Modelling global public health strategies in COVID-19 pandemic using deep reinforcement learning

Gloria Hyunjung, Kwak (✉ hkwak@cse.ust.hk)

The Hong Kong University of Science and Technology

Lowell Ling

The Chinese University of Hong Kong

Pan Hui

The Hong Kong University of Science and Technology

Research Article

Keywords: Health Policy, Public Health Intervention, Machine learning, COVID-19, Coronavirus, population health

Posted Date: May 27th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-31226/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Rationale: Unprecedented public health measures have been used during this coronavirus 2019 (COVID-19) pandemic but with a cost to economic and social disruption. It is a challenge to implement timely and appropriate public health interventions.

Objectives: This study evaluates the timing and intensity of public health policies in each country and territory in the COVID-19 pandemic, and whether machine learning can help them to find better global health strategies.

Methods: Population and COVID-19 epidemiological data between 21st January 2020 to 7th April 2020 from 183 countries and 78 territories were included with the implemented public health interventions. We used deep reinforcement learning, and the model was trained to try to find the optimal public health strategies with maximizing total reward on controlling spread of COVID-19. The results proposed by the model were analyzed against the actual timing and intensity of lockdown and travel restrictions.

Measurements and Main Results: Early implementation of the actual lockdown and travel restriction policies were associated with gradually groups of less severe crisis severity, relative to local index case date in each country or territory, not to 31st December 2019. However, our model suggested to initiate at least minimal intensity of lockdown or travel restriction even before index cases in each country and territory. In addition, the model mostly recommended a combination of lockdown and travel restrictions and higher intensity policies than the implemented policies by government, but did not always encourage rapid full lockdown and full border closures.

Conclusion: Compared to actual government implementation, our model mostly recommended earlier and higher intensity of lockdown and travel restrictions. Machine learning may be used as a decision support tool for implementation of public health interventions during COVID-19 and future pandemics.

1. Introduction

Coronavirus disease 2019 (COVID-19) was first reported by health authorities in Wuhan, China on 31st December 2019 ¹. In mainland China, the number of confirmed infections with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), increased to around 75,000 within a month from the first confirmation date of 20th January 2020 ². Korea and Italy were the next outbreak countries and currently identified cases have been reported in 183 countries and 78 territories. The massive number of patients infected within a short time period have overwhelmed many countries and territories. The lack of reliable and rapid testing, self-quarantine facilities, personal protective equipment, hospital and critical care capacity and effective treatment have created a health crisis for countries and territories that were not ready. COVID-19 have caused more than 203,000 deaths globally, and this figure is likely a conservative estimate due to under diagnosis. Furthermore, the consequence of this COVID-19 pandemic and the measures used to control it have resulted in a global financial crisis.

Determining the appropriate type and level of public health policy for each country and territory is very challenging. Different countries and territories have their unique population structure and density, economic resources, healthcare systems, governance, and culture. In addition, the index case of COVID-19 and initial spread of the virus for each country and territory is often unknown. Thus, governments were forced to apply policies with incomplete information about the burden of disease as well as the uncertainty about the biological and clinical characteristics of the virus ³. Decision making is further complicated by the response lag in new infections, hospitalizations, and mortality. In the meantime, studies have investigated the effects of these public health decisions ^{4–7}.

For example, the effect of travel restrictions on domestic and international spread of SARS-CoV-2 was studied with data from 200 countries and territories using the global epidemic and mobility model (GLEAM). The model showed that 77% reduction in cases imported to other countries due to travel restriction out of Wuhan although it only had modest effect on domestic spread in China. Yet when the efficacy of travel restrictions was assessed at different transmission scenarios, travel ban was only meaningful if combined with a 50% or higher transmission reduction ⁵. Early or preemptive lockdown has also been shown to be more effective than delayed response in China ⁶. Using simulated data, it was shown that lockdown policy reduces the number of deaths even when only 5% of population is infected ⁷. Taken together, it suggests that fast intervention, and simultaneously placing nationwide and worldwide travel bans are effective. However, there is still a lack of effective tools to provide specific decision support for individual countries and territories with different health care systems and burden of COVID-19.

In this work, we propose a data-driven approach to discover optimal lockdown and travel restriction policies for individual countries and territories with the state-of-the-art deep reinforcement learning (RL) algorithm. RL is one of three basic machine learning fields along with supervised and unsupervised learning, and it has essentially gotten the concept in the human learning process of what to do in a particular situation: how to map the situation to action.

Contrary to the concept of supervised learning to learn the correct action (label) with a description of situation (example), RL seeks an action that maximizes accumulated reward received through trial-and-error without being told what to do directly ^{8–10}. We conducted policy effectiveness studies with deep RL to learn sequential decision making so as to maximize rewards over time by accelerations and decelerations in the number of confirmed COVID-19 infections, deaths and recovered cases. The timing and intensity of lockdown and travel restriction policies were suggested by the model and compared to actual public health interventions implemented during this COVID-19 pandemic.

2. Methods

2.1 Data and Preprocessing

We included data between 21st January 2020, the first date that World Health Organization (WHO) reported situation on COVID-19, to 7th April 2020 from 183 countries and 78 territories. For each country

and territory, index case date (date of the first confirmed patient), the number of tested, confirmed infection, recovery and death were collected from Johns Hopkins coronavirus data repository, Centers for Disease Control and Prevention's reports and WHO's case reports. We also collected data on timing and intensity of domestic lockdown and international travel restrictions, population size, population density, population mid-year (aged 15 to 65 years old), gross domestic product (GDP), geological information (longitude, latitude) and life expectancy from the United Nations database, Wikipedia, and official announcements through the news 11–16. This included early actions from countries and territories implemented before the first local case of infection was confirmed. After linear interpolation from the index case date in each country and territory, data was compiled with an average value over a three-day period, to reduce bias from delayed reporting and variable viral testing capacity 17. Countries and territories were excluded from analysis if they had fewer than 100 cases of COVID-19 by 7th April 2020. The dataset was divided into a 7:2:1 ratio of training, validation and test sets.

2.1.1. Severity Level

The crude death rate due to COVID-19 reported on 7th April 2020, calculated as the number of deaths related to COVID-19 in the total population (per 1,000) was used as an indicator of the country or territory's overall crisis severity level. The severity group was divided into four levels (low/medium/high/critical level of severity). Countries or territories which did not have any deaths were designated as low severity group. The remaining were evenly divided into 3 groups according the COVID-19 crude death rate. Each severity groups' characteristics and burden of COVID-19 is shown in Table 1.

2.2. Model

The goal of reinforcement learning is to train a decision-making agent to seek to achieve its target (maximizing cumulative rewards) despite uncertainty about its environment 8. At each time stamp t , an agent has a combination of action a_t and state s_t along with compensation r_t for each case. With interacting with its environment, at each time stamp t , an agent receives state s_t and reward r_t from environment, and then chooses an action a_t . Subsequently the action a_t is sent to the environment. The environment moves to the next state s_{t+1} and finally the agent receives an evaluative feedback r_{t+1} from the environment. In this way, a RL agent tries to maximize cumulative rewards with feedback (reward) received after taking action 8–10. This RL has been widely applied in a variety of fields such as robotics, healthcare, finance and games such as AlphaGo and Atari, and has been successful in achieving human-level performance or even surpassing humans 8–10.

In this study, our agent was trained to seek an optimal policy with the Dueling Double Deep Q-Network (D3QN) which is a variant of Deep Q-Network among deep RL algorithms. We selected the model to focus on the best action (lockdown and travel policy) for a particular situation (country and territory characteristics and burden of COVID-19) without overestimation of high dimension temporal data. Also, this network was chosen to distinguish the quality of the current state (for countries and territories) and the chosen action (policy) at each time stamp 9,18,19.

After unity-based data normalization, experiments were conducted up to 100,000 episodes using all 13 variables mentioned in Data and Preprocessing section, and the final result was selected at a stabilized convergence point. We used Scikit-learn 0.20.3 library for data preprocessing, and D3QN was adapted from previous research work and modified for this paper with Keras 2.3.1 and Tensorflow 1.15.0 in Python 18,19.

2.2.1. Action and Reward

We defined a 3×3 action space for the domestic lockdown and travel restrictions. The lockdown was divided into three levels: no action (Level 0: L0), restricted public social gathering (L1) and nationwide lock down (L2). Likewise, travel policy covered no action (T0), flight suspension (T1) and closure of all the borders (T2) from each country or territory. Specifically, travel restrictions refer to measures adopted by each country or territory rather than travel bans exerted from others. We focused on how to adjust these interventions on a per-region basis, and also their crucial impact on a country and territory's severity level (the crude death rate on 7th April).

Our rewards were designed to punish accelerated increases in cases of infection and death, and to encourage rapid acceleration of recovery cases with 2:1:1 ratio. Figure 1 shows example code for a compensation formula for reward based on confirmed cases. More details on rewards can be found on Supplementary Fig. S1 to S3. The ratio was experimented arbitrarily and decided with clinical guidance.

2.2.2. Comparison of Actual Policy to Agent Decisions

After training the model to learn the implemented policies and the associated rewards, we derived the suggested initiation date and intensity level of lockdown or travel restriction for each country and territory. Difference in timing of public interventions implemented by governments and those suggested by our model was assessed by comparing the earliest date of either lockdown or travel restriction. Furthermore, we compared the overall timing and intensity level of these interventions by governments and the model over the duration of the pandemic up to 7th April 2020.

3. Results

3.1 Timing of Policies

The actual timing of lockdown and travel restriction policies for each severity level relative to 31st December 2019 and the index case date for each country and territory are shown in Fig. 2 and 3. Even before their index case, 18 countries and territories (5 of them had index case date before mid-March) applied initial lockdown measures in their community (Fig. 2. A). Similarly, 46 countries and territories applied some form of travel restrictions before index case date.

Twenty-one of them reported their first COVID-19 patient before mid-March. Full lockdowns (L2) and closure of all borders (T2) were always only applied after index case date in each country or territory. Overall, early implementation of any and full lockdown and travel restriction policies were associated with

progressively lower levels of crisis severity. These relationships were only apparent when considering timing relative to local index case date in each country or territory (Fig. 2. A, C and Fig. 3. A, C). This associations were not clear using 31st December 2019 as reference date (Fig. 2. B, D and Fig. 3. B, D).

Suggested timing to initiate either at least minimal intensity of lockdown or travel restriction by our model relative to actual policy dates is shown in Fig. 4. In general, the model suggested that at least one of the minimum restrictions should have been applied one and a half months before the actual date it was implemented. Overall, the recommended date for applying the first action at any level was in late January, even if the index case date was in mid- or late March (Fig. 5. A). Interestingly, this coincides with the travel ban from Wuhan, China on 23rd January 2020 5. In addition, proposed action timing from the model did not deviate from the actual implementation dates for some countries and territories (Fig. 5. B). In contrast, for some countries and territories, the agent suggested to delay policy implementation whereas governments took early action even though the number of cases did not grow exponentially (Fig. 5. C).

3.2. Intensity of Restrictions

The distribution of the overall intensity level of lockdown and travel restriction policies implemented by governments and suggested by the model are shown in Fig. 6. In general, the model proposed one intensity level higher for both lockdown and travel restriction than what was actually implemented by governments. Figure 6. (A, C) shows the overall actions taken by government and chosen by the agent from 31st December 2019 to 7th April 2020. No action from government (L0, T0) was common whereas our agent chose earlier and higher intensity of lockdown and travel restrictions over the same period.

Figure 6. (B, D) shows the overall actions taken by government and chosen by agents after index case or initiation of at least one policy (lockdown or travel restriction) at any given level. This is different to Fig. 6. (A, C) because it excludes the time when no action was taken before index case date but includes actions by government or agent with or without index case. This reduced unnecessary emphasis on no action in countries with late index cases. Compared to using 31st December 2019 as starting point, governments adopted higher overall intensity of both lockdown and travel ban after any events occurred (index case or implementation of policies prior to the index case). Nevertheless, no action from government was still the most common decision at 25%. In contrast, the model suggested actions that were one or two levels overall higher than those implemented by governments. Overall, about 90% were proposed to keep one or more policies at least at any level, and about 70% to keep both lockdown and travel restriction policies simultaneously. The minimum lockdown and travel restriction (L1, T1) were most commonly recommended by the model at 24% (Fig. 6. B, D). Specifically, our agent advocated one level higher lockdown from level 0 to 1 and level 1 to 2. For countries and territories that had no public interventions in general despite index cases, the model suggested low intensity lockdown such as public gathering limits or encouraging online e-learning, but stopped short of recommending a full lockdown overall. Similarly, the model recommended travel restrictions that were one to two levels higher and earlier. Nevertheless, earliest and maximal lockdown and travel bans were similar to those implemented by governments at 19%.

4. Discussion

In this study, we used deep RL on country and territory population data and serial local COVID-19 epidemiological data to develop a model to determine the optimal timing and intensity of lockdown and travel restriction for individual countries and territories. We performed timing analysis of policy implementation for each crisis severity and a deep reinforcement learning model with continuous state space and rewards to find the suitable action for each state at a particular point in time. When compared to actual government implementation during the COVID-19 pandemic, our model mostly recommended earlier and higher intensity of lockdown and travel ban.

During an emerging pandemic, it is a challenge to implement timely and appropriate public health interventions with limited data. Early on during the pandemic, SARS-CoV-2 transmission kinetics were unknown. Furthermore, the efficacy and principal of social distancing, lockdowns, and international travel restrictions have been questioned ²⁰. The results in this paper are consistent with previous studies which suggests lockdown and travel restrictions are effective in reducing the transmission of SARS-CoV-2 ^{5,6}. Our model suggests that early lockdown and travel restriction even before index case in individual countries and territories may be optimal. This may help in decision making since the index case may not always be easily identified and implementation would be based on disease burden in other countries and territories. However, adopting policies early to reduce burden of COVID-19 has to be balanced against the economic, social and health concerns ^{21–27}. Even with punishment such as fines and imprisonment for contravening public health policy, it is difficult to sustain lockdowns and border closures over long periods. The model and results of this study suggests sustained, high intensity lockdown and travel restrictions do not need to be applied to all countries and territories. Whilst this is encouraging, this may be because some countries and territories have low floating populations or other defense strategies from other countries or territories have an effect on these countries and territories.

The contributions of this study include the use of deep RL to evaluate the effects of public interventions on spread of COVID-19 using real world epidemiological and population data. This approach utilizes reward based on targets to “flatten the curve” to learn the optimal timing and intensity of policies. RL has previously been used in various fields and also been used to simulate effect of lockdown in COVID-19 ^{7,9,10}. This type of model was chosen to learn sequential decision-making with successive steps rather than supervised learning, which is commonly used but relies on the trusted labeled data. Learning from the observed values when reproducing the given policy implementation, it found the optimal policies according to the temporal and population characteristics unique to each country and territory. The result is an individualized recommendation on timing and intensity of lockdown and travel restriction for each country and territory.

The limitation of the current paper is that it was carried out with imperfect data due to the emerging COVID-19 pandemic, inconsistent testing and reporting. It is expected that more solid results will be obtained as we learn more about the transmission kinetics of the virus, the clinical characteristics of COVID-19, consistent testing and higher fidelity population data. In addition, in some countries there was

additional provincial data collected, but country-level data had to be used to maintain consistency and avoid problems caused by incomplete data. We were also unable to analyze the impact on individual travel restrictions on other countries and territories. With more detailed data on travel restrictions it may be possible to separate instances where travel bans between countries and territories have influenced each other. Only the official lockdown policies were available, and the official policy may differ from those practiced in the community. Lastly, we were not able to strike a balance between policy decisions for public health and negative impacts such as economic consequences as these remain to be determined. This means our research focuses only the population health benefits of controlling the spread of COVID-19. Nevertheless, we have shown that RL may be used to learn the effect of public health interventions.

5. Conclusion

In this study, we used deep RL to learn efficacy of lockdown and travel restrictions in controlling the COVID-19 crisis. Using local population and COVID-19 epidemiological data, we showed that the model can be trained to find the optimal strategy in specific countries and territories to maximize the expected value of total rewards over time. Compared to actual government policy implementation, the agent mainly proposed to have earlier and higher intensity of lockdown and travel restrictions to reduce the burden of COVID-19.

Declarations

Acknowledgments

We are grateful to all individuals who collect and share epidemiological data on COVID-19 outbreaks worldwide.

Data Availability

Raw data can be downloaded from each cited database or repository, and processed data and analysis files will be provided on github: <https://github.com/ucabhkw/COVID19RL20>.

Author Contributions

G.H.K. conceived and designed the study and performed data acquisition/experiments. G.H.K. and L.L. analyzed experimental data and provided substantial input to the manuscript, and P.H. approved the final version.

Competing Interests statement

Authors declare no competing interests.

References

1. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697 (2020).
2. The Center for Systems Science and Engineering, Johns Hopkins. Coronavirus COVID-19 Global Cases. <https://coronavirus.jhu.edu>.
3. Cascella, M., Rajnik, M., Cuomo, A., Dulebohn, S. C. & Di Napoli, R. Features, evaluation and treatment coronavirus (COVID-19). in *Statpearls [internet]* (StatPearls Publishing, 2020).
4. Kucharski, A. J. *et al.* Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* (2020) doi:10.1016/S1473-3099(20)30144-4.
5. Chinazzi, M. *et al.* The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* (80-.). **368**, 395–400 (2020).
6. Tian, H. *et al.* An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* (80-.). **368**, 638–642 (2020).
7. Khadilkar, H., Ganu, T. & Seetharam, D. P. Optimising Lockdown Policies for Epidemic Control using Reinforcement Learning. *arXiv Prepr. arXiv2003.14093* (2020).
8. Sutton, R. S. & Barto, A. G. Reinforcement learning: An introduction. (2011).
9. Li, Y. Reinforcement learning applications. *arXiv Prepr. arXiv1908.06973* (2019).
10. Yu, C., Liu, J. & Nemati, S. Reinforcement learning in healthcare: a survey. *arXiv Prepr. arXiv1908.08796* (2019).
11. World Health Organization. Novel Coronavirus – Situation <https://www.who.int/emergencies/diseases/novel-cor> (2020).
12. The Center for Systems Science and Engineering, Johns Hopkins. Coronavirus COVID-19 Data Repository. <https://github.com/CSSEGISandData/COVID-19> (2020).
13. United Nations. United Nations Data Retrieval System. <https://data.un.org> (2020).
14. Korea Centers for Disease Control and Prevention. The updates on novel Coronavirus infection in Korea. <https://www.cdc.go.kr> (2020).
15. 2019-20 coronavirus pandemic. https://en.wikipedia.org/wiki/2019%E2%80%9320_coronavirus_pandemic (2020).
16. Max, R., Hannah, R., Esteban, O., Joe, H. Coronavirus Disease Data. <https://ourworldindata.org/coronavirus> (2020).
17. Rotejanaprasert, C., Ekapirat, N., Areechokchai, D. & Maude, R. J. Bayesian spatiotemporal modeling with sliding windows to correct reporting delays for real-time dengue surveillance in Thailand. *J. Health Geogr.* **19**, 1–13 (2020).
18. Wang, Z. *et al.* Dueling network architectures for deep reinforcement learning. *arXiv Prepr. arXiv1511.06581* (2015).
19. Raghu, A. *et al.* Deep reinforcement learning for sepsis treatment. *arXiv Prepr. arXiv1711.09602* (2017).
20. Xiao, Y. & Torok, M. E. Taking the right measures to control COVID-19. *Lancet Dis.* (2020).

21. Nicola, M. *et al.* The Socio-Economic Implications of the Coronavirus and COVID-19 Pandemic: A Review. *J. Surg.* (2020).

22. Holmes, E. A. *et al.* Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. *The Lancet Psychiatry* (2020).

23. Cetron, M. & Landwirth, J. Public health and ethical considerations in planning for quarantine. *Yale J. Biol. Med.* **78**, 329 (2005).

24. Wilder-Smith, A. & Freedman, D. O. Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *Travel Med.* **27**, taaa020 (2020).

25. Sohrabi, C. *et al.* World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *J. Surg.* (2020).

26. Reger, M. A., Stanley, I. H. & Joiner, T. E. Suicide Mortality and Coronavirus Disease 2019—A Perfect Storm? *JAMA psychiatry* (2020).

27. Qiu, J. *et al.* A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: implications and policy recommendations. *psychiatry* **33**, (2020).

Table

Table 1. Table of population, life expectancy and GDP for each severity group; All values are expressed in median and interquartile range unless specified; Population: Population estimates (millions); Population (mid-year): Population mid-year estimates (millions); Population density (/km2); Life expectancy (years); GDP: GDP (PPP) per capita (USD); Crude death rate: the number of deaths related to COVID-19 in the total population reported on 7th April.

Severity level	Low	Medium	High	Critical
Population	1.45 (0.18 - 14.97)	31.12 (10.53 - 78.93)	42.51 (6.15 - 422.4)	8.19 (0.38 - 20.01)
Population (mid-year)	0.89 (0.11 - 8.78)	17.58 (6.42 - 42.81)	31.95 (4.13 - 270.63)	4.88 (0.26 - 11.54)
Population density	78 (24.62 - 299.66)	84.64 (25.53 - 220.65)	106.98 (46.58 - 152.72)	129.72 (70.87 - 330)
Life expectancy	71.74 (67.87 - 77.9)	69.86 (64.88 - 75.01)	77.44 (75.41 - 77.47)	81.88 (79.11 - 83.13)
GDP (x103)	7.45 (1.97 - 20.78)	5.57 (3.03 - 17.64)	18.24 (13.08 - 25.22)	35.97 (16.36 - 53.21)
Crude death rate (x10-6)	0 (0 - 0)	0.15 (0.07 - 0.22)	1.52 (1.58 - 1.88)	13 (6.14 - 29.92)

Figures

for every region in dataset:

if the rate of increase in the current confirmed case is faster than the rate of increase in the previous case :

then negative reward

else if the rate of increase in the current confirmed case has not changed from the rate of increase in the previous case after one or more policy actions :

then positive reward

else if the rate of increase in the current confirmed case is slower than the rate of increase in the previous case :

then positive reward

Figure 1

Example for a compensation formula of the confirmed case; If each country and territory had positive or negative acceleration in growth of case, it was rewarded accordingly; For conditions where there was no change in growth rate, positive reward was considered only if there was at least one action or one or more confirmed cases were found to reduce long-term no action impact before the first confirmed case was reported.

Figure 2

Distribution of lockdown policy date; A. Any lockdown policy date from the index case date for each country and territory and B. from 31st December 2019; C. Full lockdown policy date from the index case date for each country and territory and D. from 31st December 2019.

Figure 3

Distribution of travel restriction policy date; A. Any travel restriction policy date from the index case date for each country and territory and B. from 31st December 2019; C. Full travel restriction policy date from the index case date for each country and territory and D. from 31st December 2019.

Figure 4

Distribution of suggestions for action before or after the any first policy date.

Figure 5

Example of when government and agent are first involved at any level.

Figure 6

Lockdown and travel restriction policy from government and agent; A. Full period government policy from January to April; B. Government policy only the action after the infection case occurred or at least one limitation given at any level; C. Full period agent policy from January to April; D. Agent policy only the action after the infection case occurred or at least one limitation given at any level; Each tick label for lockdown and travel restriction (0, 1, 2) indicates policy level; Each number represents the percentage of each combination of actions that represents importance (e.g. darker is more important).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SRCOVIDSupplementalMaterial.pdf](#)