-Based Inference for Complex Stochastic Epidemic Models. *Statist. Sci.* **2018**, *33*, 4–18.

(11) Toni, T.; Stumpf, M. P. H. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* **2009**, *26*, 104–110.

(12) Klinger, E.; Rickert, D.; Hasenauer, J. pyABC: distributed, likelihood-free inference. *Bioinformatics* **2018**, *34*, 3591–3593.

List of Figures

ISO3	Country	Pop19	StartDate	TotDays	CloseStart	CloseEnd	Nclose	FracOpen	ISO3	Country	Pop19	StartDate	TotDays	CloseStart	CloseEnd	Nclose	FracOpen
ARE	United_Arab_E	9.77E+6	Mar 10	143	Mar 10	Jul 5	117	0.18	LKA	Sri_Lanka	2.13E+7	Mar 20	133	Jul 13	Jul 27	14	0.89
ARG	Argentina	4.48E+7	Mar 16	137	Mar 16	Jul 31	137	0.00	LVA	Latvia	1.92E+6	Mar 18	135	May 14	Jul 31	78	0.42
BGD	Bangladesh	1.63E+8	Apr 1	121	Apr 1	Jul 31	121	0.00	MAR	Morocco	3.65E+7	Mar 19	134	Mar 19	Jul 31	134	0.00
BGR	Bulgaria	7.00E+6	Mar 16	137	Mar 16	May 21	66	0.52	MDG	Madagascar	2.70E+7	Apr 2	120	Apr 2	Apr 23	21	0.83
BHR	Bahrain	1.64E+6	Mar 5	148	Mar 5	Jul 31	148	0.00	MEX	Mexico	1.28E+8	Mar 16	137	Mar 23	Jul 31	130	0.05
BOL	Bolivia	1.15E+7	Mar 27	126	Mar 27	Jul 31	126	0.00	MMR	Myanmar	5.40E+7	Apr 14	108	Apr 14	Jul 21	98	0.09
CAN	Canada	3.74E+7	Mar 7	146	Mar 20	May 11	52	0.64	MOZ	Mozambique	3.04E+7	Apr 25	97	Apr 25	Jul 31	97	0.00
CHE	Switzerland	8.54E+6	Mar 5	148	Mar 13	May 12	60	0.59	MWI	Malawi	1.86E+7	May 10	82	May 10	Jul 15	66	0.20
CHL	Chile	1.90E+7	Mar 15	138	Mar 15	Jul 2	109	0.21	MYS	Malaysia	3.19E+7	Mar 7	146	Mar 14	Jun 24	102	0.30
COD	Democratic_Re	8.68E+7	Mar 27	126	Mar 27	Jul 31	126	0.00	NGA	Nigeria	2.01E+8	Mar 26	127	Mar 26	Jul 31	127	0.00
COL	Colombia	5.03E+7	Mar 17	136	Mar 17	Jul 31	136	0.00	NLD	Netherlands	1.73E+7	Mar 6	147	Mar 16	May 12	57	0.61
CRI	Costa_Rica	5.05E+6	Mar 19	134	Mar 19	Jul 31	134	0.00	NPL	Nepal	2.86E+7	Apr 27	95	Apr 27	Jul 31	95	0.00
CUB	Cuba	1.13E+7	Mar 26	127	Mar 26	Jul 31	127	0.00	NZL	New_Zealand	4.78E+6	Mar 21	132	Mar 24	Apr 30	37	0.72
DEU	Germany	8.30E+7	Feb 29	153	Mar 18	May 5	48	0.69	OMN	Oman	4.97E+6	Mar 22	131	Mar 22	Jul 31	131	0.00
DNK	Denmark	5.81E+6	Mar 10	143	Mar 13	Apr 16	34	0.76	PAK	Pakistan	2.17E+8	Mar 17	136	Mar 17	Jul 31	136	0.00
DOM	Dominican_Re	1.07E+7	Mar 22	131	Mar 22	Jul 31	131	0.00	PAN	Panama	4.25E+6	Mar 17	136	Mar 17	Jul 31	136	0.00
ECU	Ecuador	1.74E+7	Mar 17	136	Mar 17	Jul 31	136	0.00	PHL	Philippines	1.08E+8	Mar 13	140	Mar 15	Jul 31	138	0.01
EST	Estonia	1.32E+6	Mar 14	139	Mar 16	May 16	61	0.56	POL	Poland	3.80E+7	Mar 14	139	Mar 14	May 26	73	0.47
ETH	Ethiopia	1.12E+8	Apr 8	114	Apr 8	Jul 31	114	0.00	PRT	Portugal	1.03E+7	Mar 12	141	Mar 16	May 19	64	0.55
FRA	France	6.70E+7	Feb 29	153	Mar 16	May 12	57	0.63	PRY	Paraguay	7.04E+6	Mar 27	126	Mar 27	Jul 5	100	0.21
GBR	United_Kingdo	6.66E+7	Mar 3	150	Mar 23	Jun 2	71	0.53	PSE	Palestine	4.98E+6	Mar 22	131	Mar 22	Jul 31	131	0.00
GHA	Ghana	3.04E+7	Mar 25	128	Mar 25	Jun 16	83	0.35	QAT	Qatar	2.83E+6	Mar 12	141	Mar 12	Jul 31	141	0.00
GRC	Greece	1.07E+7	Mar 8	145	Mar 10	May 12	63	0.57	ROU	Romania	1.94E+7	Mar 13	140	Mar 13	Jun 15	94	0.33
HRV	Croatia	4.08E+6	Mar 17	136	Mar 17	May 12	56	0.59	RWA	Rwanda	1.26E+7	Mar 28	125	Mar 28	Jul 31	125	0.00
HUN	Hungary	9.77E+6	Mar 19	134	Mar 19	May 4	46	0.66	SAU	Saudi_Arabia	3.43E+7	Mar 13	140	Mar 13	Jul 31	140	0.00
IDN	Indonesia	2.71E+8	Mar 14	139	Mar 26	Jun 30	96	0.31	SEN	Senegal	1.63E+7	Mar 22	131	Mar 22	Jun 25	95	0.27
IND	India	1.37E+9	Mar 12	141	Mar 18	Jul 31	135	0.04	SGP	Singapore	5.80E+6	Feb 14	168	Apr 8	Jun 2	55	0.67
IRL	Ireland	4.90E+6	Mar 13	140	Mar 13	Jun 26	105	0.25	SLV	El_Salvador	6.45E+6	Apr 4	118	Apr 4	Jul 31	118	0.00
IRN	Iran	8.29E+7	Feb 25	157	Mar 5	May 16	72	0.54	SRB	Serbia	6.96E+6	Mar 16	137	Mar 16	May 17	62	0.55
IRQ	Iraq	3.93E+7	Mar 8	145	Mar 8	Jul 31	145	0.00	SVK	Slovakia	5.45E+6	Mar 16	137	Mar 16	Jun 2	78	0.43
ISR	Israel	8.52E+6	Mar 11	142	Mar 13	May 4	52	0.63	TGO	Togo	8.08E+6	Apr 7	115	Apr 7	Jun 9	63	0.45
ITA	Italy	6.04E+7	Feb 23	159	Mar 4	Jul 31	149	0.06	THA	Thailand	6.96E+7	Mar 11	142	Mar 18	Jun 15	89	0.37
JAM	Jamaica	2.95E+6	Apr 4	118	Apr 4	Jun 8	65	0.45	TUN	Tunisia	1.17E+7	Mar 21	132	Mar 21	May 28	68	0.48
KAZ	Kazakhstan	1.86E+7	Mar 21	132	Mar 21	Jul 31	132	0.00	TUR	Turkey	8.20E+7	Mar 18	135	Mar 18	Jul 31	135	0.00
KEN	Kenya	5.26E+7	Apr 1	121	Apr 1	Jul 31	121	0.00	UGA	Uganda	4.43E+7	Apr 6	116	Apr 6	Jul 31	116	0.00
KOR	South_Korea	5.12E+7	Feb 20	162	Feb 24	May 20	86	0.47	UKR	Ukraine	4.40E+7	Mar 24	129	Mar 24	May 25	62	0.52
KWT	Kuwait	4.21E+6	Mar 3	150	Mar 3	Jul 31	150	0.00	URY	Uruguay	3.46E+6	Mar 19	134	Mar 19	Apr 23	35	0.74
LBY	Libya	6.78E+6	Apr 21	101	Apr 21	Jul 31	101	0.00	ZMB	Zambia	1.79E+7	Apr 18	104	Apr 18	Jun 1	44	0.58
LKA	Sri_Lanka	2.13E+7	Mar 20	133	Jul 13	Jul 27	14	0.89	ZWE	Zimbabwe	1.46E+7	May 22	70	Jul 8	Jul 31	23	0.67

Figure 1: Country summary statistics

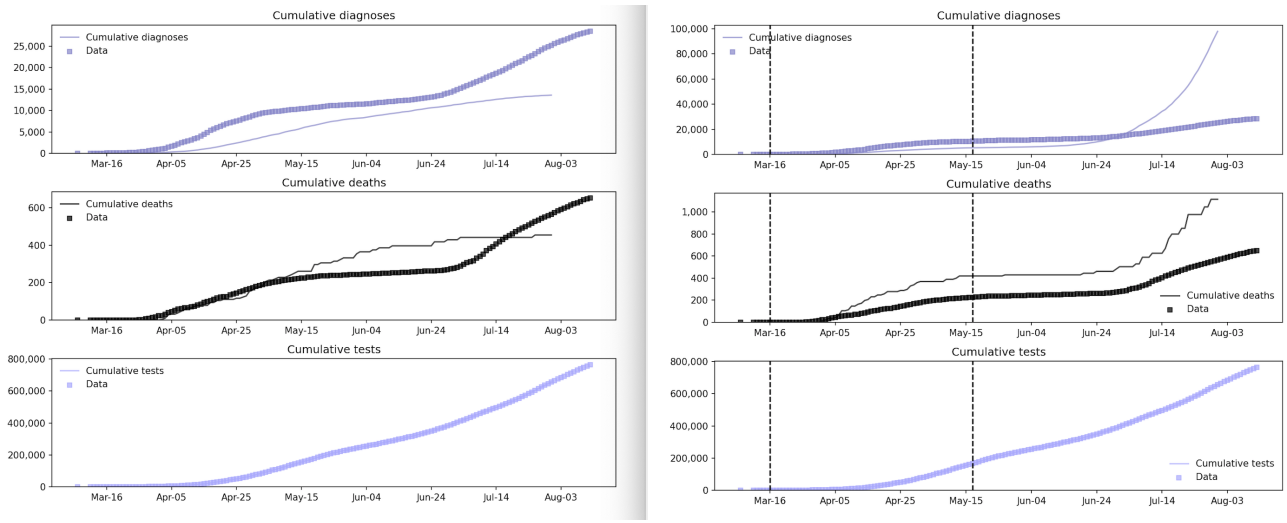


Figure 2: Model predictions for SRB, without and with intervention

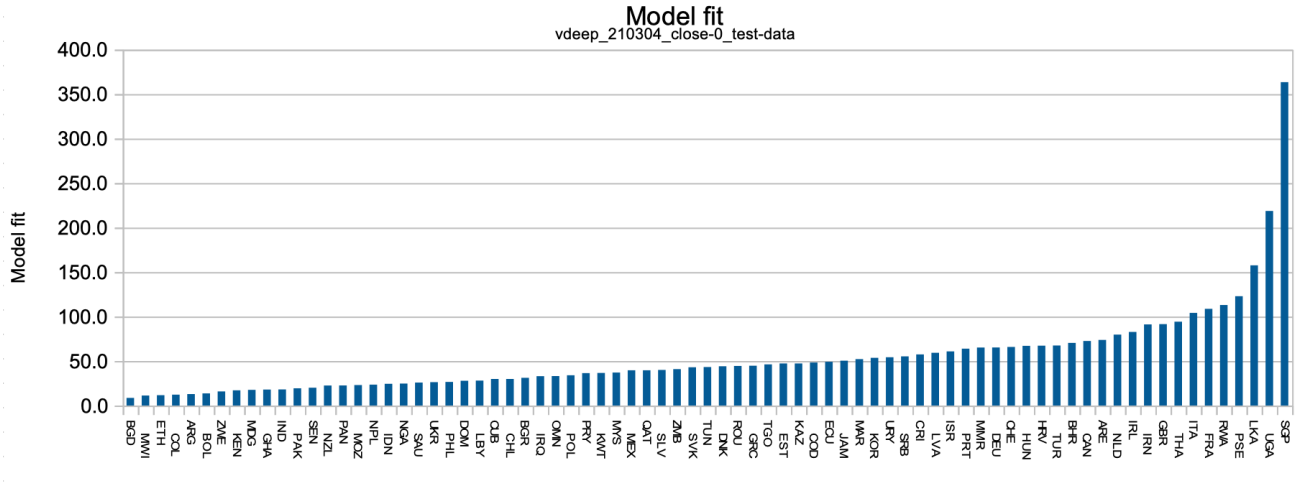


Figure 3: Base model fits

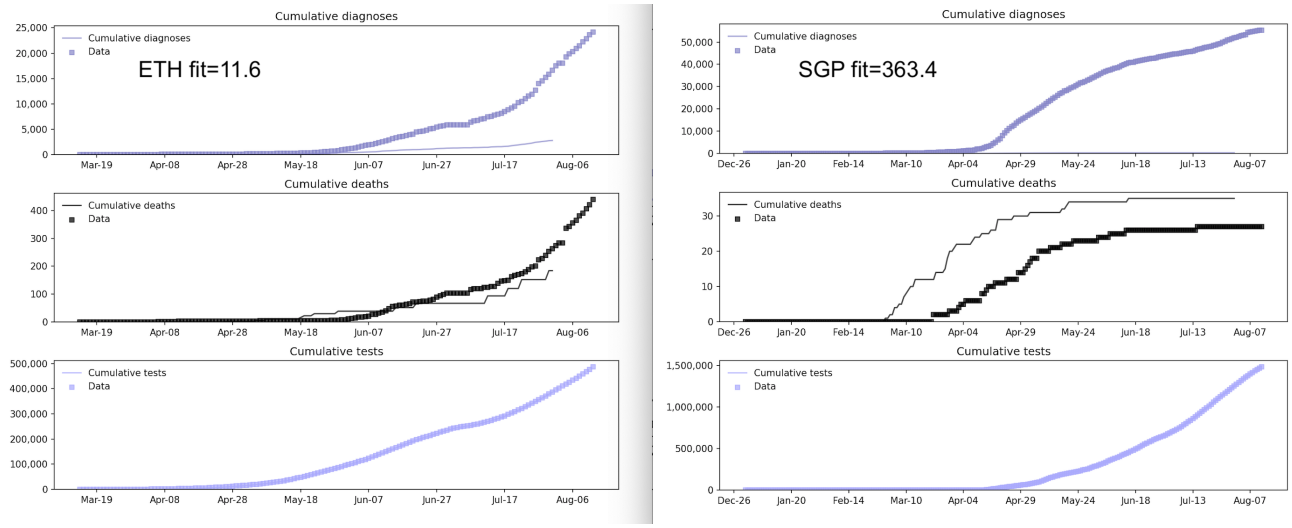


Figure 4: Base model: COL and SGP

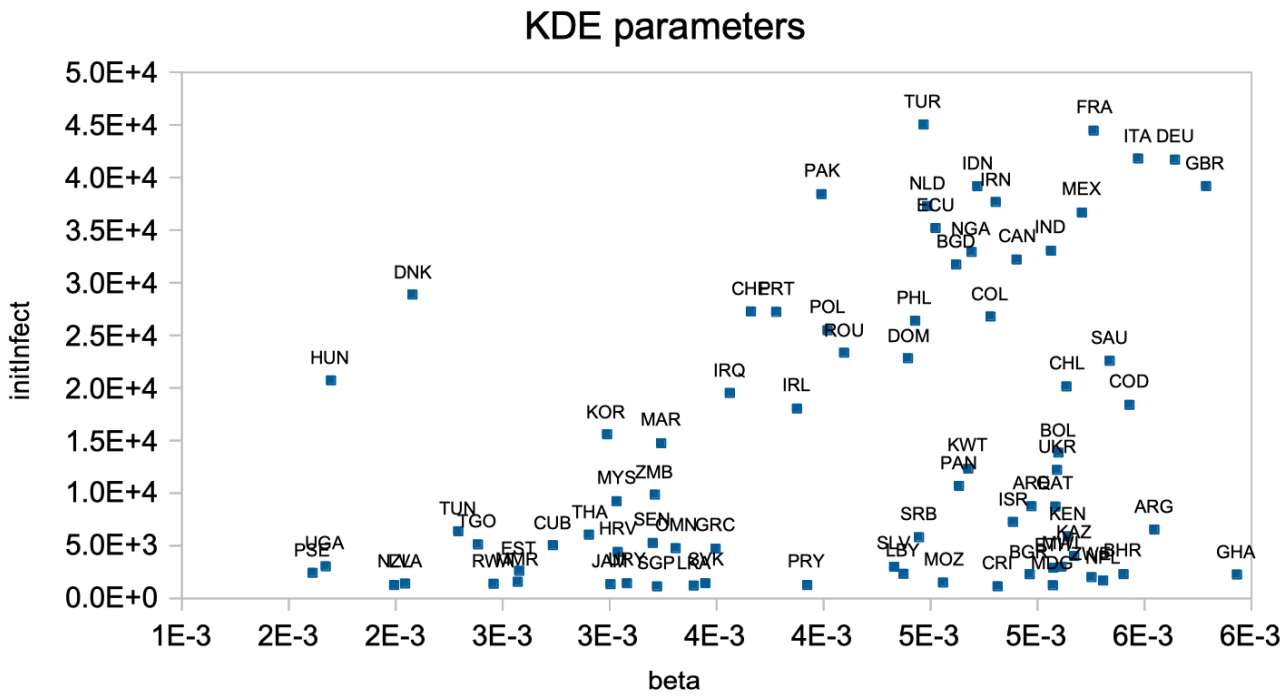


Figure 5: Base model key parameter distribution

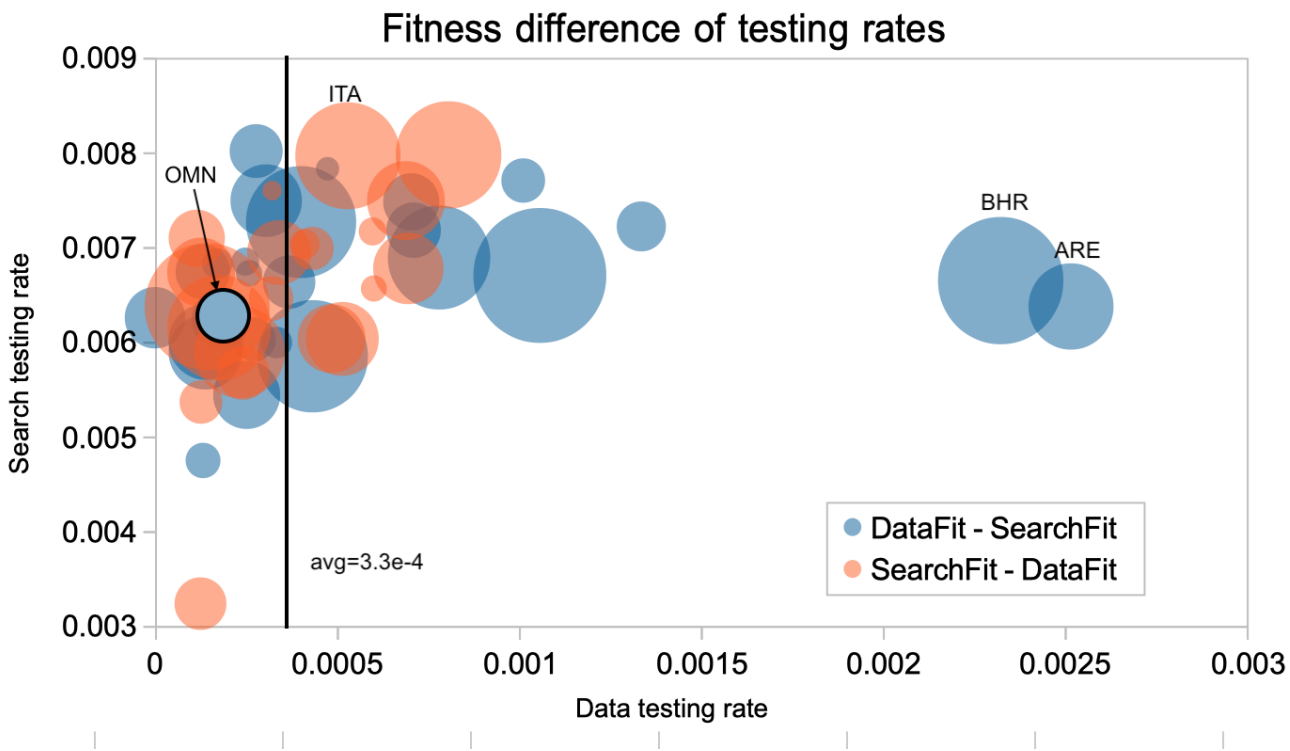


Figure 6: Testing rate from data vs. calibration

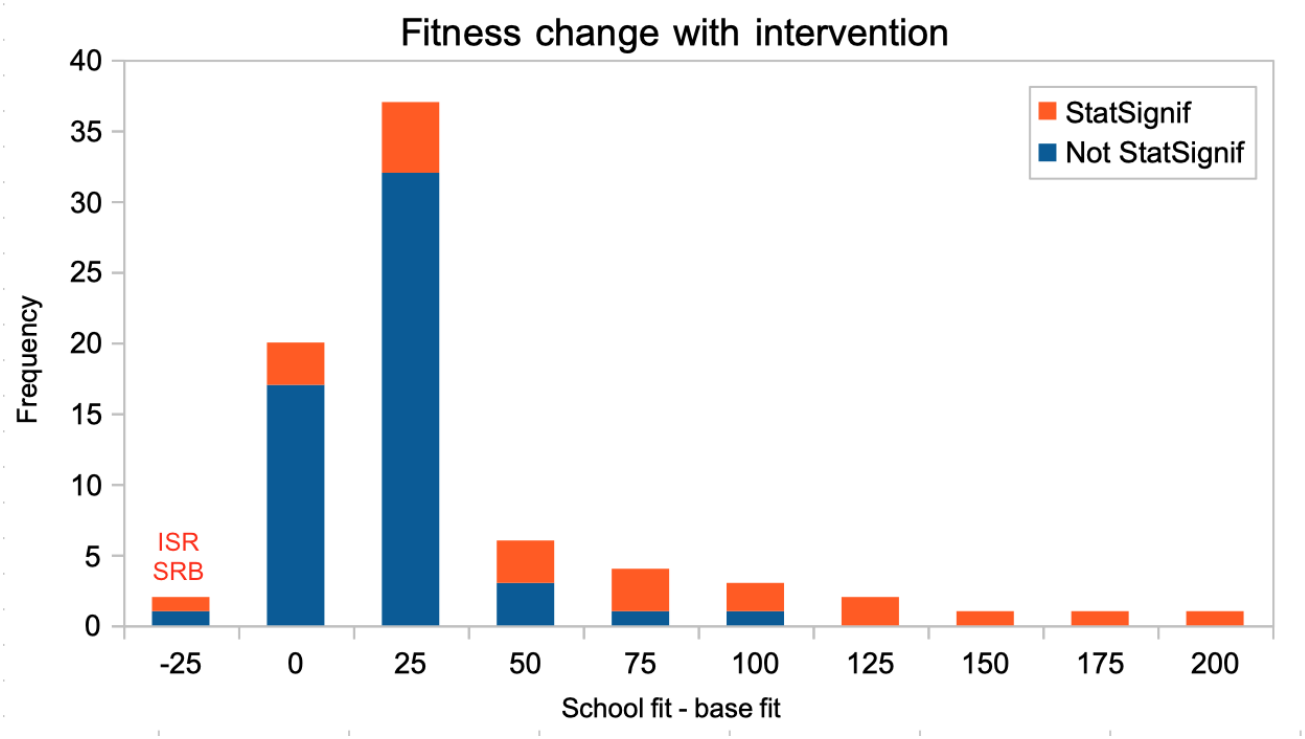


Figure 7: Change in fitness with intervention

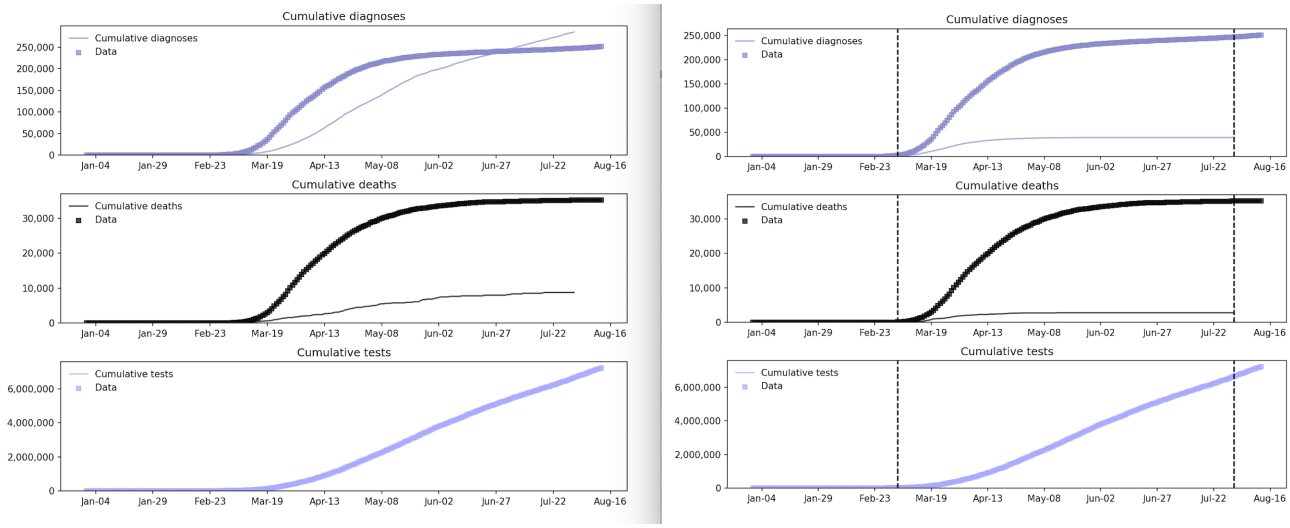


Figure 8: Intervention change ITA

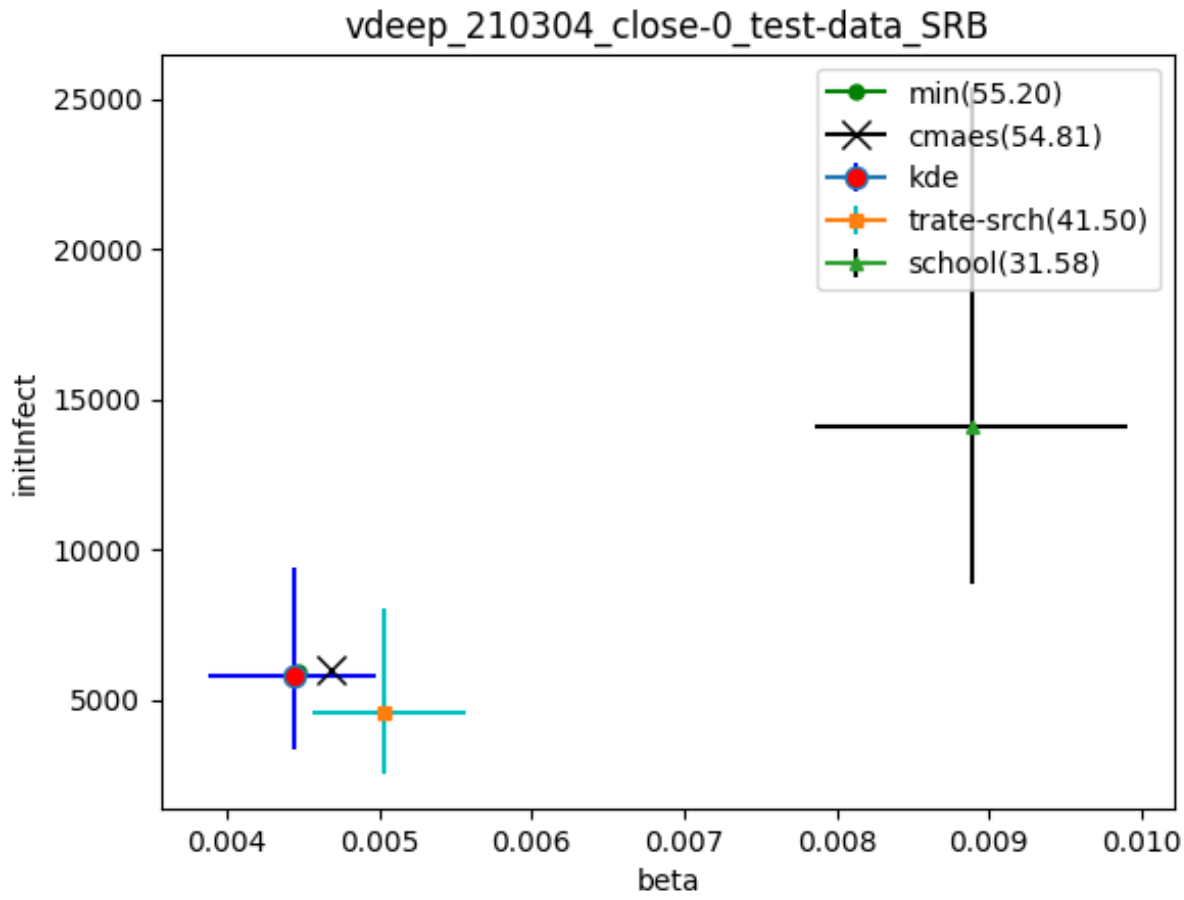


Figure 9: Calibrated parameters with credible intervals - SRB

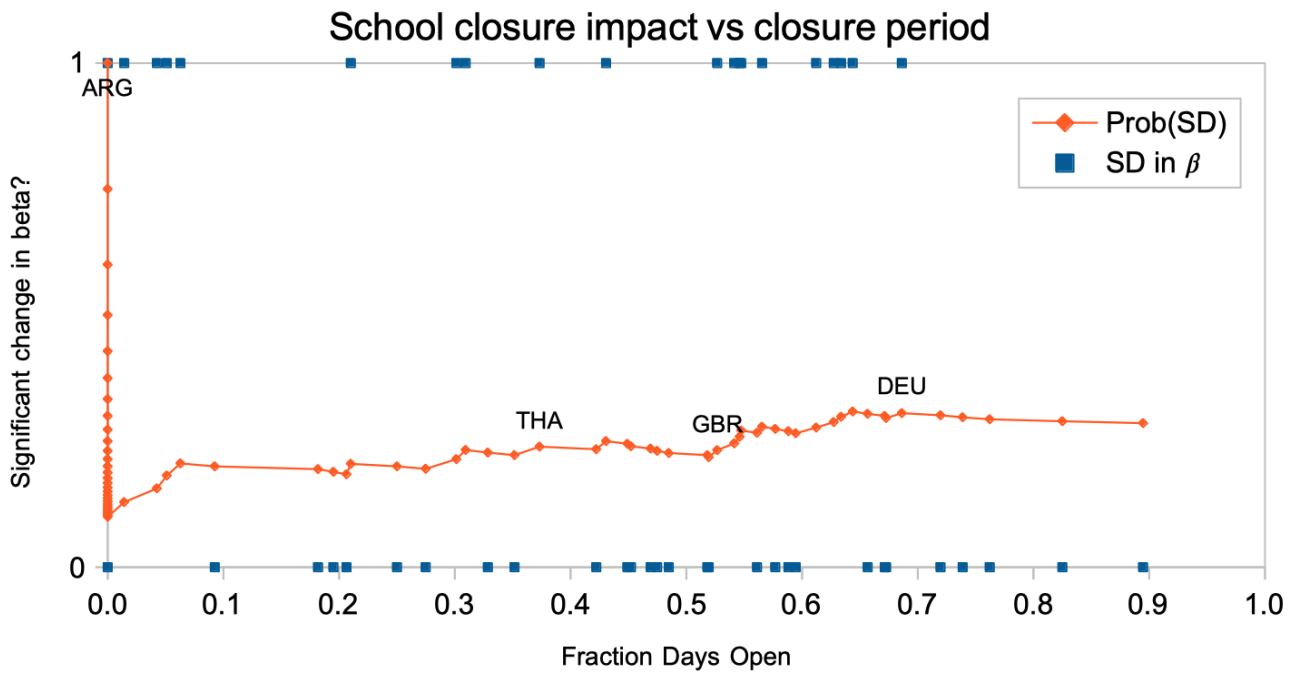


Figure 10: Impact of closiure period on statistical significance

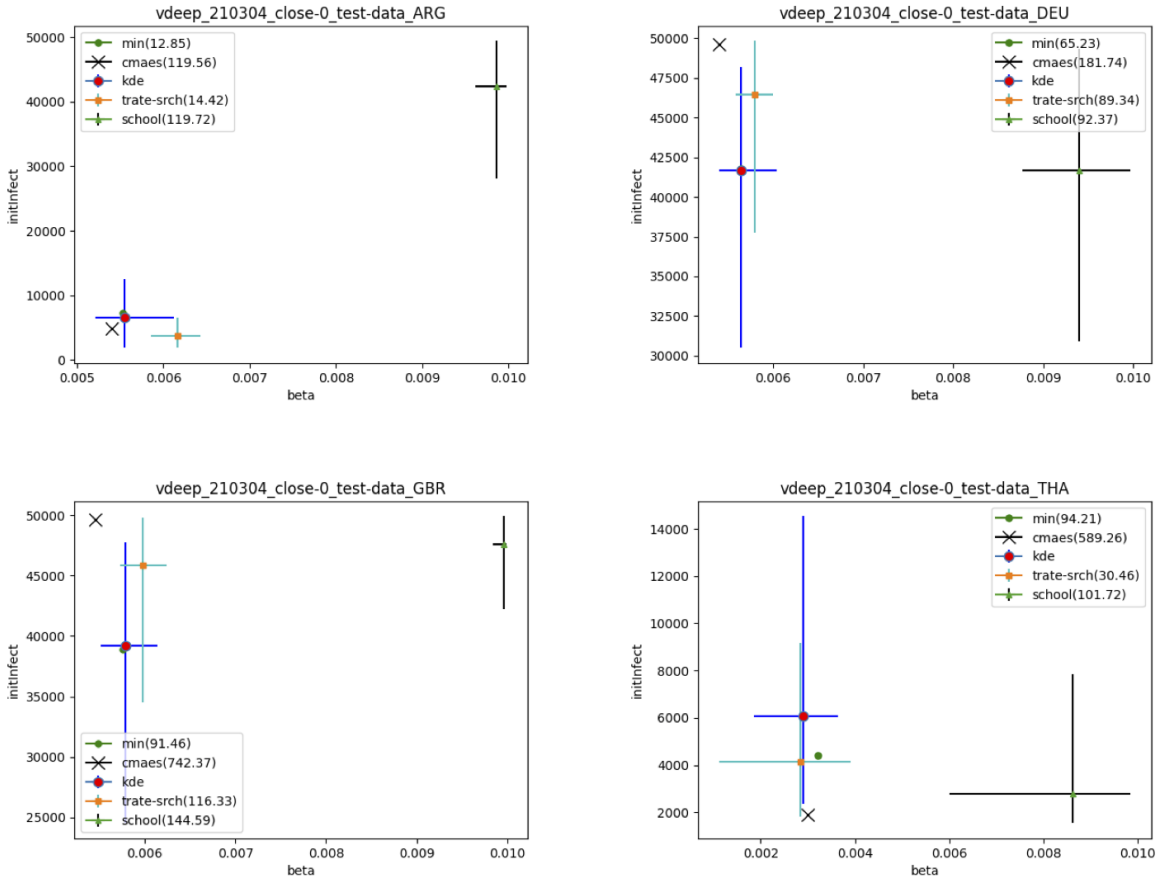


Figure 11: Examples of statistical significance in β

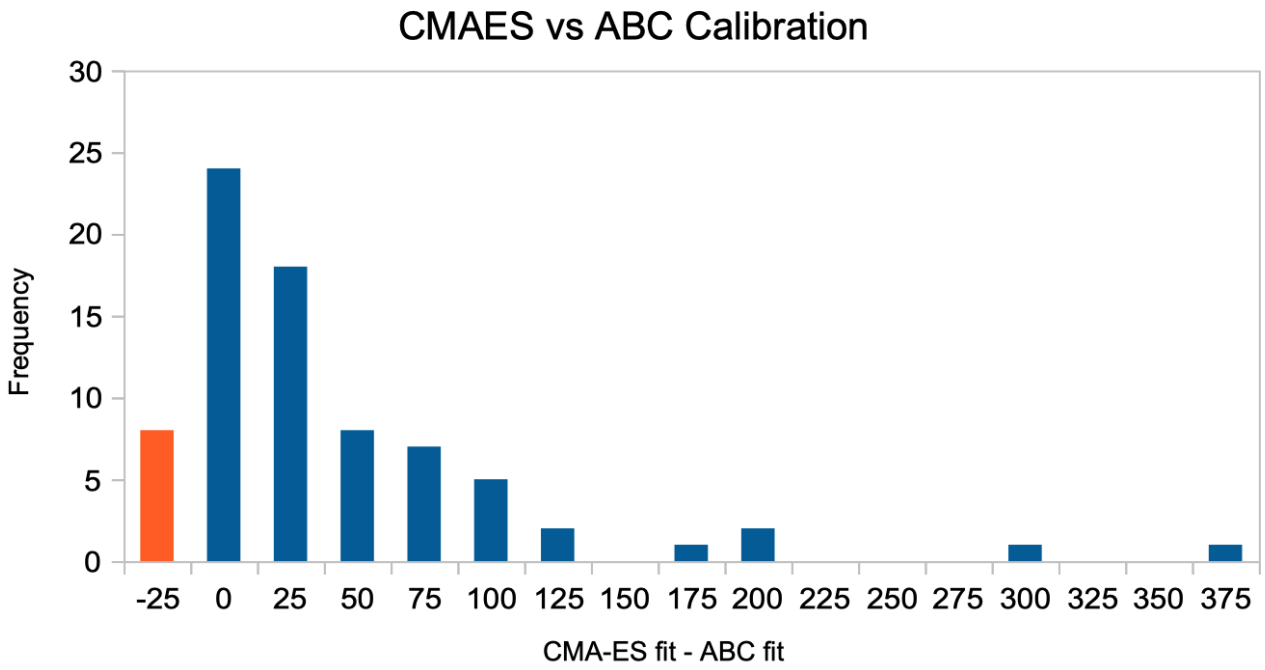


Figure 12: CMA-ES vs ABC search

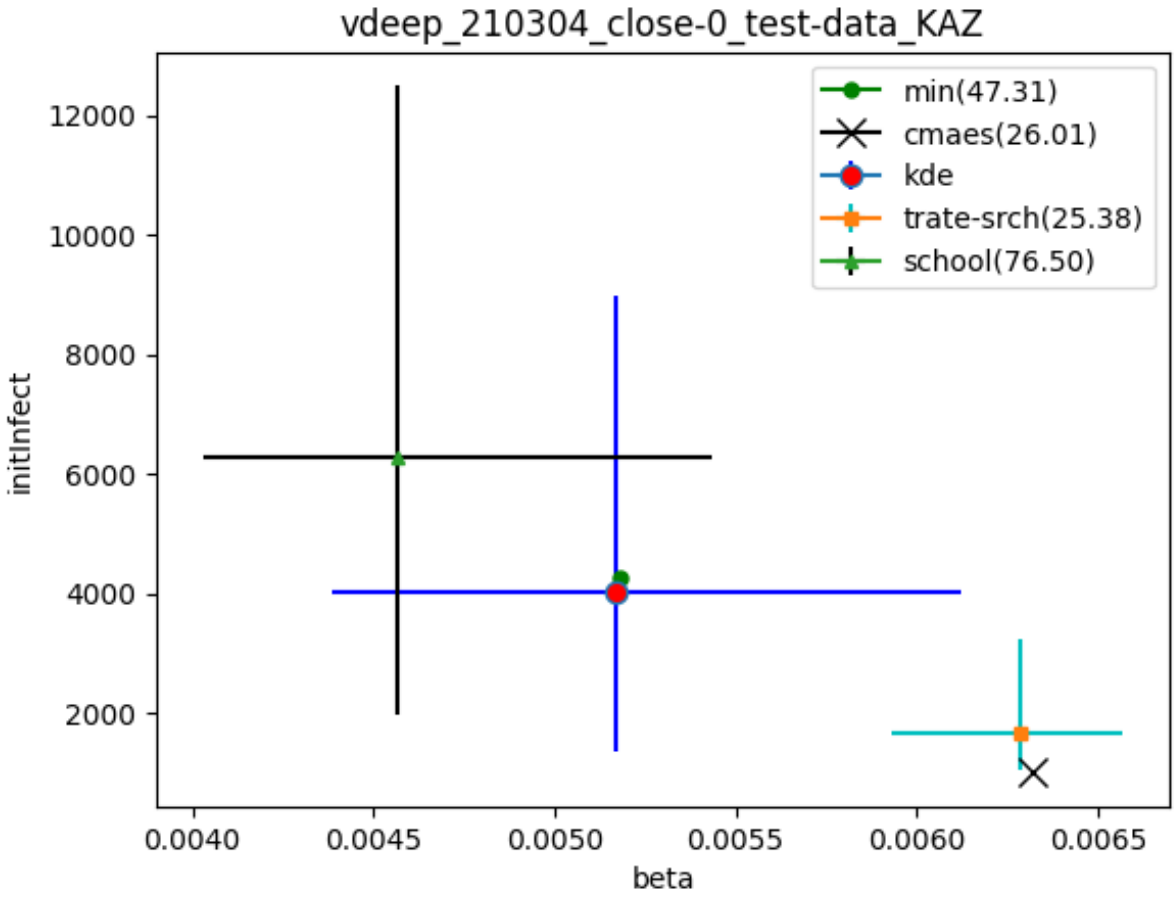


Figure 13: CMA-ES alternative - KAZ

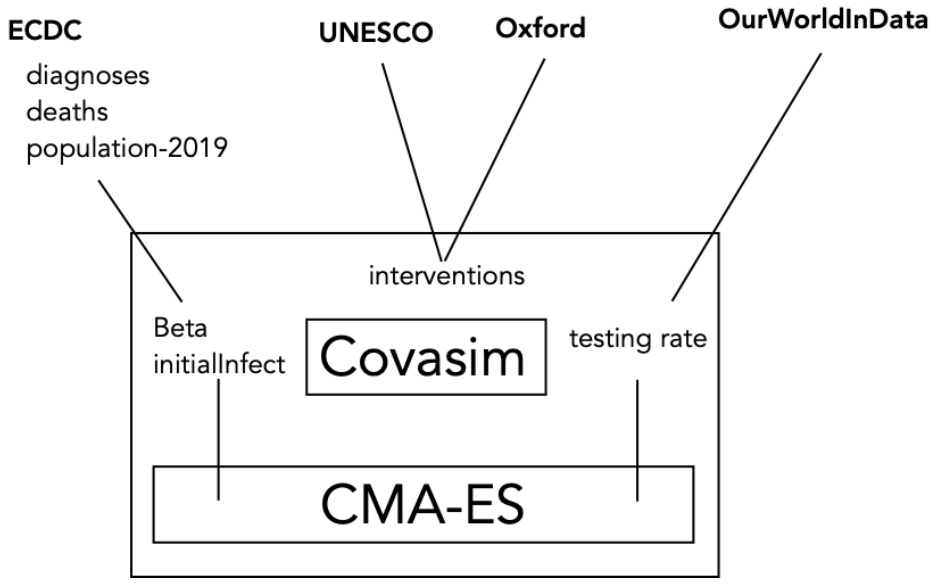


Figure 14: Data sources

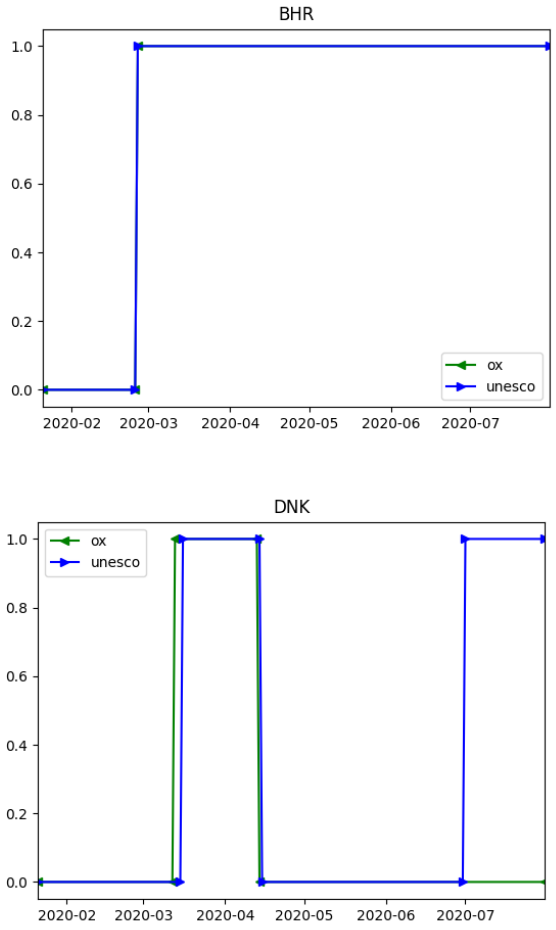
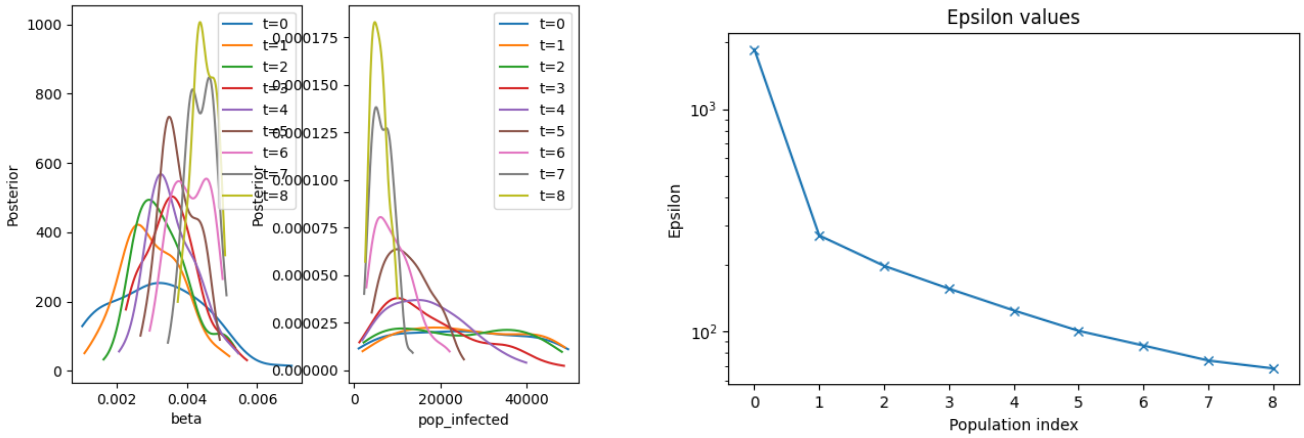


Figure 15: Overlapping intervention data



(a) Posterior history

(b) Epsilon convergence

Figure 16: pyABC search for SRB parameters