Data-driven Optimization Model for Global Covid-19 Intervention Plans

Anonymous authors
Paper under double-blind review

Abstract

In the wake of COVID-19, every government huddles to find the best interventions that will reduce the number of infection cases while minimizing the economic impact. However, with many intervention policies available, how should one decide which policy is the best course of action? In this work, we describe an integer programming approach to prescribe intervention plans that optimizes for both the minimal number of daily new cases and economic impact. We present a method to estimate the impact of intervention plans on the number of cases based on historical data. Finally, we demonstrate visualizations and summaries of our empirical analyses on the performance of our model with varying parameters compared to two sets of heuristics.

1 Introduction

As the pandemic continues to rage, developing data-driven approaches to predict future infection cases and prescribe actionable recommendations to prevent further infections becomes imperative. Thanks to the large-scale and recent data curation related to the pandemic, many insights can be taken from historical trends to understand the infection rate and prevention plans' effectiveness. Based on these insights, we can help decision-makers understand the impact of different intervention plans (IP) and prescribe the most beneficial policies. In order to tackle this problem, we set out to formulate an integer program, with parameters computed from historical data, to find the optimal intervention policies. The final section of this paper highlights the performance of our proposed model compared with existing heuristics. This work results from participating in the XPRIZE Pandemic Response Challenge (1). We will be using the same dataset and standard predictor as provided in the challenge.

Data We use the Oxford dataset (2) to get information on the number of cases and historical IPs for regions around the world. This dataset contains case numbers and historical IPs taken across 280 regions around the globe since the rise of the pandemic. We can see a sample of the historical IPs prescribed in the US and Canada in Appendix A.

Standard Predictor In this study, we make use of the predictor made available by (1) to help us estimate parameters and make future prescriptions. The authors use a 21-day lookback window to capture the number of cases and the IPs separately. The model enforces monotonicity on the IPs as increasing their stringency should lead to fewer infection cases. They also use a 7-day rolling window to average out any inconsistencies in daily case counts. The authors then train a model comprising two separate recurrent LSTM layers as two distinct pathways that eventually lead to a single prediction. The standard predictor uses the Oxford dataset (2) until Jan 11, 2021.

Problem Definition Let $\mathcal{P}$ represent the set of IPs, in this problem, $|\mathcal{P}| = 12$, and $\mathcal{L}_p$ represent the set of available restriction level for plan $p \in \mathcal{P}$. The list of intervention plans and their associated restriction level is shown in Appendix B. For each region, given the different IPs and the restriction levels for each IP, the problem is to find the optimal set of IPs that minimizes both the daily new cases and the stringency cost for each region.

2 Exact Approach to Prescribing Optimal Intervention Plans

We present an exact algorithm to solve the task of prescribing optimal intervention plans. To formulate the problem as an integer program, we must first estimate several parameters. The formulation uses the following variables:

$$\text{Let } x_{p,l} = \begin{cases} 1 & \text{if IP } p \text{ at level } l \text{ is prescribed} \\ 0 & \text{otherwise} \end{cases}$$

2.1 STRINGENCY COSTS

Let $S_{p,l}$ represent the stringency cost associated with IP p at level l. In this work, we experimented with three sets of stringency costs: fixed, random, and realistic. For all of these costs, we only generate a cost for each IP at level one. We multiply the cost at level one with the current level to reflect increasing costs with increasing restrictions for each increasing level. The fixed costs assume that the stringency cost for each IP is 1. We generate the random costs at random. The realistic costs are chosen based on the relative costs of implementing these IPs in Canada on a scale of 1 to 10 (10 being the highest cost), as shown in Appendix B. We then apply the same costs to all the regions and expect it to add value given its relatively horizontal and realistic application.

2.2 IMPACT OF INTERVENTION PLANS ON THE NUMBER OF INFECTION CASES

Let $C_{p,l}$ represent the impact IP p at level l has on the number of infection cases. We estimate $C_{p,l}$ as follows. For each region, first, we use the standard predictor to compute baseline infection cases (B) by setting all IPs to level zero, simulating a scenario where a region prescribes no IP. Next, for each IP and each restriction level, we use the standard predictor to compute the estimated number of infection cases ($E_{p,l}$) by only activating that specific IP at that restriction level. For each region, we then compute $C_{p,l}$ for each IP and its associated level as shown in Equation 1.

$$C_{p,l} = \frac{(E_{p,l} - B)}{B} * 100 \tag{1}$$

To test the impact of these weights, we perform this estimating process for one and seven days in the future, starting from three different days (Aug 2, 2020; Jan 2, 2021; Jan 15, 2021), creating in total six sets of impacts that we test with our optimization model.

2.3 OBJECTIVE FUNCTION NORMALIZATION

Since we are optimizing for two different objectives: the number of cases, and stringency costs, we need to normalize the two numbers so the optimization model does not lean towards one or the other. We first normalize the stringency costs at level one for the 12 IPs to sum to 1. The total stringency cost can be written as in Objective 2. Since the restriction level for any IP is at most 4, the total stringency cost is constrained to be at most 4.

$$\sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}_p} S_{p,l} * x_{p,l} \tag{2}$$

As for the number of cases, the impacts are at most 0.1 and on average at 0.02. We then normalize the number of infections to the current case numbers. For each region, let β be the number of current new cases. Let α be the number of new cases in the previous day, which we compute using the standard predictor. The number of new cases can be written as in Objection 3. As we prescribe for multiple days, β remains the same while α will vary. As α, the number of predicted future new cases increases, Objective 3 will increase, forcing the model to pay more weights to the number of cases. On the contrary, if α decreases, the model will start to shift more weights to the stringency costs.

$$\frac{1}{\beta} \left(\alpha + \sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}_p} C_{p,l} * x_{p,l} * \alpha \right) \tag{3}$$

2.4 INTEGER PROGRAM FORMULATION

To prescribe for several days and multiple regions, we take an iterative approach. For each prescription day d and for each region r, we solve the integer program shown in Figure 1. This formulation uses

binary variables $x_{p,l} = 1$ if IP p at level l is prescribed. The objective function 4 minimizes the total number of new cases and stringency costs. In Equation 5, we make sure that for each IP, we select only one restriction level at a time.

$$\text{minimize Objectives (2) (3)} \tag{4}$$

$$\text{s.t.} \sum_{l \in \mathcal{L}_p} x_{p,l} = 1 \qquad\qquad \forall p \in \mathcal{P} \tag{5}$$

$$x_{p,l} \in \{0, 1\} \qquad\qquad \forall p \in \mathcal{P}, l \in \mathcal{L}_p \tag{6}$$

Figure 1: Integer Program Model

2.5 Consecutive Constraint

In the real world, it is natural that when an IP is put in place, it will last for at least several days to minimize the impact on people's lives. By analyzing historical IPs across all the regions (as shown in Appendix A), we have created a minimum number of consecutive days that each intervention plan at a specific level must hold, which we set a ceiling of 7 days, denoted as $D_{p,l}$. At the time of prescribing for a day, based on the previous seven days of prescribed IPs, we create a new parameter $F_{p,l} = 1$ if IP p at level l has not been prescribed to enough consecutive days, thus forcing the model to set $x_{p,l} = 1$, and 0 otherwise. We formulate this new constraint as Equation 7 and add it to the inter program formulation.

$$D_{p,l} \le x_{p,l} \qquad\qquad \forall p \in \mathcal{P}, l \in \mathcal{L}_p \tag{7}$$

3 Experimental Results

In this section, we present our integer program's performance, with and without the consecutive constraint, and with varying weights, compared with two sets of heuristics provided by XPRIZE[1].

Blind-greedy Heuristic In the blind-greedy heuristic, the algorithm does not take into consideration of past IPs. It starts by setting all IPs to level zero, then iteratively, it increases a restriction level of the next least costly IP. In this way, each subsequent prescription sets higher levels of restrictions but at the same time also becomes more costly. The result set of prescriptions from the blind-greedy method should yield a Pareto front highlighting the trade-off between the number of cases of stringency costs.

Random Heuristic For each region and each prescription day, this heuristic assigns a valid level for each IP at random.

We make prescriptions for 28 days from Jan 12, 2021. We then measure the mean stringency and the mean number of new cases predicted by the standard predictor. The optimization solver used is the SCIP solver (3) imported from Google OR Tools (4). We set the solver to use its default parameters.

In Figure 2, we can see the performance of prescribing for 28 days of various models using the realistic stringency costs in the US and Canada. As expected, the random models perform the worst. The blind_greedy models generate a Pareto front highlighting the trade-off between the two factors at hand.

We can observe that the real IPs tend to be on the more expensive side for both the US and Canada. A few of our optimization models' weights find less costly solutions for the US while leading to fewer cases. In Canada, some of our solutions trade-off a few extra cases to reduce the stringency impact significantly. Most of the real IPs and our optimization solutions do better than our baseline heuristics. We report the detailed numbers in Appendix D.

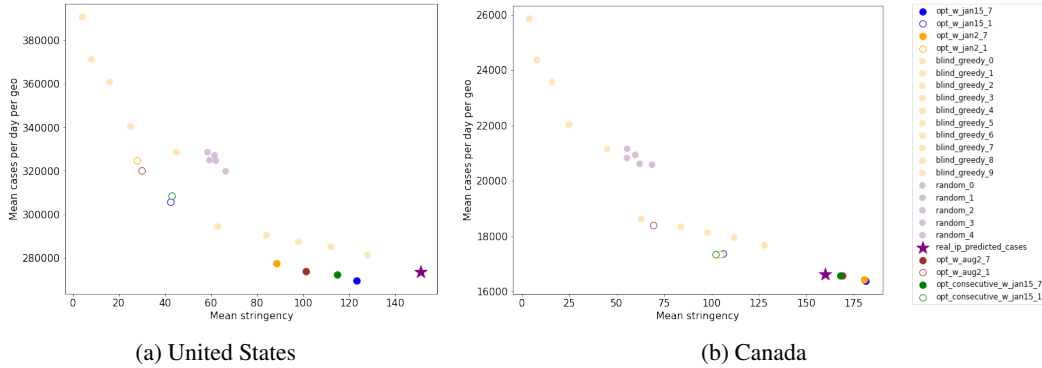[1]https://github.com/leaf-ai/covid-xprize/

(a) United States

(b) Canada

Figure 2: Performance of different models with realistic stringency costs in the US and Canada

3.1 Live Dashboards of the Prescriptor

Furthermore, we have designed live interactive dashboards that allow users to study all the different models' efficacy across regions and different sets of stringency costs. The first dashboard[2] compares the trade-off between the number of cases and stringency costs as seen in Figure 4 in Appendix C. The second dashboard [3] simulates the case numbers based on the different IPs over a 28-day period as seen in Figure 5 in Appendix C.

We compare our prescription model with the different parameters to the blind_greedy and random baselines. We also show real prescriptions as a baseline - to simulate the impact of listening to our models and comparing what transpired in real life. Users can choose different regions, weight types, models and study these factors' effect on the case numbers over 28 days starting Jan 12, 2021.

These dashboards have an immediate short-term application to the stakeholders dealing with public health decisions as economies start to re-open in 2021. For example, based on the first dashboard with Jan 15 weights prescriptions, the Canadian government could reduce the average economic impact by approximately 22% for a slight increase in the number of cases over a 28-day horizon.

4 Conclusions and Future Work

We have described a data-driven integer program approach to optimal intervention plan prescription. We presented the performance of our models in comparison with other heuristics and introduce a framework to visualize the impact of intervention plans on the number of cases and economic costs on a region-by-region basis.

Relevant stakeholders in public health, epidemiological research, and governments can readily implement the recommendations to manage the spread of the pandemic while minimizing the impact on society. We look forward to presenting our framework to our local government to get feedback and to get to a broader audience through this platform to save lives while mitigating societal impact.

This work is still preliminary and there are multiple ways we look forward to extending our work. First of all, we believe we can improve more on the normalization of the two costs in the integer program objective function by implementing algorithms from the multi-objective optimization field. Further, we can also improve on the estimated $C_{p,l}$ by performing a linear regression on the historical IPs. We have also started exploring Reinforcement Learning methods to tackle this problem, where we first model the spread of the disease as an environment for an agent to explore, we then integrate Deep Q-Learning (5) and Soft Actor Critic (SAC) algorithms (6) to optimize IPs for each region. We believe this might be a promising approach and will investigate further. Finally, we realize that the performance of the prescription model is heavily dependent on the performance of the infection cases predictor. We believe in our future work, we can also improve on the standard predictor or incorporate the two models in an interactive system (7).

---

[2]https://tabsoft.co/3ejM0iq
[3]https://tabsoft.co/3d71wOd

REFERENCES

[1] Risto Miikkulainen, Olivier Francon, Elliot Meyerson, Xin Qiu, Elisa Canzani, and Babak Hodjat. From prediction to prescription: Evolutionary optimization of non-pharmaceutical interventions in the covid-19 pandemic. *arxiv:2005.13766*, May 2020.

[2] Thomas Hale, Anna Petherick, Toby Phillips, and Samuel Webster. Variation in government responses to covid-19. *Blavatnik school of government working paper*, 31:2020–11, 2020.

[3] Gerald Gamrath, Daniel Anderson, Ksenia Bestuzheva, Wei-Kun Chen, Leon Eifler, Maxime Gasse, Patrick Gemander, Ambros Gleixner, Leona Gottwald, Katrin Halbig, Gregor Hendel, Christopher Hojny, Thorsten Koch, Pierre Le Bodic, Stephen J. Maher, Frederic Matter, Matthias Miltenberger, Erik Mühmer, Benjamin Müller, Marc E. Pfetsch, Franziska Schlösser, Felipe Serrano, Yuji Shinano, Christine Tawfik, Stefan Vigerske, Fabian Wegscheider, Dieter Weninger, and Jakob Witzig. The SCIP Optimization Suite 7.0. Technical report, Optimization Online, March 2020.

[4] Laurent Perron and Vincent Furnon. OR-Tools. `https://developers.google.com/optimization/`.

[5] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533, 2015.

[6] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International Conference on Machine Learning*, pages 1861–1870. PMLR, 2018.

[7] Adam N Elmachtoub and Paul Grigas. Smart" predict, then optimize". *arXiv preprint arXiv:1710.08005*, 2017.

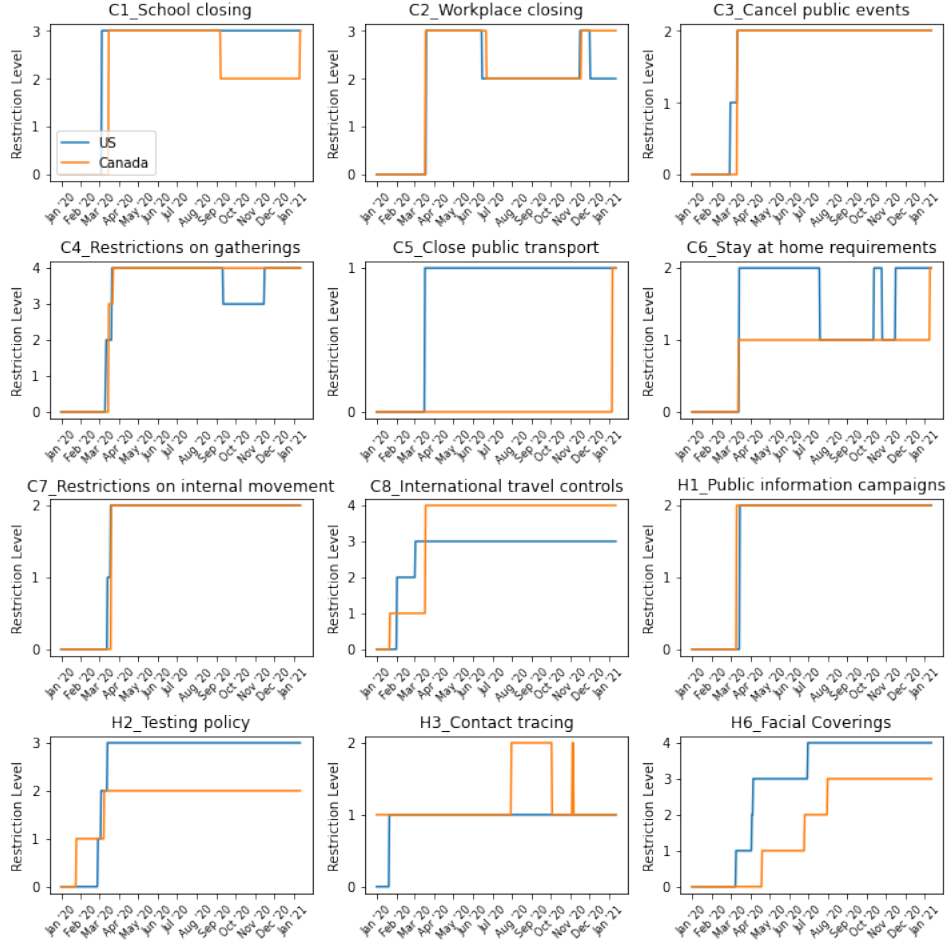# A   HISTORICAL INTERVENTION PLANS IN THE US AND CANADA

Figure 3: Trends of IPs in the US and Canada

B   INTERVENTION PLANS

Intervention Plan	Restriction Levels	Realistic Stringency Cost
C1_School closing	(0,1,2,3)	9
C2_Workplace closing	(0,1,2,3)	6
C3_Cancel public events	(0,1,2)	2
C4_Restrictions on gatherings	(0,1,2,3,4)	5
C5_Close public transport	(0,1,2)	8
C6_Stay at home requirements	(0,1,2,3)	7
C7_Restrictions on internal movement	(0,1,2)	7
C8_International travel controls	(0,1,2,3,4)	8
H1_Public information campaigns	(0,1,2)	2
H2_Testing policy	(0,1,2,3)	3
H3_Contact tracing	(0,1,2)	7
H6_Facial Coverings	(0,1,2,3,4)	2

Table 1: Restriction levels and realistic stringency cost of IPs

C DASHBOARDS

The first dashboard[4] compares the trade off between number of cases and stringency costs, a snapshot is shown in Figure 4. The second dashboard [5] simulates the case numbers based on the different IPs over a 28-day period, a snapshot is shown in Figure 5.
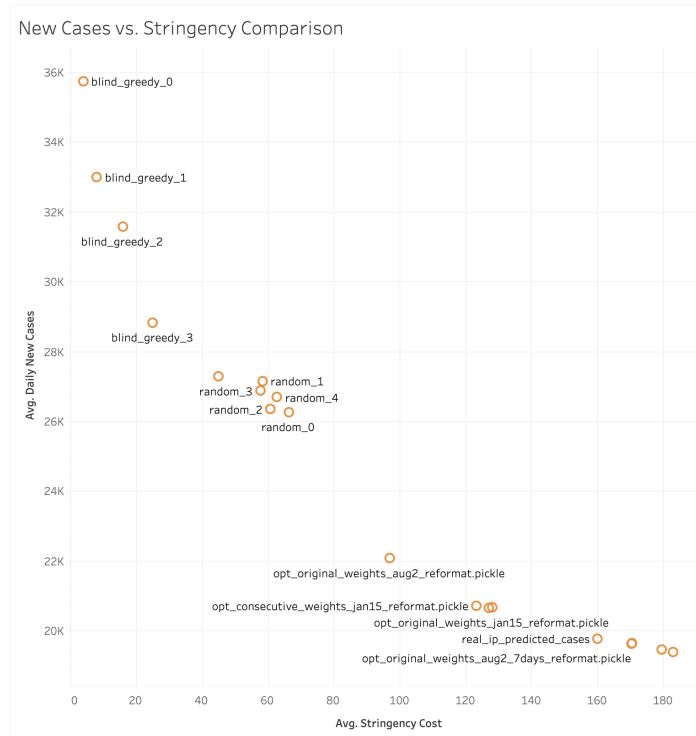

Figure 4: Comparison of daily average cases and Stringency of different models
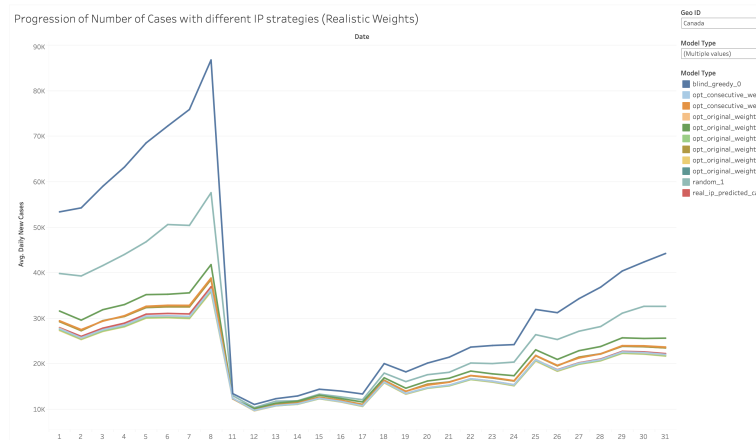

Figure 5: Impact of different IPs over the a seven day horizon

D   PERFORMANCE RESULTS

We present the detailed number of our experimental results here. There are in total 10 blind-greedy models with 5 random models. We experimented with 6 sets of different $C_{p,l}$, predicting 1 day and 7 days starting from 3 different dates: Aug 2, 2020, Jan 2, 2021; and Jan 15, 2021. We further experimented the consecutive constraint with two sets of $C_{p,l}$.

D.1   UNITED STATES

Stringency Costs	fixed		random		realistic	
Measure	**cases**	**costs**	**cases**	**costs**	**cases**	**costs**
opt_consecutive_w_jan15_7	273939.20	21.33	269403.72	19.95	272183.34	114.81
opt_consecutive_w_jan15_1	307375.74	7.71	307685.89	8.58	308403.71	42.95
opt_w_jan15_7	268771.90	24.24	268738.99	19.97	269436.13	123.10
opt_w_jan15_1	300796.03	8.67	300927.83	8.09	305727.94	42.43
opt_w_jan2_7	276585.64	17.43	272252.66	17.09	277367.53	88.38
opt_w_jan2_1	326332.17	5.10	322437.11	4.03	324601.12	28.00
opt_w_aug2_7	273110.33	20.48	270715.27	18.18	273755.98	101.14
opt_w_aug2_1	318821.35	5.67	317196.32	4.70	320039.63	29.90
blind_greedy_0	350768.63	3.00	276585.64	17.43	390854.80	4.00
blind_greedy_1	306824.18	6.00	272252.66	17.09	371281.95	8.00
blind_greedy_2	302727.20	8.00	276585.64	17.43	360932.75	16.00
blind_greedy_3	292355.06	12.00	272252.66	17.09	340350.61	25.00
blind_greedy_4	285647.76	14.00	293142.51	9.98	328540.60	45.00
blind_greedy_5	281362.98	17.00	289671.00	11.56	294287.71	63.00
blind_greedy_6	278267.20	19.00	287302.63	14.04	290403.70	84.00
blind_greedy_7	272759.52	23.00	284184.68	17.82	287388.78	98.00
blind_greedy_8	270096.78	25.00	280900.00	23.18	285031.79	112.00
blind_greedy_9	265914.71	28.00	278425.29	25.90	281126.78	128.00
random_0	322917.85	10.95	324300.98	9.79	327159.45	61.52
random_1	328073.25	10.62	322851.06	11.23	328588.25	58.48
random_2	324572.39	10.57	332052.66	10.76	324622.28	62.05
random_3	330125.04	10.00	326852.32	10.02	319797.71	66.38
random_4	326739.40	10.38	327730.99	10.34	324854.00	59.29
real_ip_predicted_cases	273586.82	29.00	273586.82	28.92	273586.82	151.00
actual_cases	*174680.05*	*29.00*	*174680.05*	*28.92*	*174680.05*	*151.00*

Table 2: Performance of various model for prescribing 28 days from Jan 12, 2021 in the United States

D.2 CANADA

Stringency Costs	fixed		random		realistic	
Measure	cases	costs	cases	costs	cases	costs
opt_consecutive_w_jan15_7	16525.46	31.19	16789.03	27.48	16588.19	168.00
opt_consecutive_w_jan15_1	17273.91	19.57	17450.57	14.24	17346.64	102.38
opt_w_jan15_7	16410.78	32.57	16480.23	31.43	16383.82	181.43
opt_w_jan15_1	17239.80	20.43	17451.10	14.11	17356.17	106.00
opt_w_jan2_7	16476.53	32.24	16556.99	31.07	16438.41	180.76
opt_w_jan2_1	17307.67	19.67	17476.61	13.88	17340.07	104.57
opt_w_aug2_7	16571.76	30.81	16780.44	27.34	16572.66	169.00
opt_w_aug2_1	18190.33	14.05	17892.12	8.34	18388.79	69.24
blind_greedy_0	22811.46	3.00	22151.84	0.09	25858.13	4.00
blind_greedy_1	19543.13	6.00	20593.91	1.47	24362.63	8.00
blind_greedy_2	19242.32	8.00	20275.72	2.83	23577.58	16.00
blind_greedy_3	18484.12	12.00	19852.64	4.35	22028.85	25.00
blind_greedy_4	17996.32	14.00	19300.23	6.05	21147.75	45.00
blind_greedy_5	17685.69	17.00	17710.85	8.81	18624.70	63.00
blind_greedy_6	17461.80	19.00	17447.73	13.65	18341.68	84.00
blind_greedy_7	17064.48	23.00	17228.14	16.15	18122.44	98.00
blind_greedy_8	16872.91	25.00	16905.64	21.79	17951.33	112.00
blind_greedy_9	16572.61	28.00	16745.72	26.02	17668.37	128.00
random_0	21057.65	9.95	20743.72	11.14	20582.15	68.67
random_1	20808.50	11.62	20705.43	11.53	21155.11	55.57
random_2	20777.58	10.95	20742.32	10.97	20614.5	62.14
random_3	20963.39	10.90	20952.72	11.15	20822.47	55.48
random_4	20673.91	11.76	20774.19	11.00	20937.03	59.86
real_ip_predicted_cases	16615.24	29.00	16615.24	29.52	16615.24	160.00
actual_cases	*5488.857*	*29.00*	*5488.86*	*29.52*	*5488.857*	*160.00*

Table 3: Performance of various model for prescribing 28 days from Jan 12, 2021 in Canada

D.3 WORLD

Stringency Costs	fixed		random		realistic	
Measure	**cases**	**costs**	**cases**	**costs**	**cases**	**costs**
opt_consecutive_w_jan15_7	6769.99	26.03	6769.69	23.67	6777.57	142.09
opt_consecutive_w_jan15_1	7405.65	19.57	7338.26	17.31	7464.95	107.01
opt_w_jan15_7	6665.29	28.15	6684.85	25.74	6666.85	154.33
opt_w_jan15_1	7263.72	20.56	7254.68	18.00	7329.61	110.60
opt_w_jan2_7	6810.59	25.85	6799.07	23.50	6820.37	141.28
opt_w_jan2_1	7781.06	17.31	7690.06	15.08	7786.65	93.88
opt_w_aug2_7	6794.48	24.79	6794.49	22.21	6803.79	134.86
opt_w_aug2_1	7829.05	14.92	7734.96	13.01	7863.44	80.52
blind_greedy_0	9863.78	3.00	10795.07	0.44	11533.58	4.00
blind_greedy_1	8135.79	6.00	10029.18	1.34	10706.43	8.00
blind_greedy_2	7980.38	8.00	9444.77	2.63	10277.08	16.00
blind_greedy_3	7591.98	12.00	8775.29	4.27	9441.59	25.00
blind_greedy_4	7344.57	14.00	8349.93	6.39	8973.26	45.00
blind_greedy_5	7188.00	17.00	8017.03	8.89	7663.01	63.00
blind_greedy_6	7075.65	19.00	7572.93	11.80	7518.95	84.00
blind_greedy_7	6877.36	23.00	7326.67	15.23	7407.82	98.00
blind_greedy_8	6782.21	25.00	7085.89	19.14	7321.33	112.00
blind_greedy_9	6633.68	28.00	6901.52	23.47	7178.85	128.00
random_0	8841.54	10.99	8797.98	10.95	8824.25	60.56
random_1	8808.71	11.02	8764.34	10.99	8849.60	60.10
random_2	8779.83	11.02	8829.24	11.07	8810.69	60.65
random_3	8835.40	10.95	8786.90	10.99	8789.41	60.44
random_4	8800.00	11.03	8792.29	11.03	8792.33	60.65
real_ip_predicted_cases	5536.39	21.64	5536.39	21.55	5536.39	111.10
actual_cases	*3144.09*	*21.64*	*3144.09*	*21.55*	*3144.09*	*111.10*

Table 4: Performance of various model for prescribing 28 days from Jan 12, 2021 in the world