

Generalized modeling of interventions across 136 countries

Richard K. Belew,^{*,†} Cliff Kerr,^{*,‡} Jasmina Panovska-Griffiths,^{*,¶} and Dina Mistry^{*,‡}

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

[‡]*Institute for Disease Modeling, Bellevue, WA*

[¶]*University College London*

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

Draft: 25 December 2020

Abstract

COVID-19 continues to spread around the world and modeling plays an important role in informing policy¹. 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 there². We extend this methodology to consider data from 136 countries for which intervention data is available to retrospectively model interventions employed 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. For 87 of the countries evaluated there is also data concerning testing rates in each country that can be compared to other estimates.

A Covasim model was calibrated for each of 136 countries, with each country's population, age distribution, disease emergence and school closing dates specified, then calibrated against data concerning COVID-19 diagnosed cases and related deaths. All countries' simulations were run with and without school interventions. The hypothesis considered is: Does incorporating the additional knowledge of school interventions into a model makes it a better predictor of data concerning diagnoses and deaths?

Because assumptions regarding background testing rates being performed are critical for interpretation of statistics regarding the number of diagnosed patients, three conditions were considered: assume the same *fixed* testing rated for all countries; *searching* for an estimated testing rate as part of model calibration; and use *data* for the 87 countries that have this data available.

An analysis of the global optimization used to calibrate these models to parameters is considered in detail. General lessons for using and interpreting automatic parameter calibration methods are suggested. The main difficulty, of course, is in establishing what makes one model a *significantly* better predictor. For now at least we can *not* reject the null hypothesis: As a general rule, adding school interventions can not be shown to improve the model.

1 Introduction

COVID-19 continues to spread around the world and modeling plays an important role in informing policy¹. 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-19³

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

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 currently available, only models at the level of individual countries and interventions ordered nationally are considered here.

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. Because global optimization of simulation parameters is a common aspect of many modeling efforts⁵, we evaluate the optimization process implemented using the CMA-ES process in some detail. 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.

2 Results

The first section below reports on “base model” experiments that do *not* include school intervention and which use a *fixed* testing rate (=3.3e-4, the average across all countries with testing data; cf. Section 2.2). Experiments contrasting models searching for the testing

rate, and ones using data for testing rate are in Section 2.2. Experiments contrasting the base model against simulations including school interventions are considered in Section 2.3.

2.1 Base model

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 1 shows the range of resulting model fits across all 136 countries, with lower fitness scores assigned to simulations with a better match of predictions for the number of diagnoses and deaths. There is not any absolute criterion as to what makes for a “good” fit, but 130 countries have fit values less than 500.

Figure 2 shows the result of the simulation for two countries, COL and LAO. 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 simply shows the number of tests that would have been given over time, assuming the fixed testing rate and country’s population; it isn’t useful in this basic model but will provide contrast in Sections 2.2. COL and LAO were selected as being a particularly good “fit” with the data (COL), and particularly bad (LAO). Note that this should be considered the “best fit found” with respect to the calibration optimization over the two free parameters, β and INITINFECT, the number of initial infections, and using sum-squared-error as the measure of mismatch.

2.2 Using available data for testing rate

For 87 countries there exists data on testing rates via OWID. Experiments using this source for testing rates can be compared to using a fixed testing rate, and to rates found via calibration.

2.2.1 Comparing data against fixed testing rate

Using data-based testing rates vs. the fixed rate is as likely to improve as harm the model. As shown in Figure 3, while there seems a general trend towards increased testing rates improving fit, the correlation is not strong ($r^2 = 0.37$).

2.2.2 Comparing data against calibrated testing rate

Figure 4 shows the log ratio $\log(\frac{\text{testRate}_{\text{data}}}{\text{testRate}_{\text{search}}})$ of the testing rate from data vs. search for all countries. In most cases the testing rate identified via search is lower, and sometimes much lower, than that provided in data; i.e., most countries are below the line showing the ratio=1.0.

Three countries, BGR,CHE and NGA, sampled from across the range of results, have been highlighted. Comparisons of model results for these countries is shown in Figure 5, and details are provided in Table 1. Figure 6 shows how *daily* testing data compares to the constant (average) testing *rate* identified via search for BGR; note that BGR has a data/search testing rate ratio near 1.0. Nevertheless search's slightly higher testing rate allows an improved value due to a better match with diagnosis data. In the CHE models, using data for testing rates improved the fit considerably (although barely noticeable in the plotted curves). With NGA, on the other hand, using a much larger testing rate than in the data finds only a slightly better model fit.

Table 1: Search vs. data for testing rate - example countries

		Search				Data			
ISO3	trRatio	infect	beta	testrate	value	infect	beta	testrate	value
BGR	1.07E+0	1418	5.34E-3	2.42E-4	231	1539	5.13E-3	2.59E-4	131
CHE	3.85E+0	42012	3.70E-3	1.12E-4	576	34946	4.00E-3	4.33E-4	62
NGA	2.74E-3	32249	4.62E-3	2.59E-3	193	9740	4.81E-3	7.10E-6	225

Overall, using data-based testing rates with a model improved their fit somewhat, in comparison to the value found via calibration search; average improvement $\sim 100^1$. As shown

¹Note that “improvement” by data-based models is reflected in negative differences in value, since smaller

in Figure 7, while there seems a general trend towards increased testing rates improving fit, the correlation is not strong ($r^2 = 0.34$).

Note that the difference between predicted testing rate and data regarding was not used by the calibration process to guide the search for model parameters.

2.3 Incorporating school closure interventions

The next set of experiments contrast runs with and without school closure interventions modeled. Figure 8 shows the curves for two countries, ITA and GBR for which incorporation of school interventions (base runs on the left, interventions on the right) improves their fit. The dashed lines in the right-hand figures reflect the beginning and ending dates of the interventions.

Figure 9 shows the range of differences in fit value for all countries, from improved fits (left end) to models that had worse fit (right end). Countries where both models (with or without school interventions) had very poor fit scores > 1000 have been highlighted in red. In these cases, both models are considered broken and contrasting them tells us little. For the large majority of countries ($n=120$), the difference between the two models was small, < 100 . Figure 10 contrasts the results for two other countries, FRA and CAN, for which adding interventions to the model harmed their fit to data. Details of all four runs are included in Table 2.

Table 2: With and without interventions modeled - example countries

ISO3	None			Intervention		
	infect	beta	value	infect	beta	value
CAN	49979	5.12E-3	168	49916	9.11E-3	409
FRA	49692	5.84E-3	581	49739	1.04E-2	853
GBR	49869	6.02E-3	525	49980	1.30E-2	274
ITA	46062	5.66E-3	632	49249	1.72E-2	232

values reflect better models.

2.3.1 Interpretation

The changes related to modeling school interventions (vs. not) described above were based on the simplest alternative regarding testing rates, viz. using the same, fixed test rate (3.3e-4). The same basic pattern is shown across the two variations on the 87 countries where data on testing rates was used instead.

While it is tempting to take positive differences in data like that from ITA and GBR showing that additional modeling effort seems to improve the model, but consideration of the differences in model parameters `initInfect` and `beta` makes this seem suspect. The changes in `initInfect` are all quite small, relative to the upper bound of the allowed calibration search space. The role of initial boundaries for calibration are discussed in Section 2.4. In all cases `beta` has increased in the intervention models.

That no clear impact of modeling school interventions can be observed is at least somewhat surprising. A key feature of Covasim is its separation of individuals into layers with distinctly intra-layer connection networks. As modeled here, school interventions make significant changes to all four interaction layers (cf. Table 3), and still don't change much.

2.4 Model calibration

2.4.1 Global optimization

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 here⁵. 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 warns there can be no guarantee of finding the actual optimum value⁶. In the context of modeling, the use of calibration to identify some parameter

values generates several new questions:

1. Since we may have multiple criteria for good solutions, how can we determine the Pareto-optimal combination? For example caring to match both death data and diagnosis data, how do we determine the Pareto optimal combination?
2. How much effort should be put into the search? Ie, how much value is there in extending computational effort towards improved results? What sort of reward is given for deeper search?
3. Conversely, what makes a solution “good enough”?

For most of the experiments reported here, the answers to these are:

1. A simple weighted sum of deaths and diagnosis mismatch is used, with deaths counting twice as heavily
2. A constant search effort (population size=32, 25 generations) was allowed the CMAEvolution optimizer
3. There is no early/late termination based on the quality of solutions found

In this section we describe experiments using many more trial and search ranges focused on a single country, AUS.

2.4.2 Search dynamics

In the experiments reported here we will generate a sample of 2500 trials of particular parameter pairs of initial infection and beta. The samples were generated by a CMAES search initiated with the bounds shown and initialized at the location of the red plus sign.

On the left is a plot of the value surface, showing all sample points with darker blues indicating better fit values, a red plus sign showing the initial search parameter, and a green X showing the labeled best value found. Looking closely you can see several good value “ridges”

that have been sampled more densely. On the right is a more detailed contour plot showing an interpolated surface across these samples. It shows a highly textured, multi-modal region near the discovered optimum.

Convergence to value optimum The samples were generated adaptively, towards better solutions across CMA-ES populations of size 100, in a series of 25 generations. The progression of this search process can be tracked according to several key statistics as they vary over generations.

A first statistic is the model fit value parameter shown In Figure 12. This shows fit value average, together with standard deviation bars. The later generations 10-25 have been broken out in an inset graph, the better to see the smaller values. After dramatic early improvements, model fit does not improve much after about 12 generations.

Beta leads initialInfections

Figure 13 considers the values of beta and initial infection independently. Values of beta converge more quickly and have stabilized by about generation 10. Values of initial infection continue to increase throughout all generations. Note in Figure 12, however, that increasing initial infections does not change model fit much.

Deaths dominate diagnoses

Figure 14 looks at the component measures on which model fits is evaluated: match with death data and match with diagnosis data. As expected, because it is weighted more heavily, matching death data drives the search. It is interesting to note that improvement with respect to in the two separate values (diagnoses and death) is well correlated across the generations of the search.

2.4.3 Bounding search

As mentioned earlier, global optimization procedures generally require bounds within which they are to search. Figure 15 can be compared to that originally shown in Figure 11. Here

a range of initial infection sizes twice as large (50000 vs 25000) was allowed in the search. Again, the first plot shows similar ridges leading from the initial starting value, to an optimum with about the same number of initial infections but a smaller beta. A second, vertical ridge is indicated leading to the very top boundary imposed by the a priori constraint.

Back-of-the-envelope calculations regarding possible values for initial infections² suggests that the second, upper attractor containing the optimum in this experiment should be discounted as implausible.

There is a clear bifurcation shown in the second contour plot. The lower upper bound on initial infection (25000) was selected to force search to include only the lower attractor.

3 Methods

3.1 Data sources

ECDC used as primary organizing data source. ISO3 codes were used to identify and merge country data.³

- ECDC
- OWID testing⁷
- UNESCO interventions
- OxCGRT interventions
- Age distribution data from the Neher Lab as distributed as part of Covasim⁴

²r0 = number of contacts * days infectious * infection probability per contact per day so if we assume r0 = 2, 20 contacts, 8 days infectious, we get beta = 0.0125 [Cliff, 3 Oct]

³Two countries missing from ECDC lists, TWN and HNG. 2019 population populations for these two was obtained from 'pop19': 23773876 }WorldOMeters

⁴cf. `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 17 shows two countries' examples: MNG, a country with nearly overlapping accounts; and UZB 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 > 14 days were considered. 136 countries had school interventions that satisfied these criteria.

3.1.2 Testing rate data

Testing data from OurWorldInData⁷ 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.

Table 18 summarizes the 136 countries included in this study: their name, ISO-3 code,

and whether or not they had testing rate data available about them.

3.2 Covasim

Covasim⁸ 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.

As data regarding deaths is believed to generally more accurate than that about positive tests, 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 again enhanced or attenuated. Following², 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 basic Covasim values and the result of school interventions are shown in Table 3.

3.3 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

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

“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 the “calibration” outer-loop around the basic Covasim simulation to search for key parameter values that allow the model to best “fit” against data. Preliminary experiments across a large number of trials on a single country (cf. Section 2.4) identified 25 generations of 32 individuals each as being reasonable; i.e., 800 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 INITINFECTION, the number of initial infections. Based on the experiments reported in Section 2.4, initial search bounds for the optimization were set:

- initial infected: 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 4, a third parameter corresponding to testing rate is also included. Bounds on its search were established using the data for the 87 countries available:

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

4 Next steps

- Use TPE (Tree-structured Parzen Estimator) optimization ala Optuna

4.1 Model validation

1. {lauerReich20}: Infectious Disease Forecasting for Public Health¹⁰
2. {bergmeir18}:A note on the validity of cross-validation for evaluating autoregressive time series prediction¹¹
3. {bergmeirBenitez12}: On the use of cross-validation for time series predictor evaluation¹²

4.2 Integration of genomic data: molecular epidemiology

- Rockett¹³
 1. examine the added value of near real-time genome sequencing of SARS-CoV-2 in a subpopulation of infected patients during the first 10 weeks of COVID-19 containment in Australia and compare findings from genomic surveillance with predictions of a computational agent-based model (ABM)
 2. based on AceMOD
 3. 21 January and 28 March 2020, 1,617 cases of COVID- 19 were diagnosed and reported to the NSW Ministry of Health. All patients resided in metropolitan Sydney.

5 Conclusions

5.1 Lessons learned

- Be careful with standardization of country naming and data:

- Note excluded countries from the ECDC lists: TWN and HNG
- cf `covasim.normalizeCVAgeData()`
- Select which parameters you choose to calibrate, and their initial bounds, carefully.
- Differences in Covasim value/fit < 100 are not significant; larger differences may not be either.

5.2 Honest priors

Efforts towards placing models like the ones considered here on firm statistical foundations require better answers to a number of questions:

- How do the bounds established on global optimization, and artifacts of the global search process, impact our understanding of the priors they impose on model results?
- How best to capture the sensitivity of our results given uncertainty in testing rates?
- What accuracy can we claim for model parameter estimation?

Promising approaches include Approximate Bayesian Computation (ABC) and history matching¹⁴.

5.3 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 descriptions 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.

The well-engineered, open source Covasim codebase allows independent model components to be investigated and incorporated separately. Basic agent-based models can be related directly to compartment model analogs. Careful experimentation building from these may be able to extend the results into mathematically intractable regimes. The publishing of full model implementation details, like that included with the Panovska-Griffiths publication² and code repository is an excellent example. 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) Panowska-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) Lauer, S. A.; Brown, A. C.; Reich, N. G. Infectious Disease Forecasting for Public Health. *arXiv preprint arXiv:2006.00073* **2020**,

- (11) Bergmeir, C.; Hyndman, R. J.; Koo, B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis* **2018**, *120*, 70 – 83.
- (12) Bergmeir, C.; Benítez, J. M. On the use of cross-validation for time series predictor evaluation. *Information Sciences* **2012**, *191*, 192 – 213, Data Mining for Software Trustworthiness.
- (13) Rockett, R. J. et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. *Nature Medicine* **2020**,
- (14) 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.

List of Figures

1	Base model fits	20
2	Base model: COL and LAO	21
3	Model fit related to data vs. fixed testing rate	22
4	Search vs. data for testing rate - all countries	23
5	Search vs. data testing rates for BGR, CHE and NGA	24
6	Testing rate from search vs. data for BGR	25
7	Model fit related to data/search testing rate	26
8	Interventions improving models	27
9	Improved fit	28
10	Interventions harming models	29
11	Value surface	30
12	Value convergence	31
13	Parameter convergence	32
14	Independent deaths and diagnoses	33
15	Values across larger bounded region	34
16	Data sources	35
17	Overlapping intervention data	36
18	Countries' data	37

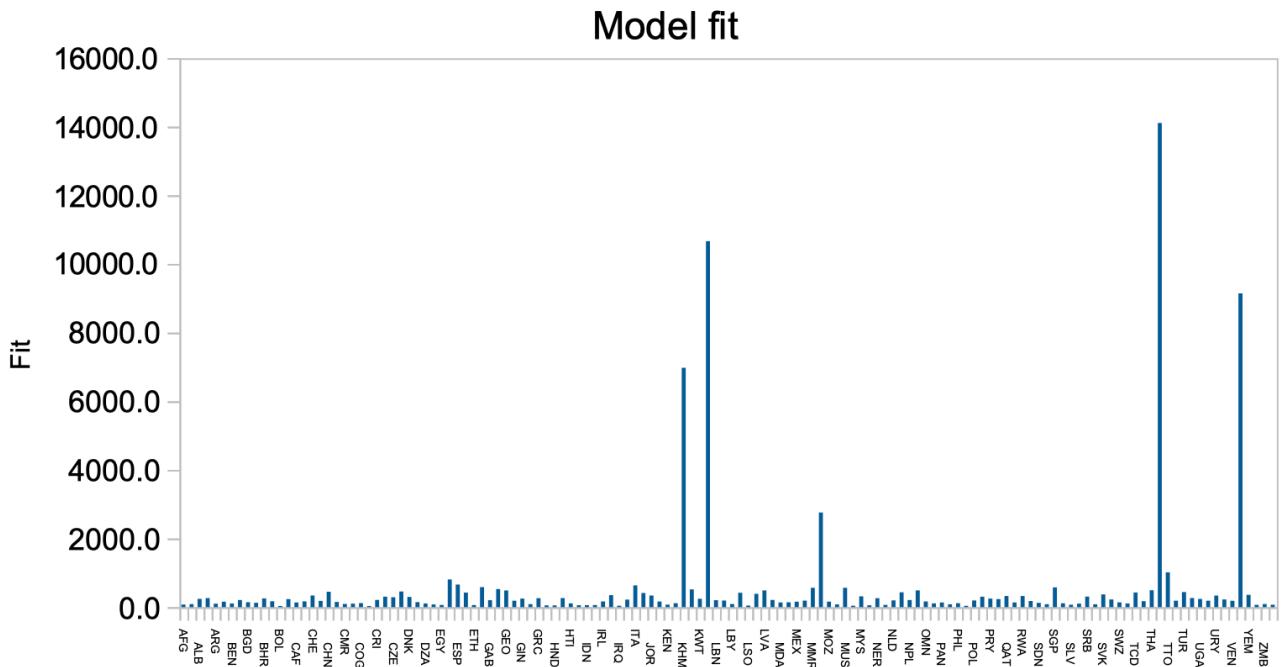


Figure 1: Base model fits

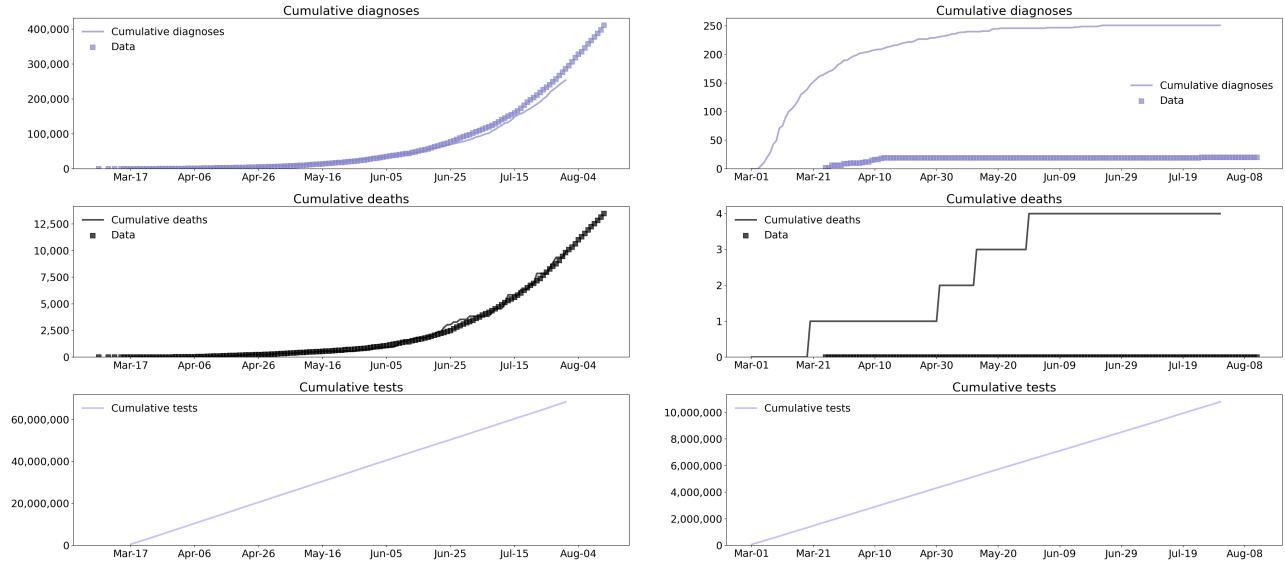


Figure 2: Base model: COL and LAO

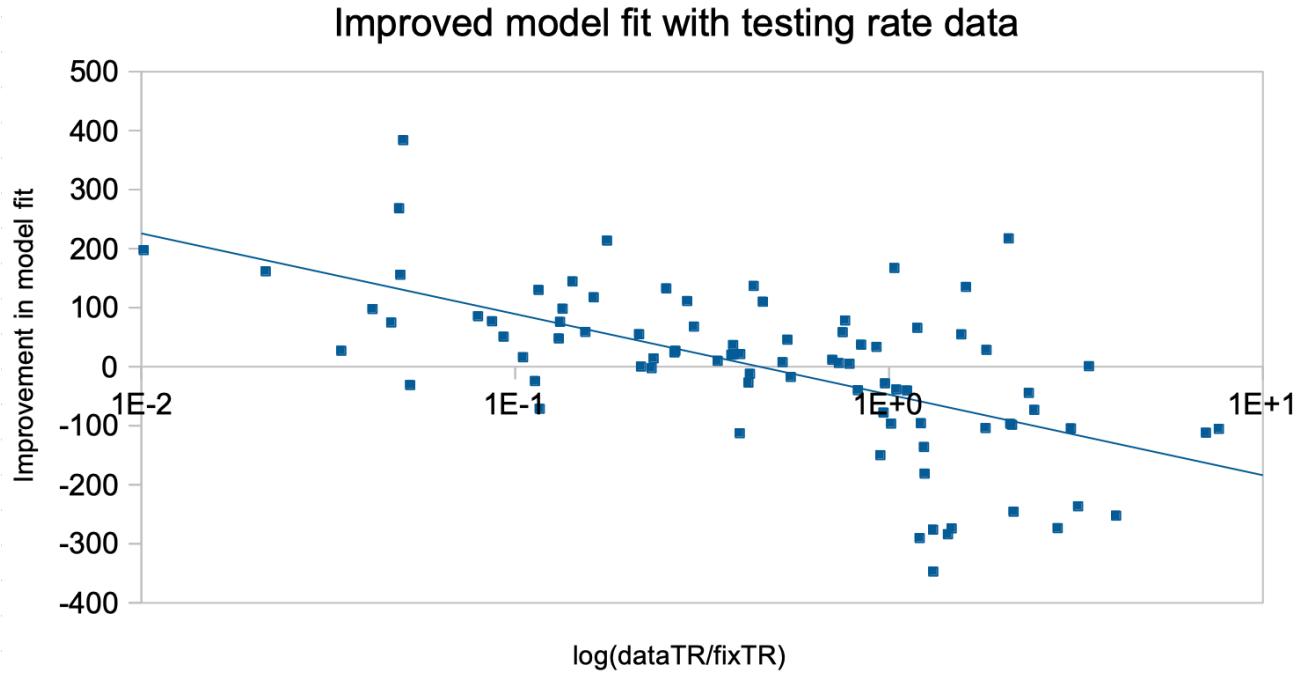


Figure 3: Model fit related to data vs. fixed testing rate

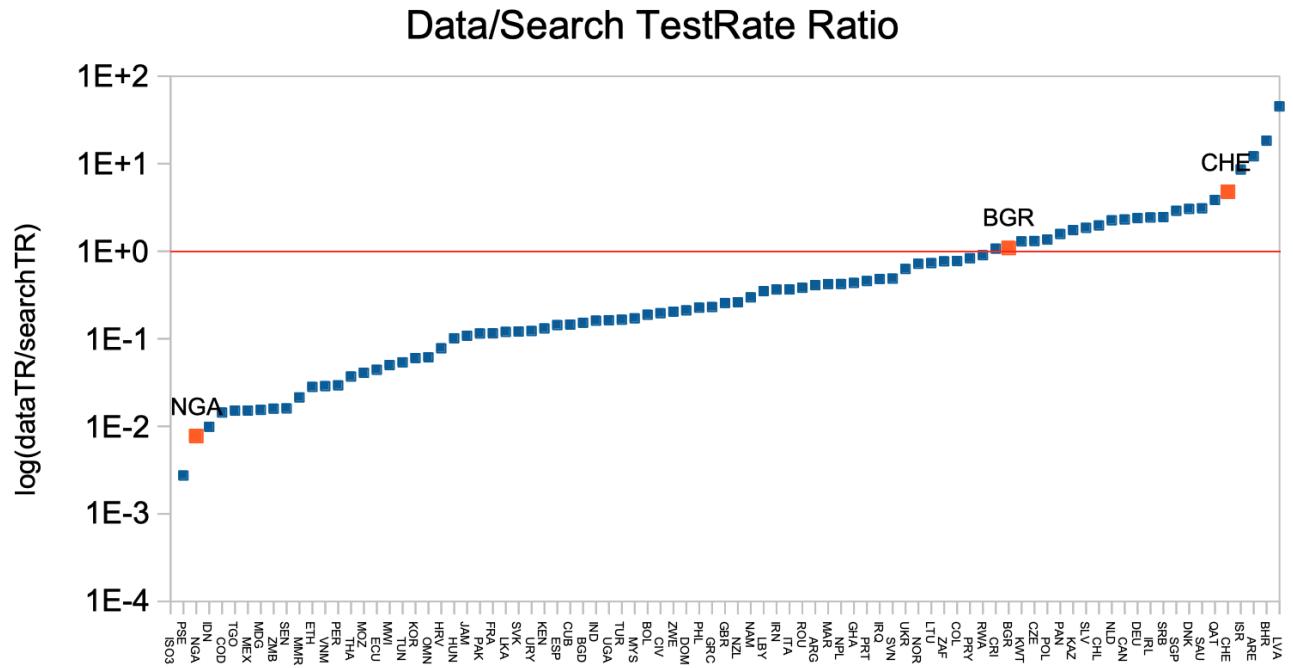


Figure 4: Search vs. data for testing rate - all countries

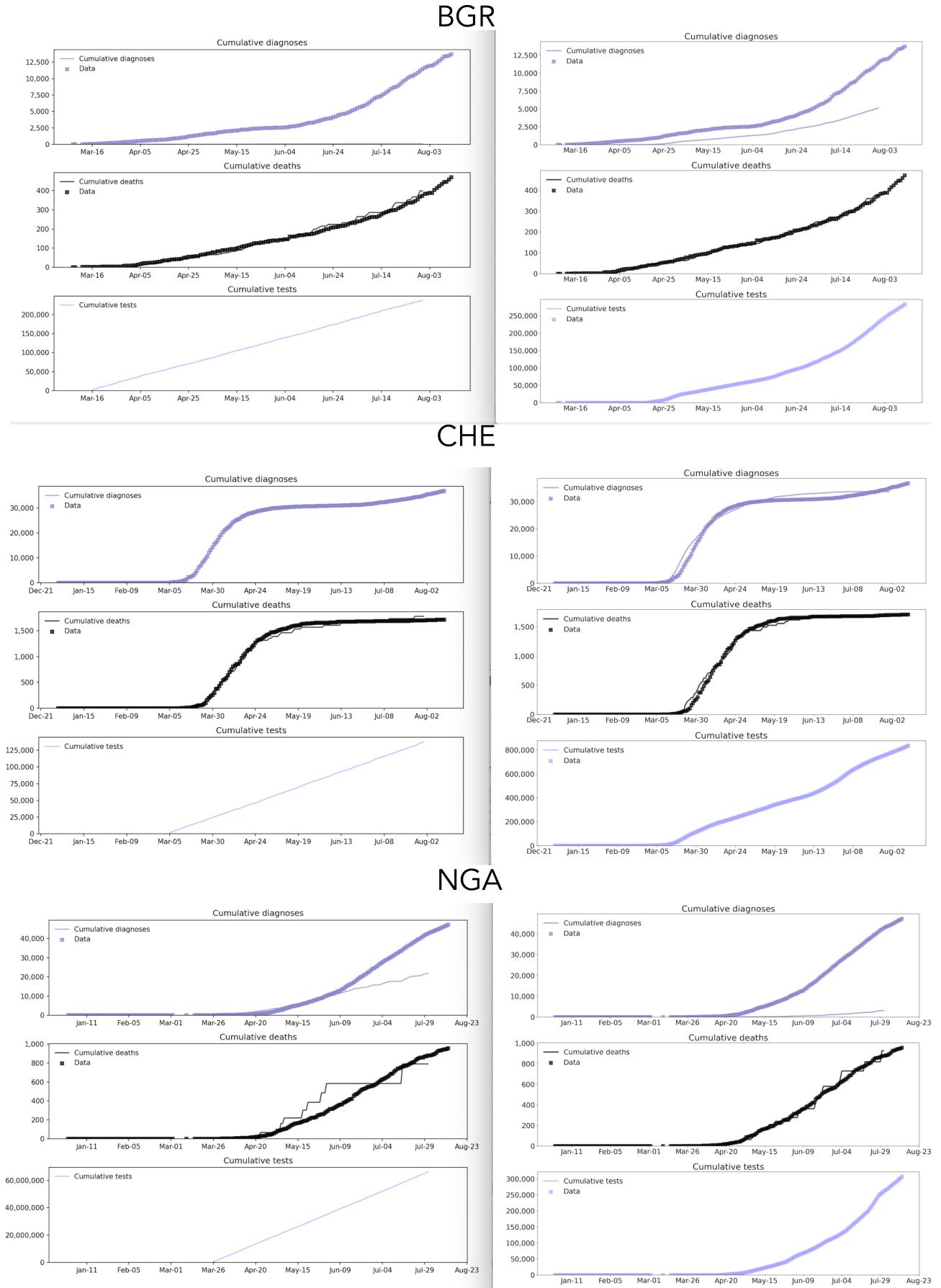


Figure 5: Search vs. data testing rates for BGR, CHE and NGA

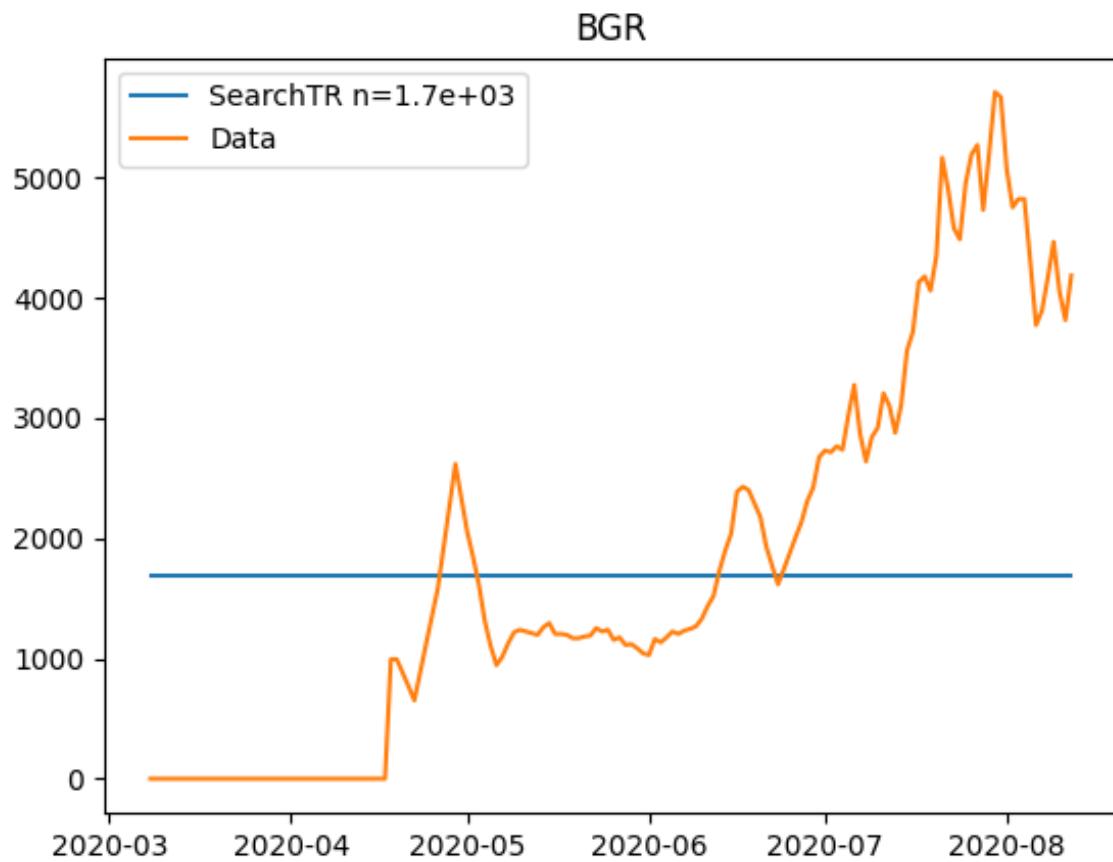


Figure 6: Testing rate from search vs. data for BGR

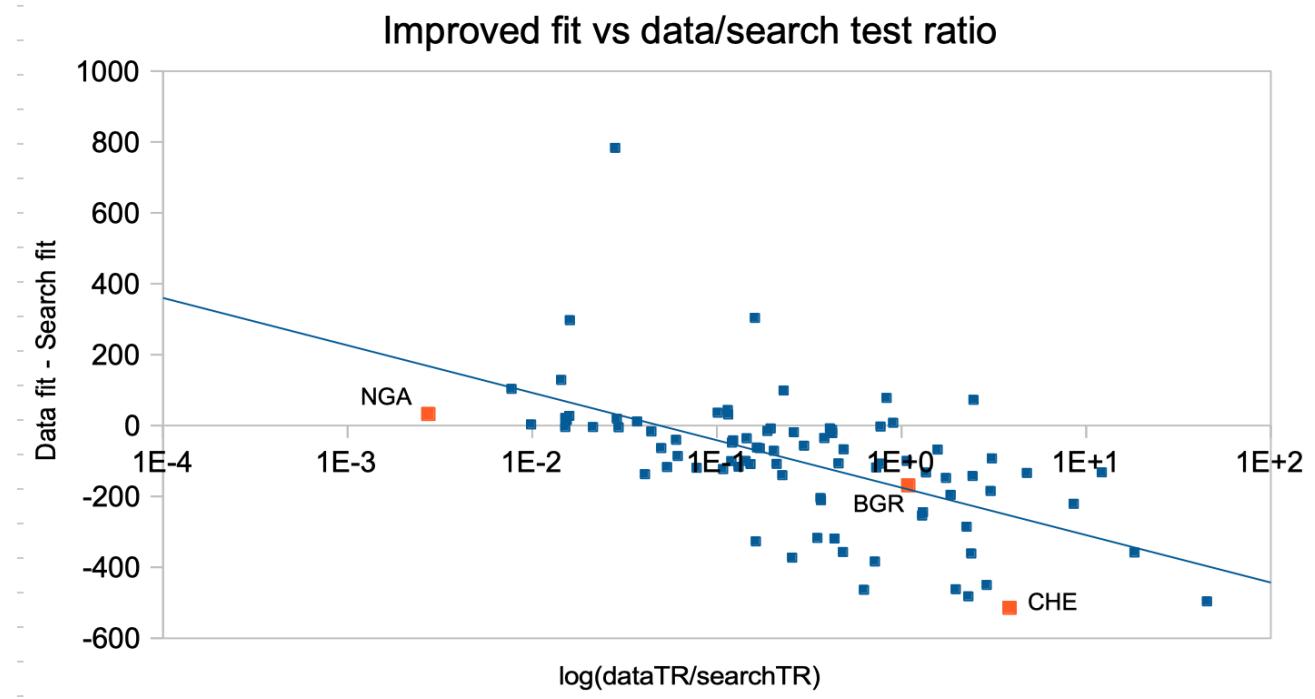
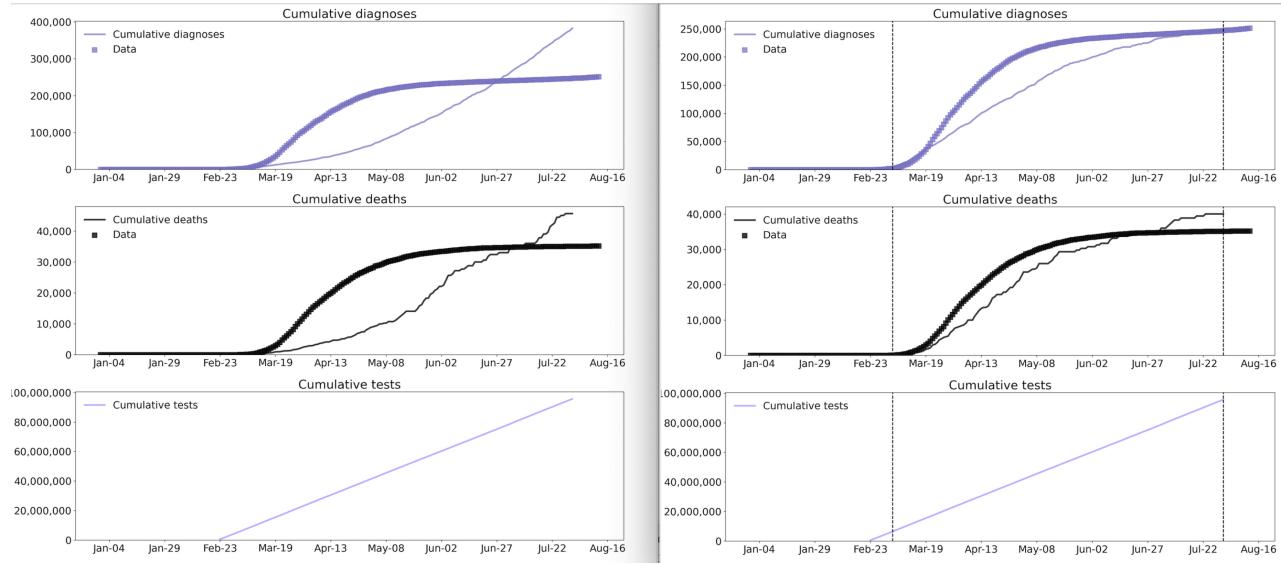


Figure 7: Model fit related to data/search testing rate

ITA



GBR

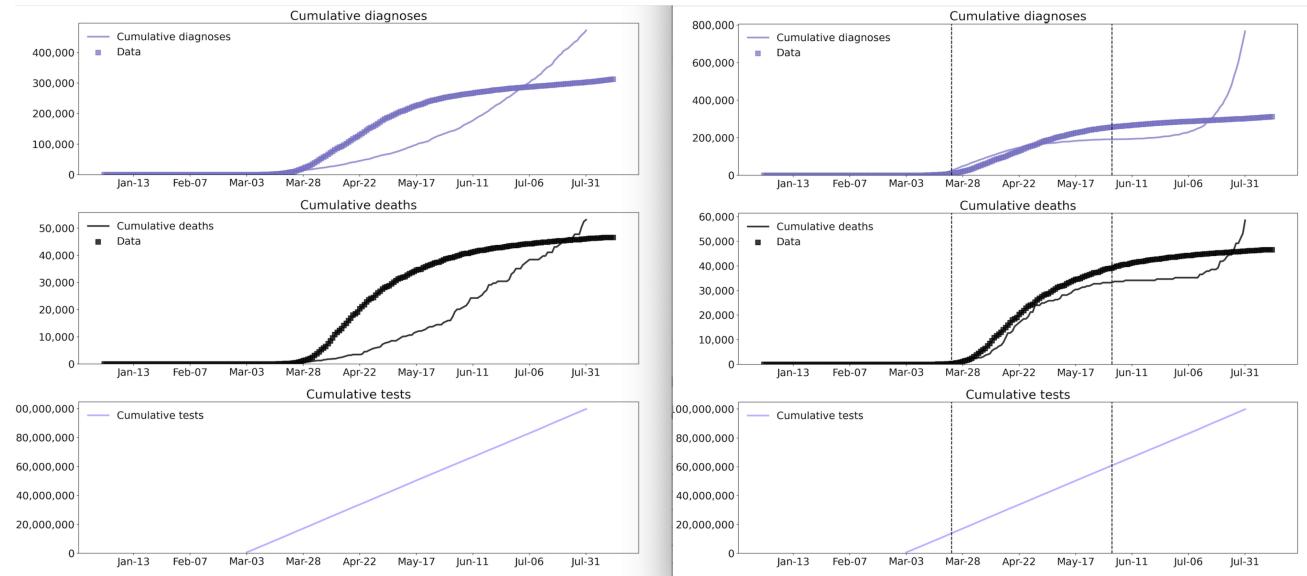


Figure 8: Interventions improving models

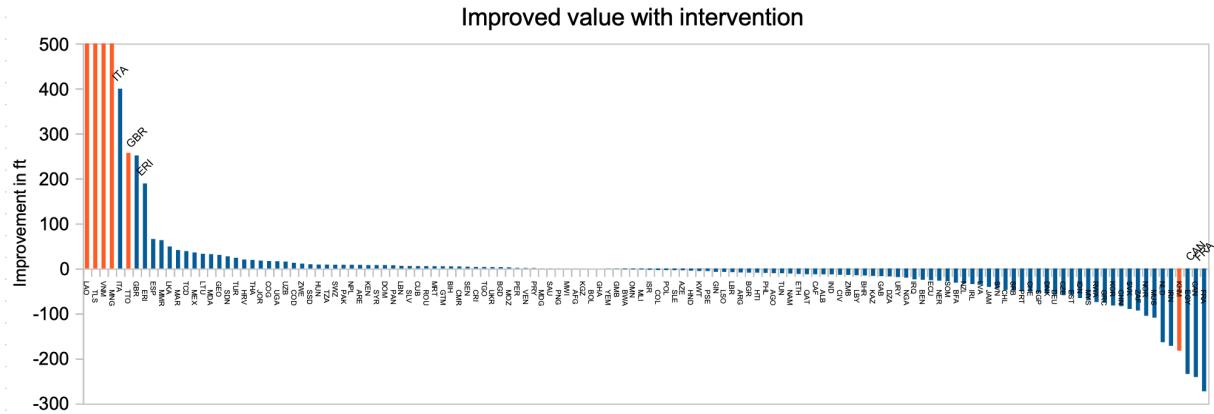
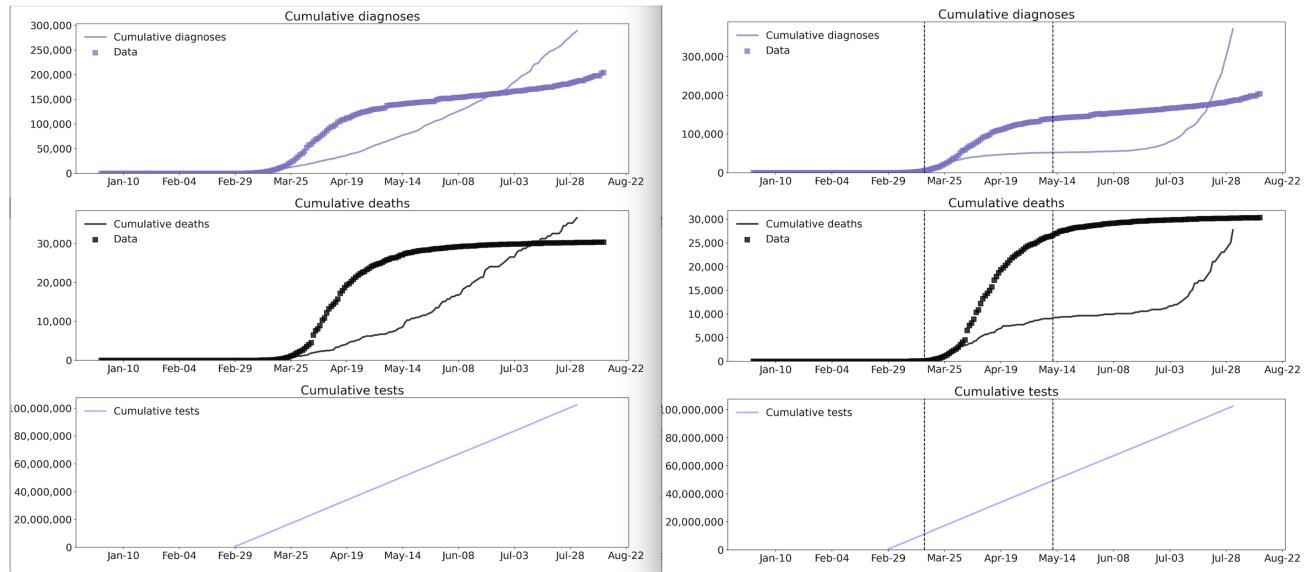


Figure 9: Improved fit

FRA



CAN

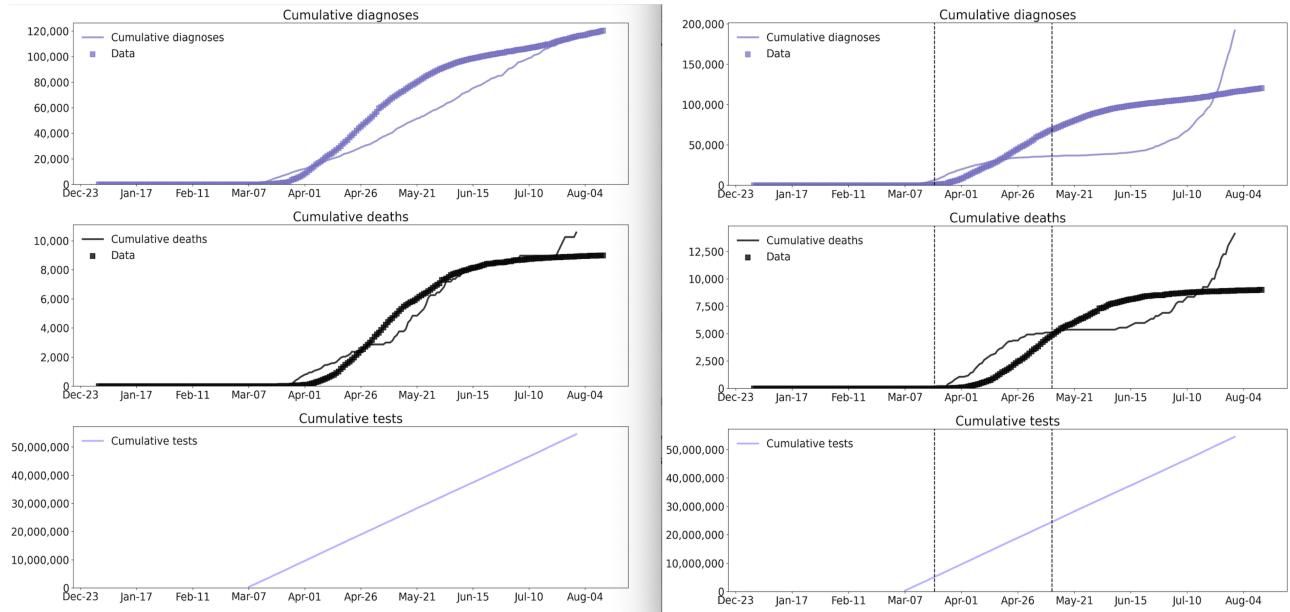


Figure 10: Interventions harming models

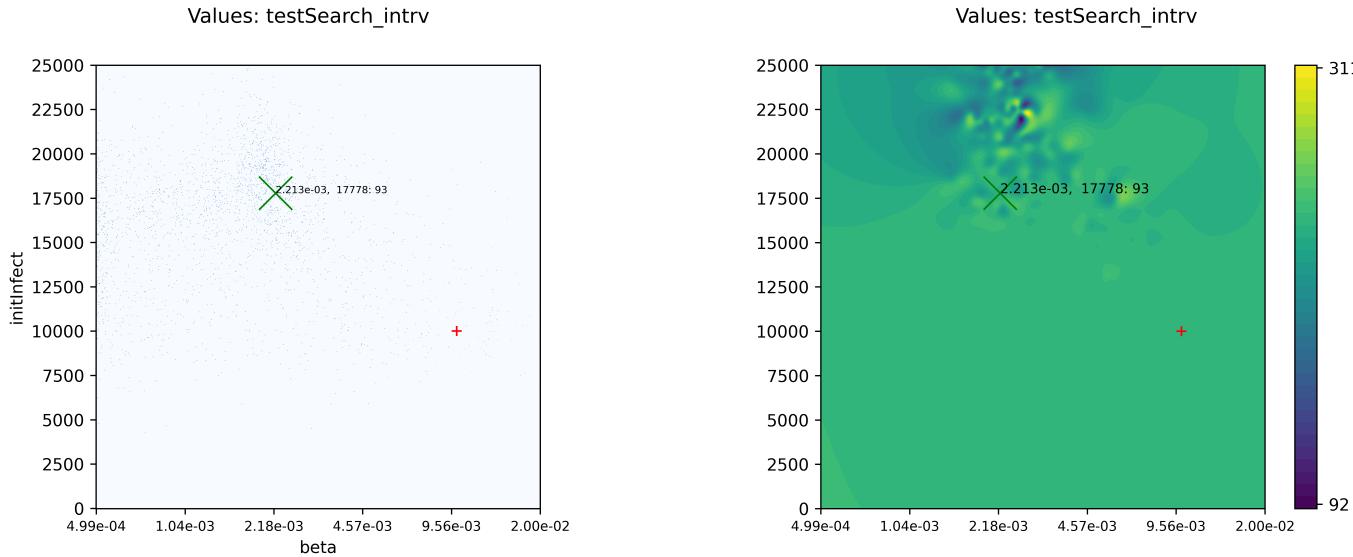


Figure 11: Value surface

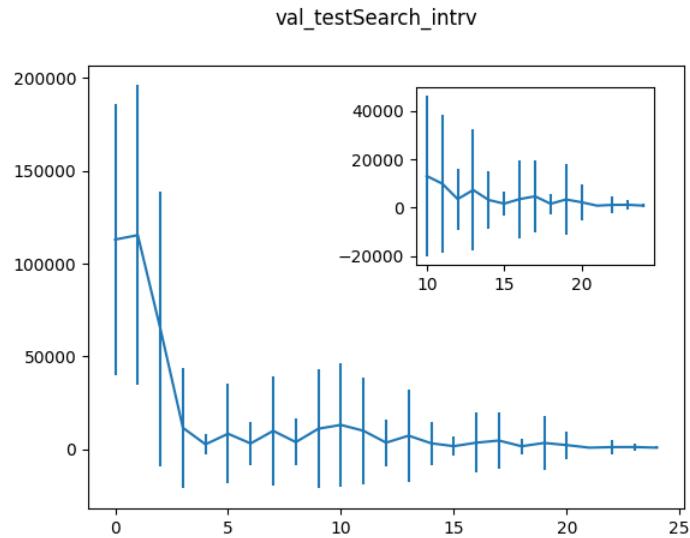


Figure 12: Value convergence

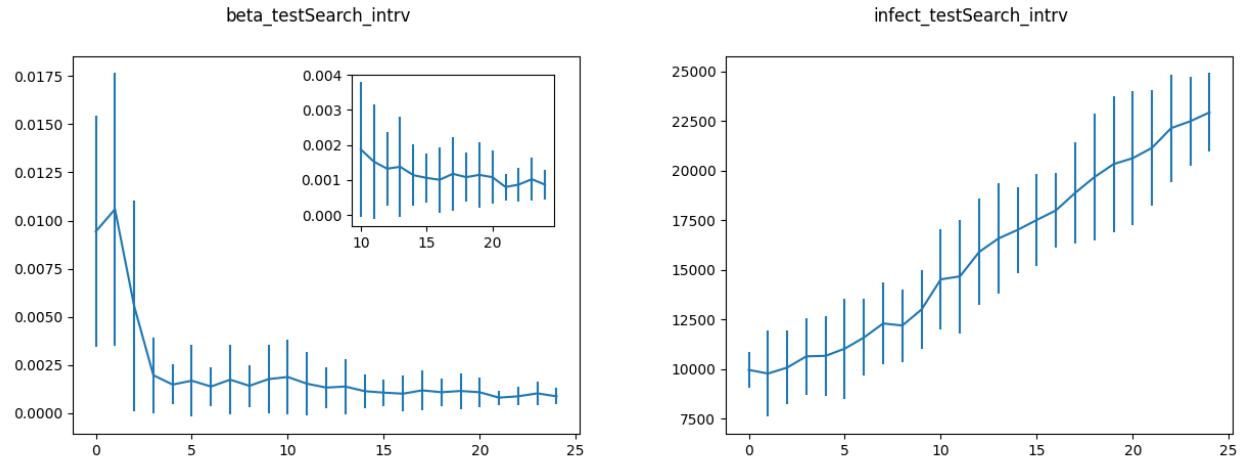


Figure 13: Parameter convergence

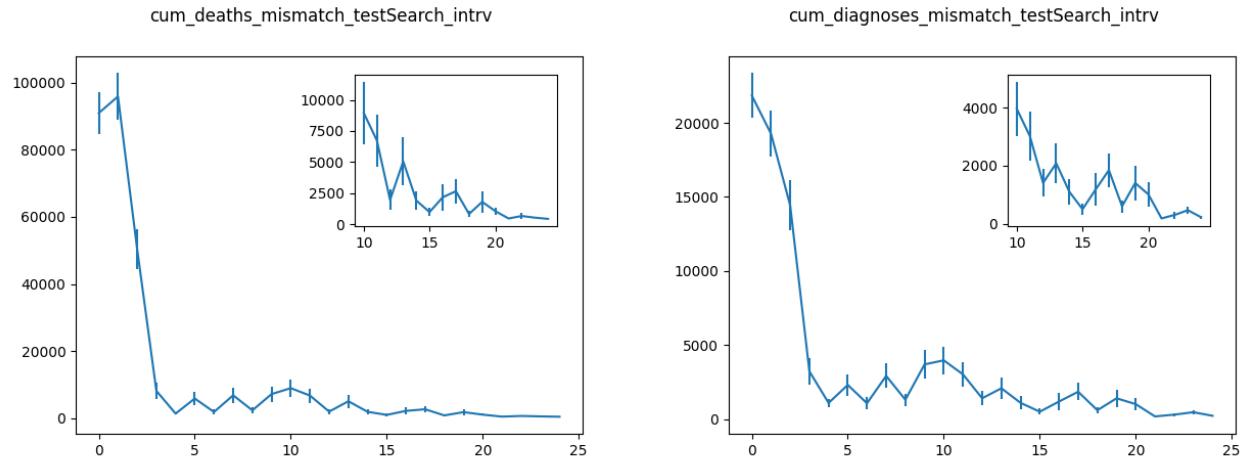


Figure 14: Independent deaths and diagnoses

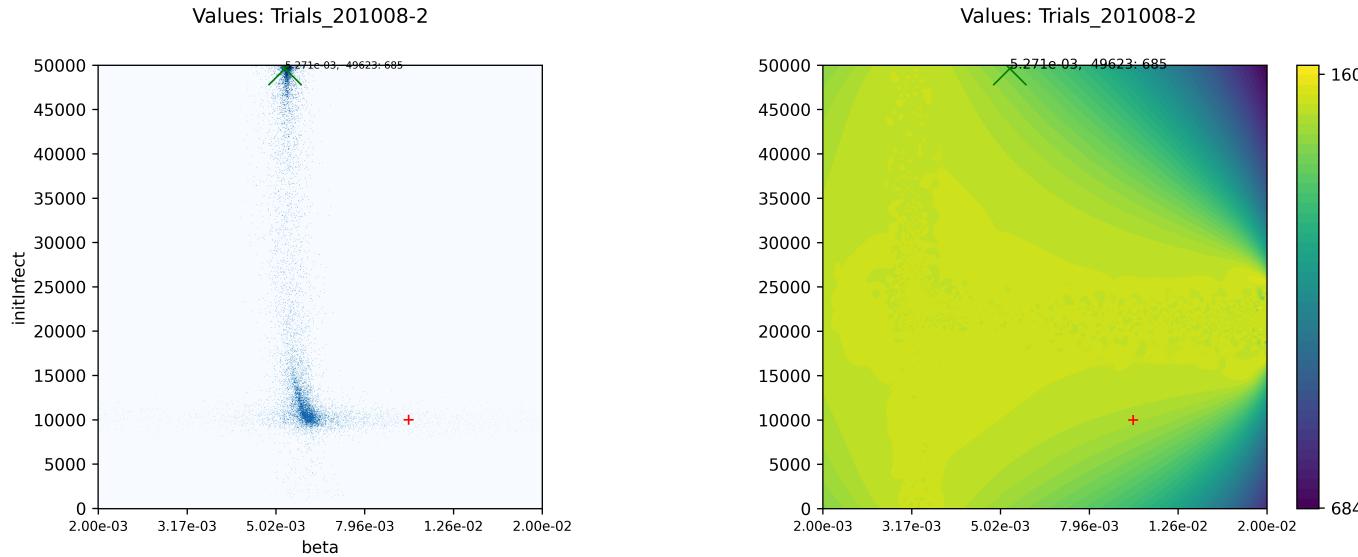


Figure 15: Values across larger bounded region

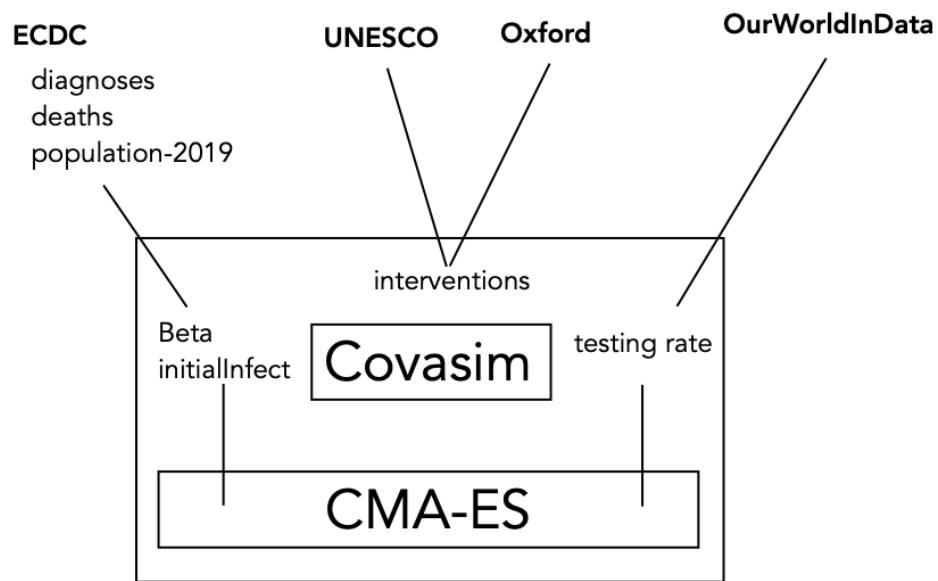


Figure 16: Data sources

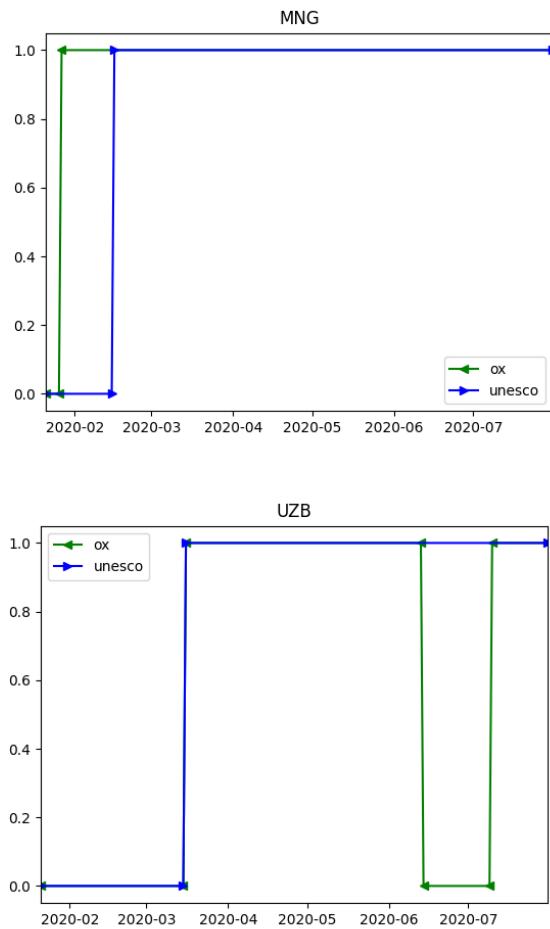


Figure 17: Overlapping intervention data

ISO3	Country	Intrvn	Test	ISO3	Country	Intrvn	Test	ISO3	Country	Intrvn	Test
AFG	Afghanistan	1	0	HND	Honduras	1	0	NZL	New_Zealand	1	1
AGO	Angola	1	0	HRV	Croatia	1	1	OMN	Oman	1	1
ALB	Albania	1	0	HTI	Haiti	1	0	PAK	Pakistan	1	1
ARE	United_Arab_Em	1	1	HUN	Hungary	1	1	PAN	Panama	1	1
ARG	Argentina	1	1	IDN	Indonesia	1	1	PER	Peru	1	1
AZE	Azerbaijan	1	0	IND	India	1	1	PHL	Philippines	1	1
BEN	Benin	1	0	IRL	Ireland	1	1	PNG	Papua_New_Gui	1	0
BFA	Burkina_Faso	1	0	IRN	Iran	1	1	POL	Poland	1	1
BGD	Bangladesh	1	1	IRQ	Iraq	1	1	PRT	Portugal	1	1
BGR	Bulgaria	1	1	ISR	Israel	1	1	PRY	Paraguay	1	1
BHR	Bahrain	1	1	ITA	Italy	1	1	PSE	Palestine	1	1
BIH	Bosnia_and_Her.	1	0	JAM	Jamaica	1	1	QAT	Qatar	1	1
BOL	Bolivia	1	1	JOR	Jordan	1	0	ROU	Romania	1	1
BWA	Botswana	1	0	KAZ	Kazakhstan	1	1	RWA	Rwanda	1	1
CAF	Central_African_I	1	0	KEN	Kenya	1	1	SAU	Saudi_Arabia	1	1
CAN	Canada	1	1	KGZ	Kyrgyzstan	1	0	SDN	Sudan	1	0
CHE	Switzerland	1	1	KHM	Cambodia	1	0	SEN	Senegal	1	1
CHL	Chile	1	1	KOR	South_Korea	1	1	SGP	Singapore	1	1
CHN	China	1	0	KWT	Kuwait	1	1	SLE	Sierra_Leone	1	0
CIV	Cote_d'Ivoire	1	1	LAO	Laos	1	0	SLV	El_Salvador	1	1
CMR	Cameroon	1	0	LBN	Lebanon	1	0	SOM	Somalia	1	0
COD	Democratic_Repub	1	1	LBR	Liberia	1	0	SRB	Serbia	1	1
COG	Congo	1	0	LBY	Libya	1	1	SSD	South_Sudan	1	0
COL	Colombia	1	1	LKA	Sri_Lanka	1	1	SVK	Slovakia	1	1
CRI	Costa_Rica	1	1	LSO	Lesotho	1	0	SVN	Slovenia	1	1
CUB	Cuba	1	1	LTU	Lithuania	1	1	SWZ	Eswatini	1	0
CZE	Czechia	1	1	LVA	Latvia	1	1	SYR	Syria	1	0
DEU	Germany	1	1	MAR	Morocco	1	1	TCD	Chad	1	0
DNK	Denmark	1	1	MDA	Moldova	1	0	TGO	Togo	1	1
DOM	Dominican_Repu	1	1	MDG	Madagascar	1	1	THA	Thailand	1	1
DZA	Algeria	1	0	MEX	Mexico	1	1	TLS	Timor_Leste	1	0
ECU	Ecuador	1	1	MLI	Mali	1	0	TTD	Trinidad_and_Tobago	1	0
EGY	Egypt	1	0	MMR	Myanmar	1	1	TUN	Tunisia	1	1
ERI	Eritrea	1	0	MNG	Mongolia	1	0	TUR	Turkey	1	1
ESP	Spain	1	1	MOZ	Mozambique	1	1	TZA	United_Republic_of	1	0
EST	Estonia	1	1	MRT	Mauritania	1	0	UGA	Uganda	1	1
ETH	Ethiopia	1	1	MUS	Mauritius	1	0	UKR	Ukraine	1	1
FRA	France	1	1	MWI	Malawi	1	1	URY	Uruguay	1	1
GAB	Gabon	1	0	MYS	Malaysia	1	1	UZB	Uzbekistan	1	0
GBR	United_Kingdom	1	1	NAM	Namibia	1	1	VEN	Venezuela	1	0
GEO	Georgia	1	0	NER	Niger	1	0	VNM	Vietnam	1	1
GHA	Ghana	1	1	NGA	Nigeria	1	1	YEM	Yemen	1	0
GIN	Guinea	1	0	NLD	Netherlands	1	1	ZAF	South_Africa	1	1
GMB	Gambia	1	0	NOR	Norway	1	1	ZMB	Zambia	1	1
GRC	Greece	1	1	NPL	Nepal	1	1	ZWE	Zimbabwe	1	1
GTM	Guatemala	1	0								

Figure 18: Countries' data