ORIGINAL PAPER

Risk of a second wave of Covid-19 infections: using artificial intelligence to investigate stringency of physical distancing policies in North America

Shashank Vaid[1] · Aaron McAdie[2] · Ran Kremer[3] · Vikas Khanduja[4] · Mohit Bhandari[5]

Abstract
Purpose Accurately forecasting the occurrence of future covid-19-related cases across relaxed (Sweden) and stringent (USA and Canada) policy contexts has a renewed sense of urgency. Moreover, there is a need for a multidimensional county-level approach to monitor the second wave of covid-19 in the USA.
Method We use an artificial intelligence framework based on timeline of policy interventions that triangulated results based on the three approaches—Bayesian susceptible-infected-recovered (SIR), Kalman filter, and machine learning.
Results Our findings suggest three important insights. First, the effective growth rate of covid-19 infections dropped in response to the approximate dates of key policy interventions. We find that the change points for spreading rates approximately coincide with the timelines of policy interventions across respective countries. Second, forecasted trend until mid-June in the USA was downward trending, stable, and linear. Sweden is likely to be heading in the other direction. That is, Sweden's forecasted trend until mid-June appears to be non-linear and upward trending. Canada appears to fall somewhere in the middle—the trend for the same period is flat. Third, a Kalman filter based robustness check indicates that by mid-June the USA will likely have close to two million virus cases, while Sweden will likely have over 44,000 covid-19 cases.
Conclusion We show that drop in effective growth rate of covid-19 infections was sharper in the case of stringent policies (USA and Canada) but was more gradual in the case of relaxed policy (Sweden). Our study exhorts policy makers to take these results into account as they consider the implications of relaxing lockdown measures.

Keywords Artificial intelligence · Covid-19 · Bayesian (SIR) · Kalman filter · Machine learning

Understanding the implications of covid-19 policy interventions remains a significant hurdle [1]. As of mid-May, on the one hand, the covid-19 pandemic continues to devastate economies—some estimates indicate that under current circumstances the economic downturn could be as much as 2.0% global economic growth per month [1]. On the other hand, as of May 18, 2020, over 4.8 million people have been infected and over 318,000 have died [2]. Clearly, policy makers are struggling to balance trade-offs between economic recovery and infection-related mortality [3]. While stringency in policy interventions appears to be now globally accepted, there remain variations in government responses to a phased implementation and reduction in physical distancing policies. Sweden acted relatively slowly seeking voluntary physical distancing from citizens. But Germany, South Korea, Hong Kong, among others followed a more stringent approach toward physical distancing [4, 5]. The USA and Canada also broadly followed a more stringent policy intervention.

Attempts are now being made to visualize counterfactual scenarios that compare the impact of less and more intervention strategies tailored around social distancing, school closures, and border control [6]. Efforts are also underway to systematically track and compare worldwide government responses through a stringency index that monitors 16 interventions in over 100 countries [7]. While various organizations

✉ Shashank Vaid
  vaids1@mcmaster.ca

1 DeGroote School of Business, McMaster University, 1280 Main St. West, Hamilton, Ontario L8S 4M4, Canada

2 Clinical Analytics, Seattle Children's Hospital, Seattle, WA, USA

3 Kofkin Faculty of Engineering, Bar-Ilan University, Ramat Gan, Israel

4 Addenbrooke's Hospital, University of Cambridge, Cambridge, UK

5 Department of Surgery, Division of Orthopaedic Surgery, McMaster University, Hamilton, Ontario, Canada

are establishing mechanisms to monitor policy interventions, leading scientific opinion and commentaries appear to be struggling with how various policy interventions, biased in favour of either relaxed or stringent policies, are impacting occurrence of future cases. Furthermore, in the current landscape of widespread economic hardship and quarantine fatigue, many states in the USA are drawing down suppression measures, sometimes even in spite of increases in daily new case-loads. These conditions in concert create a risk for a large second wave of infections. Close monitoring of where health care systems are currently being strained or have excess capacity, as well as where new cases are increasing or decreasing, could help inform wise policy choices at a local level.

Thus, it appears that accurately forecasting the occurrence of future covid-19-related cases across relaxed (Sweden) and stringent (United States and Canada) policy contexts has a renewed sense of urgency. Moreover, there is a need for a multidimensional county-level approach to monitor the second wave of covid-19 in the USA. We developed a robust modeling framework using multi-method artificial intelligence approach based on timeline of policy interventions to evaluate impacts of stringency on a potential second wave of infections.

Method

We modeled influence of covid-19 policy intervention bias and present a one month forecast of covid-19 cases across relaxed (Sweden) and stringent (United States and Canada) contexts. Though we do not study the impact of individual policies, we do account for the timelines of governmental interventions that cluster various policies related to covid-19. Such timelines proxy for the presence or absence of stringency. We see stringency through clustering of policies around dates of key policy interventions, in USA (March 12, 2020, March 23, 2020, and April 7, 2020), in Canada (March 12, 2020, March 20, 2020, and March 31, 2020), and in Sweden (March 1, 2020, March 12, 2020, and March 30, 2020). We use an Artificial Intelligence framework that triangulated results based on the three approaches—Bayesian susceptible-infected-recovered (SIR), Kalman filter, and machine learning. We identified timelines of policy interventions based on publicly available information [6, 7].

The policy interventions included in our analysis are school closing, workplace closing, cancelled public events, restriction on gatherings, closing public transport stay at home requirements, restrictions on internal movements, international travel controls, income support, debt contract relief, international support, public information campaigns, testing policy, contact tracing, emergency investment in health care, and investment in vaccines. Of these, Sweden had not actioned the following policy interventions at all: cancelled public events,

closed public transport, and stay at home requirements. Of those that were actioned, Sweden did better than its North American counterparts only in income support, debt contract relief, and contact tracing. Overall, Sweden's policy has been much more relaxed. However, relative stringency of these policies across the USA, Canada, and Sweden measured by the Oxford COVID-19 government response tracker (OxCGRT) stringency index can be seen in Fig. 1 [7].

Bayesian SIR model

We started by performing Bayesian inference for the key epidemiological parameters of an SIR model using Markov chain Monte Carlo (MCMC) sampling [8]. To model influence of covid-19 policy intervention bias and present a one month forecast of covid-19 cases across stringent (USA and Canada) and relaxed (Sweden) policy interventions, we built on outstanding work by Dehning et al. (2020) (Figs. 2a, 3a, and 4a). The Bayesian SIR model helped us investigate influence of government response measures as seen through timelines associated with relaxed and stringent policy interventions. We assumed a temporal change in the covid-19 spreading rates across these jurisdictions as a function of specific timeline of policy interventions. We calculated the spreading rate based on (a) key policy interventions clustered around specific timelines (i.e., dates), (b) reproduction numbers (R_0) for Sweden, USA, and Canada on respective dates, and (c) recovery rate that had a median recovery time of eight days [6–9] The key parameters of the model included the following: three normally distributed change points indicating specific policy interventions, three log-normally distributed spreading rates across aforementioned change points, log-normally distributed recovery rate, log-normally distributed reporting delay of eight days, and initially infected population of 100.

Kalman filter

Next, we took a robustness check to test the Bayesian SIR model's results using a Kalman filter approach. Kalman filter was pioneered by Rudolf Emil Kalman in 1960, originally designed and developed to solve the navigation problem in Apollo Project. Since then, it has found numerous technology-based applications related to guidance, navigation, vehicle control, object tracking, trajectory optimization, time series analysis in signal processing, econometrics, among others. To forecast covid-19 cases across USA, Canada, and Sweden over a month, we used Kalman filter's recursive algorithm that used time series measurement while accounting for statistical noise, and produced estimations of unknown variables (Figs. 2b, 3b, and 4b). We used daily cases (confirmed, death, and recovered) as a time series and fit a model that tracked the series. Our data was derived from Johns

Fig. 1 Mapping the Oxford COVID-19 government response tracker (OxCGRT) stringency index across USA, Canada, and Sweden
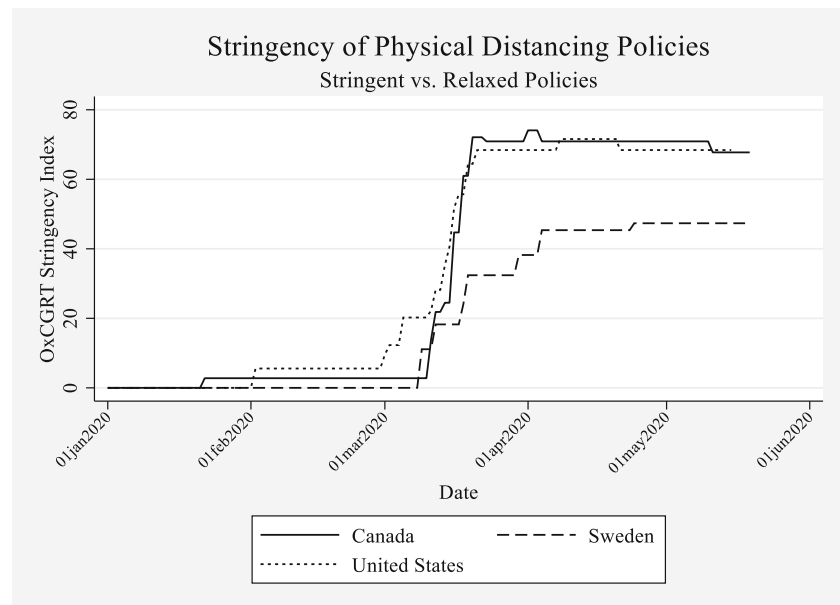
Hopkins University's Coronavirus Resource Center. Our algorithm was adaptive as it did not need a lot of historical/training data. Each day the algorithm was updated with new observation; after fast parameter estimation, it generated predictions for the next timestamp. To predict longer periods, we imitated an iterative online system that trained the model every day and predicted the next day. Then, instead of getting the feedback of the real value, we updated the next timestamp value with the last prediction plus some calculated noise, and predicted again. Eventually, we fitted a linear model to predict the spread of COVID-19 along the time where the Kalman predictors were the main features. See equations in the Appendix.

## Inferring potential for second wave across US counties

Finally, while we forecasted covid-19 cases across Sweden, USA, and Canada over a month, we also extrapolated available information using a multidimensional county-level machine learning approach to monitor the second wave of covid-19 in the USA. To that end, we projected all US counties with more than 100,000 residents onto axes of current hospital capacity and week-over-week change in new cases (Fig. 2c). The primary data source for our analysis was the New York Times data repository maintained on GitHub. This data source gave by county daily cumulative case and death counts. We differenced the vector of daily cumulative cases to get new cases per day and calculated the rolling sum and mean of that quantity across various time windows to support further analysis. We then incorporated hospital bed count data from the US Homeland Infrastructure Foundation-Level Data (HFLID) repository. We processed this dataset into a total count of

general acute care and critical care beds by county and joined it to the daily cases dataset. With daily cases and total hospital beds, we then built metrics to estimate current bed capacity and change in new cases for each county. Estimating bed capacity required assumptions of the rate of hospitalization among reported covid-19 cases, as well as an assumption of the proportion of total hospital beds that are available to treat covid-19 patients, and finally an assumption of length of stay in the hospital. We assumed that 20% of reported cases would require hospitalization based on a survey of states that report total hospitalizations. Next, we assumed that 39% of total beds would be available to treat covid-19 patients. This assumption was informed by taking the mean of available beds across hospital referral regions in a dataset put together by The Harvard Global Health Institute (HGHI). Finally, a 12-day length of stay was assumed, also informed by the HGHI methods.

The 12-day rolling sum of new cases, starting from three days prior to mitigate reporting delays, multiplied by the hospitalization rate gave an estimate of the number of currently hospitalized patients in each county. We subtracted that quantity from the estimate of available beds and then divided by available beds to get a proportion of open beds to available beds. We then created a metric for the change in new reported cases over time. We investigated multiple time horizons but found the weekly interval to be both reasonably low latency and stable. The daily data tended to have a lot of variability, which was smoothed out by summing cases over a week. We again indexed the rolling sum three days back to mitigate reporting delays.

With these two metrics created, we leveraged a two-dimensional plane to project the county-level data. In order to provide a geographic context, we mapped a colour dimension to
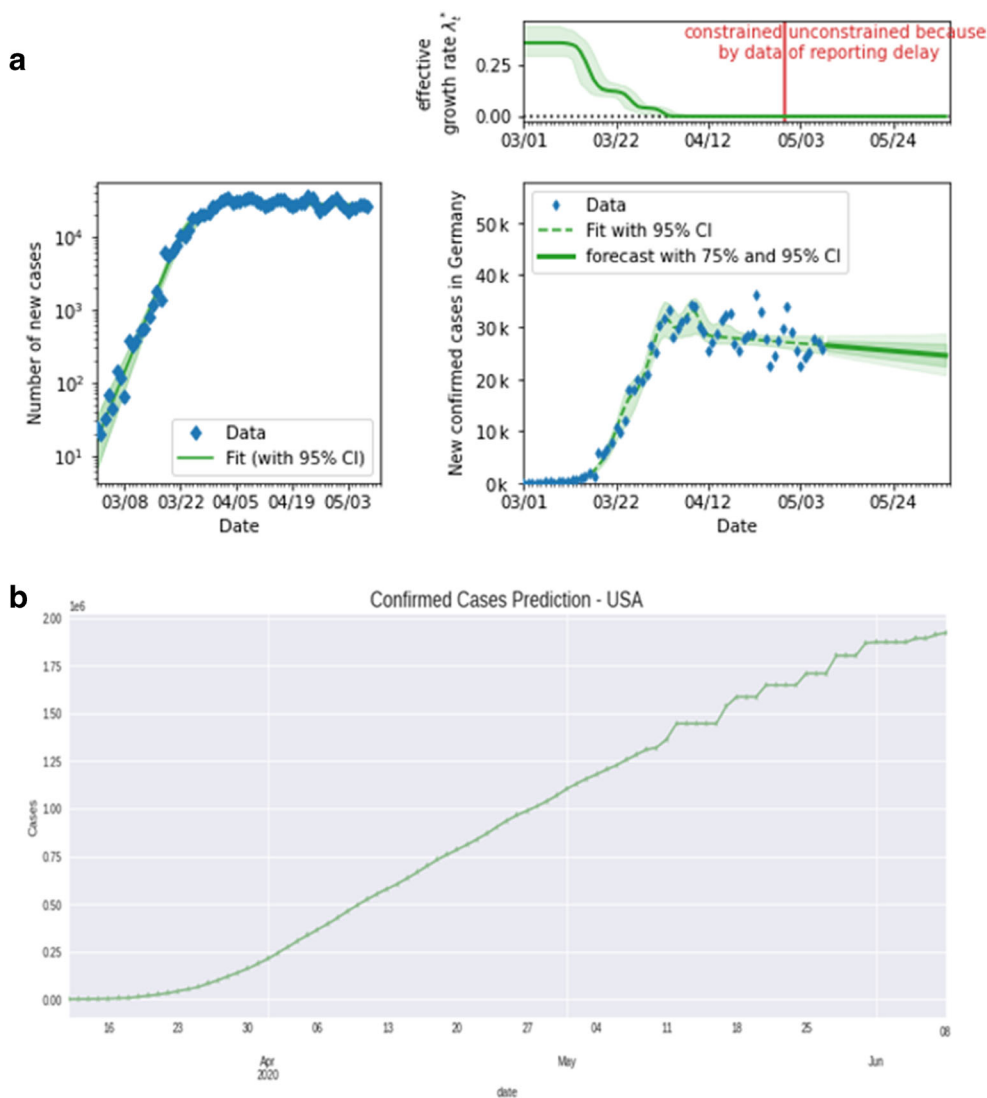
Fig. 2 **a** (Clockwise) Effective growth rate of covid-19 cases for the USA, Bayesian SIR model forecast of new confirmed covid-19 cases for the USA, and number of new covid-19 cases in the USA. **b** Kalman filter prediction of confirmed covid-19 cases in the USA. **c** Estimated bed capacity relative to weekly change in new cases in US counties. USA counties that fall in the top right quadrant of the chart (e.g., New England, Mountain States, Midwest, and South Atlantic) may be at the most risk for a surge in infections

the nine US Census Bureau regions. Finally, we joined US Census Bureau population estimates from the 2017 census and mapped that data to the size dimension. A static snapshot of the final chart is shown below. See equations in the Appendix.

Results

As of May 10, 2020, the forecasted trend until mid-June in the USA was downward trending, stable, and linear. Furthermore, the effective growth rate of covid-19 infections dropped in response to the approximate dates of key policy interventions. Specifically, the first change point for the spreading rate may have occurred approximately around March 15, 2020, and the second change point may have occurred approximately around March 25, 2020. Regardless of the impacts, exponential growth

in the USA was still implied, and a Kalman filter robustness check projects that the USA is likely to have close to two million covid-19 cases by mid-June. US counties in New England and Mountain states may be at the most risk for a second wave surge in infections. However, counties in East North Central and West North Central (typically referred to as the Midwest) and in South Atlantic USA (North Carolina, Virginia) are also at risk of a second wave (Figs. 2a, b, and c).

In Canada, the forecasted trend until mid-June appears to be more stable and linear. Infection rates declined coincident with Canada's relatively stringent policy interventions. However, Kalman filter results indicate that Quebec will still be seeing a growth in covid-19 infections (Figs. 3a, and b).

Sweden's forecasted trend until mid-June appears to be non-linear and upward trending. Furthermore, the drop in the effective growth rate of covid-19 infections appeared to
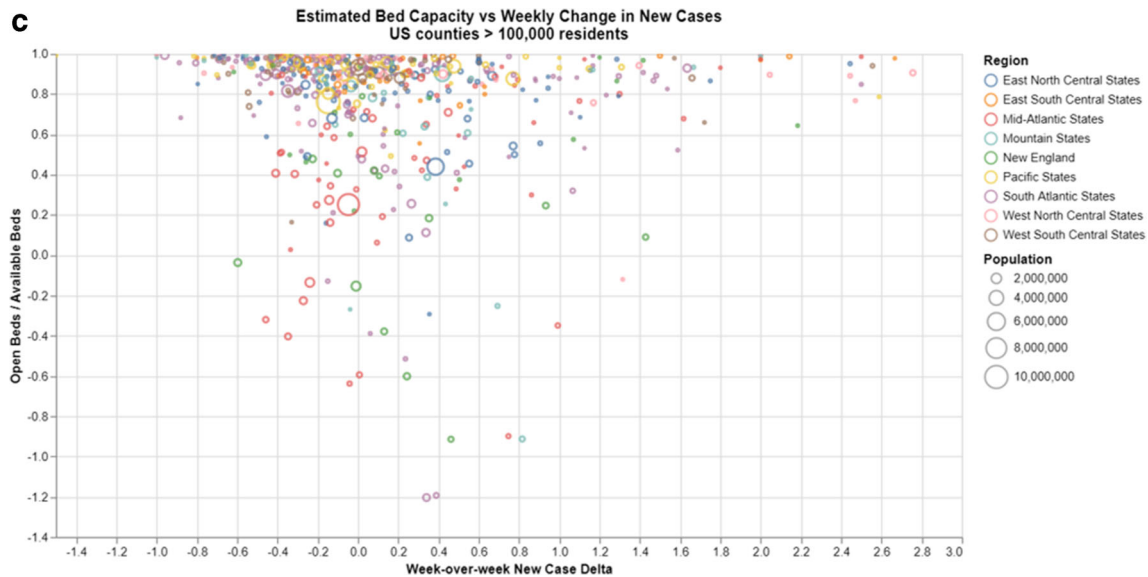
Fig. 2 (continued)

have been only gradual around the dates of key policy inter-
ventions. Our findings suggest that infections did not slow
down as rapidly in Sweden compared with USA and
Canada. We infer that the likely reason is Sweden's relatively
relaxed policy intervention (Figs. 1, 4a, b).

Discussion

Using an Artificial Intelligence framework, we modeled influ-
ence of covid-19 physical distancing policies across relaxed
(Sweden) and stringent (USA and Canada) contexts. Our find-
ings suggest three important insights. First, the effective growth
rate of covid-19 infections dropped in response to the approx-
imate dates of key policy interventions. The drop was sharper in
the case of the USA and Canada that actioned stringent policies
but was more gradual in the case of Sweden that actioned a
more relaxed policy. We find that the change points for spread-
ing rates approximately coincide with the timelines of policy
interventions across respective countries. Second, forecasted
trend until mid-June in the USA was downward trending, sta-
ble, and linear. Sweden is likely headed in the other direction.
That is, Sweden's forecasted trend until mid-June appears to be
non-linear and upward trending. Canada appears to fall some-
where in the middle—trend for the same period is flat. Third, a
Kalman filter based robustness check indicates that by mid-
June the USA will likely have close to two million virus cases,
while Sweden will likely have over 44,000 covid-19 cases.

Specifically, we studied 16 policy interventions in the
USA, Canada, and Sweden that ranged from school and work-
place closing and restrictions on public gatherings and trans-
port to testing and contact tracing policies. We found that
Sweden had neither cancelled public events and closed public
transport nor actioned stay at home requirements. Sweden

outperformed its North American counterparts only in income
support, debt contract relief, and contact tracing. Overall,
Sweden's policy has been much more relaxed and this could
drive a higher number of covid-19 cases by mid-June.

In the case of the USA, the Bayesian SIR model predicts that
by mid-June, there will likely be between 20,000 and 28,000
covid-19 new confirmed cases reported daily. However, rela-
tive to the past, the forecasted trend appears to be more down-
ward trending, stable, and linear. Furthermore, we notice that
the effective growth rate of covid-19 infections appears to drop
around dates seeing key policy interventions characterized by
Dehning et al. (2020) as forms of mild distancing, strong dis-
tancing, and contact ban on March 12, 2020, March 23, 2020,
and April 7, 2020, respectively. Thus, as of May 10, 2020, we
find evidence that the virus was clearly slowed down through
USA's relatively stringent policy interventions. Coinciding
with such policy interventions, the first change point for the
spreading rate may have occurred approximately around
March 15, 2020, and the second change point may have oc-
curred approximately around March 25, 2020. However, it can
be fairly assumed that despite these actions, exponential growth
was still implied. As a robustness check, our Kalman filter
approach indicates that the USA is likely to have close to two
million covid-19 cases by mid-June. However, the Kalman
filter also does indicate a gradual flattening of the curve.

Furthermore, we believe that when the human element is
considered, US counties that fall in the top right quadrant of
the chart (e.g., New England, Mountain States, and Midwest)
may be at the most risk for a surge in infections. These are
places where hospitals currently have capacity and decision
makers might feel better with relaxing suppression measures,
even though new cases are currently on the rise.

In the case of Canada, the Bayesian SIR model predicts that
by mid-June, the country may still have between 1000 and
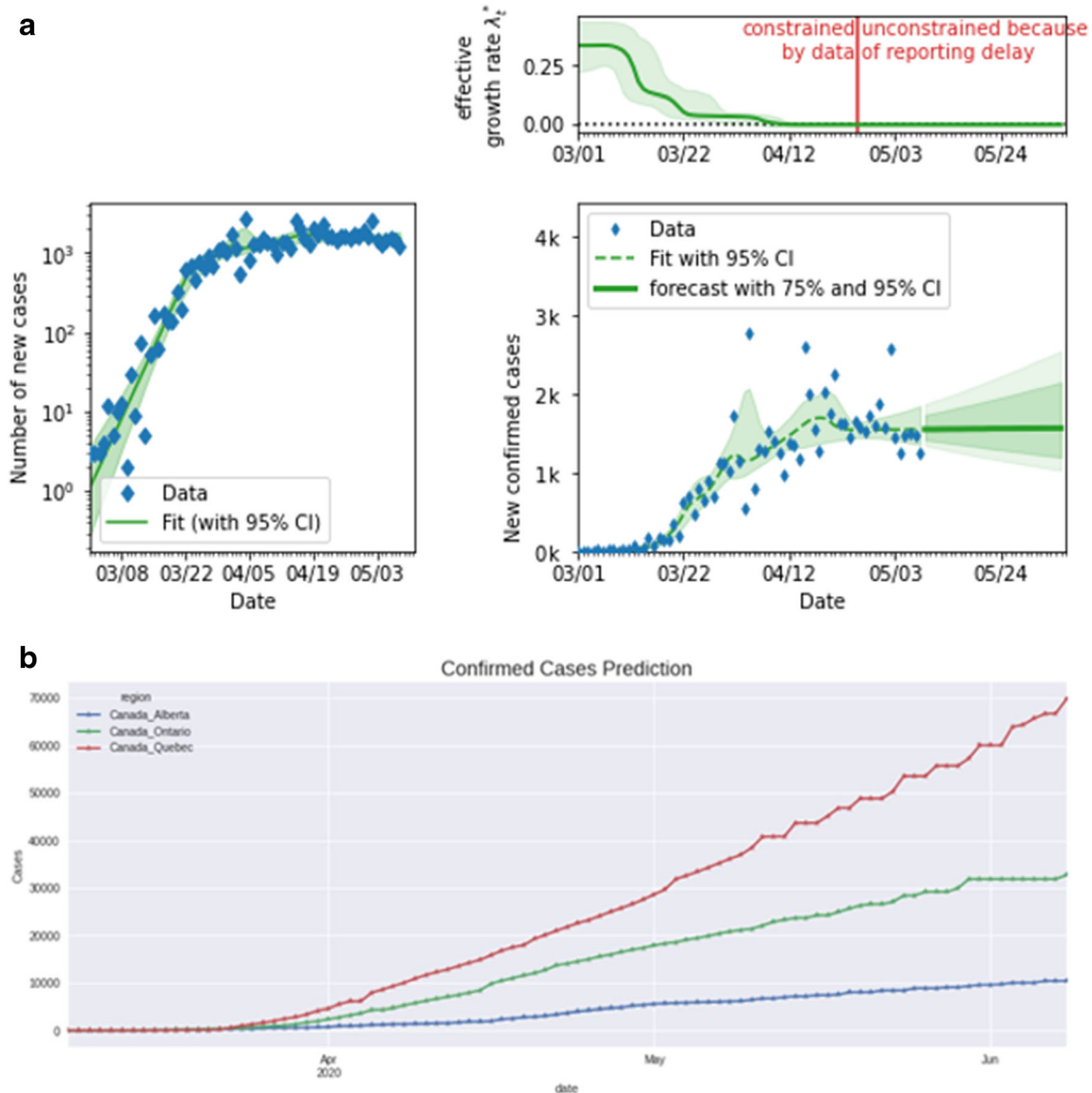
Fig. 3 **a** (Clockwise) Effective growth rate of covid-19 cases for Canada, Bayesian SIR model forecast of new confirmed covid-19 cases for Canada, and number of new covid-19 cases in Canada. **b** Kalman filter prediction of confirmed covid-19 cases in Alberta, Ontario and Quebec

2500 covid-19 new confirmed cases reported daily. Thus, as of May 10, 2020, we found evidence that the rate of new infections was decreasing (i.e., flattening of the curve) through Canada's relatively stringent policy interventions on March 12, 2020, March 20, 2020, and March 31, 2020. Coinciding with such policy interventions, the first change point for the spreading rate may have occurred approximately around March 14, 2020, and the second change point may have occurred approximately around March 22, 2020. However, it can be fairly assumed that despite these actions exponential growth was still implied. To progress toward flattening the curve, Canada needed to undertake the third policy intervention, which it did take around March 31, 2020. As a robustness check, our Kalman filter approach indicates that the Canadian provinces of Alberta, Ontario, and Quebec are likely to have > 10,000, > 32,000, and > 69,000 covid-19 cases by mid-June. Of

these Canadian provinces, only Quebec appears to still be showing a growth in covid-19 infections.

In the case of Sweden, the Bayesian SIR model predicts that by mid-June, the country may still have upward trending of 600 covid-19 new confirmed cases reported daily. However, the forecasted trend appeared to be non-linear and upward trending. Furthermore, we noticed that drop in the effective growth rate of covid-19 infections appeared to have been only gradual around dates seeing key policy interventions, as described above (March 1, 2020, March 12, 2020, and March 30, 2020). Thus, as of May 10, 2020, we find evidence that the virus did not slow down as rapidly in Sweden. We infer that the likely reason is Sweden's relatively relaxed policy intervention. Coinciding with such relaxed policy interventions, the first change point for the spreading rate may have occurred approximately around March 7, 2020, and

Fig. 4 **a** (Clockwise) Effective growth rate of covid-19 cases for Sweden; Bayesian SIR model forecast of new confirmed covid-19 cases for Sweden; and Number of new covid-19 cases in Sweden. **b** Kalman filter prediction of confirmed covid-19 cases in Sweden
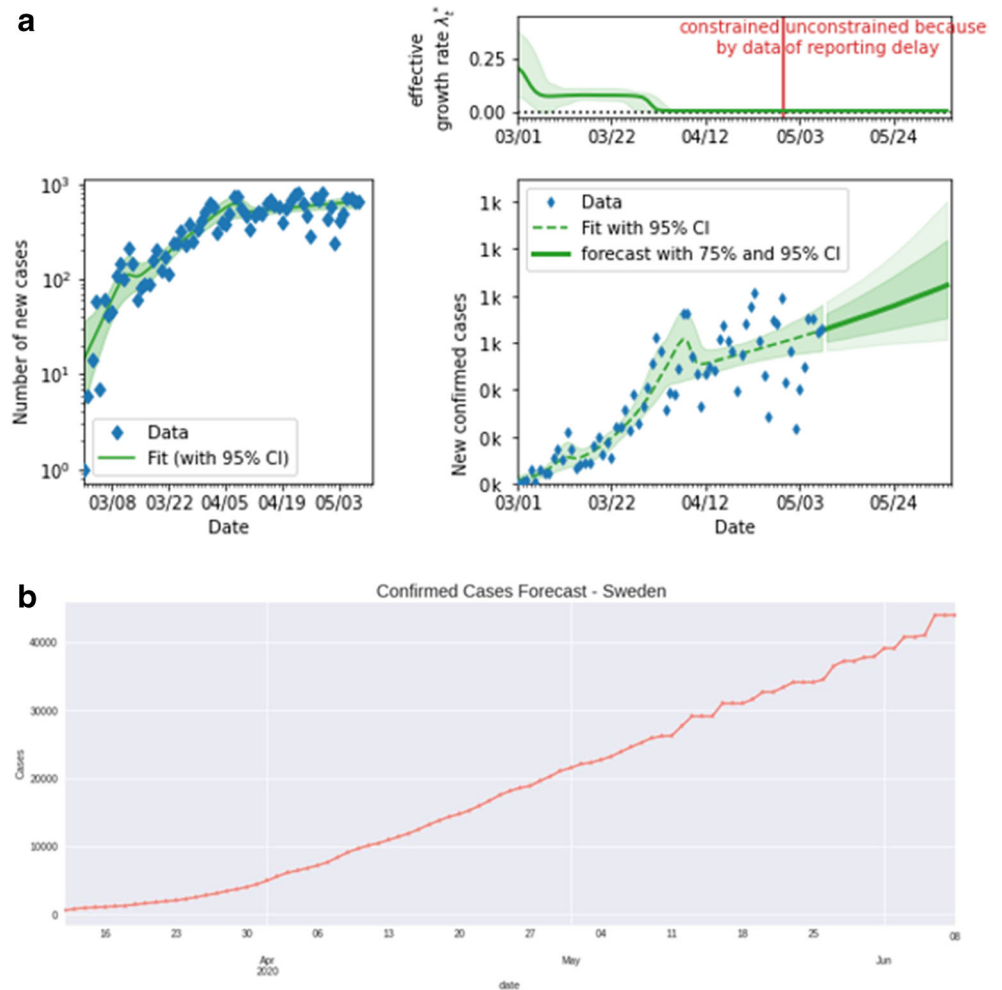
the second change point may have occurred approximately around March 27, 2020. However, despite these actions exponential growth was still implied. As a robustness check, our Kalman filter approach indicates that Sweden will likely have > 44,000 covid-19 cases by mid-June.

Our analysis has several strengths. We show that stringent physical distancing policies are influencing the downward covid-19 trend (as in the case of USA) or are flattening the curve (as in the case of Canada). We also find that the relaxed physical distancing policy might be influencing the upward covid-19 trend (as in the case of Sweden). We base our analysis on changing covid-19 spreading rates triggered by specific policy interventions. Our greatest limitation is that we are making a forecast on a rapidly evolving scenario. Though we forecast the big picture and triangulate our analysis using a novel AI framework based on timeline of policy interventions, we may be missing out on impact of individual policies across the USA, Canada, and Sweden.

Furthermore, we also make a novel rapid assessment of the current status of the outbreak. We determine US counties where hospital resources are stretched thin, those with capacity, and those with large increases or decreases in new cases. The data

visualization could be operationalized for decision making by setting thresholds for turning suppression measures on or off based on hospital capacity and new cases. These thresholds would correspond to regions of the chart, and counties falling within those regions could consider the policy recommendation. We feel that the assumptions made to calculate the metrics are reasonable given the landscape of uncertainty and lack of reliable data on true infection and adverse outcome rates; however, there are several improvements required to make the tool sufficiently robust to inform policy decisions. We know that age and presence of comorbidities have a significant effect on hospitalization rate, so a model of county-by-county hospitalization rate that used those factors as inputs could improve the estimate of currently hospitalized patients. Similarly, bed capacity could be modeled at the county level, analogous to the work that the HGHI team did for hospital referral regions. Lastly, the week-over-week case delta metric currently assumes a stable rate of testing and is susceptible to shifts when that assumption is violated. It could be improved to account for varying test rates if that data was available at the county level.

In conclusion, our study has made some headway in understanding the implications of covid-19 policy interventions.

Though we do not study the impact of individual policies, we do account for the timelines of governmental interventions that cluster various policies related to covid-19. We show that fall in effective growth rate of covid-19 infections was sharper in the case of the USA and Canada that actioned stringent policies but was more gradual in the case of Sweden that actioned a more relaxed policy. Our study exhorts policy makers to take these results into account as they consider the implications of relaxing lockdown measures.

Compliance with ethical standards

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

APPENDIX

Equations for Kalman filter

The following equations are based on Bar Ilan University module, 'Algorithms for Statistical Signal Processing'. We greatly appreciate Prof. Sharon Gannot's generous support in this regard. Kalman filter is a recursive algorithm suitable for dealing with multivariable systems, time varying systems, and non-stationary random processes. It can overcome the shortcomings and limitations of the classical Wiener filtering theory. Consider the dynamic system as follows:

$$d_n = \Phi_n d_{n-1} + w_n; \quad n = 1, 2, \cdots$$

Where d_n is the state vector, Φ_n state transition matrix and w_n is the model noise.

The initial conditions are given d_0 a random vector where

$$E\{d_0\} = m_0$$
$$E\left\{(d_0-m_0)(d_0-m_0)^T\right\} = P_0$$

w_n model noise is a white noise where

$$E\{w_n\} = 0$$
$$E\{w_n w_m^T\} = \begin{cases} Q_n & m = n \\ 0 & m \neq n \end{cases}$$
$$E\{w_n d_0^T\} = 0; \quad n = 1, 2, \cdots$$

The measurement equation is given as follows:

$$y_n = H_n d_n + v_n$$

Where H_n is the measurement matrix and v_n is the measurement noise.

Let $\widehat{d}_{n|n}$ be the estimator of d_n from the measurements

$$y_n; \quad n = 1, 2, 3, \cdots$$

The following are the estimation error and the covariance:
$$e_{n|n} = \widehat{d}_{n|n} - d_n$$

$$P_{n|n} = E\left\{\left(\widehat{d}_{n|n}-d_n\right)\left(\widehat{d}_{n|n}-d_n\right)^T\right\} = E\left\{e_{n|n}e_{n|n}^T\right\}$$

The posteriori estimate covariance matrix.

Initialization:

$$\widehat{d}_{0|0} = E\{d_0\} = m_0$$
$$P_{0|0} = E\left\{(d_0-m_0)(d_0-m_0)^T\right\} = P_0$$

Propagation equations:

$$\widehat{d}_{n|n-1} = \Phi_n \widehat{d}_{n-1|n-1}$$
$$P_{n|n-1} = \Phi_n P_{n-1|n-1} \Phi_n^T + Q_n$$

Update equations:

$$\widehat{d}_{n|n} = \widehat{d}_{n|n-1} + K_n\left(y_n - H_n\widehat{d}_{n|n-1}\right)$$
$$P_{n|n} = (I-K_nH_n)P_{n|n-1}(I-K_nH_n)^T + K_nR_nK_n^T$$
$$= (I-K_nH_n)P_{n|n-1}$$

Kalman gain

$$K_n = P_{n|n-1}H_n^T\left(H_nP_{n|n-1}H_n^T + R_n\right)^{-1}$$

Equations to infer potential for second wave across US counties

The calculation for estimated bed capacity is shown below.

$$\text{Est.Capacity} = \frac{\text{total beds}*0.39 - \sum_{i=-15}^{-3}\text{new cases}_i*0.2}{\text{total beds}*0.39}$$

The proportional change in weekly cases was calculated as shown below.

$$\Delta_{\text{new cases}} = \frac{\sum_{i=-18}^{-11}\text{new cases}_i - \sum_{i=-10}^{-3}\text{new cases}_i}{\sum_{i=-18}^{-11}\text{new cases}_i}$$

References

1. Global economic effects of covid-19 (2020). https://fas.org/sgp/crs/row/R46270.pdf
2. Data on covid-19 coronavirus pandemic (2020). https://coronavirus.jhu.edu/map.html
3. Baldwin R (2020). https://voxeu.org/article/covid-remobilisation-and-stringency-possibility-corridor
4. WHO warns that coronavirus cases have jumped in countries that eased lockdowns (2020). https://www.cnbc.com/2020/05/11/who-warns-that-coronavirus-cases-have-jumped-in-countries-that-eased-lockdowns.html
5. Whose coronavirus strategy worked best? Scientists hunt most effective policies (2020). https://www.nature.com/articles/d41586-020-01248-1
6. Visualizing the impact of SARS-CoV-2 intervention strategies (2020). https://covidvis.berkeley.edu/
7. Oxford covid-19 government response tracker (2020). https://covidtracker.bsg.ox.ac.uk/
8. Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, Priesemann V (2020) Inferring covid-19 spreading rates and potential change points for case number forecasts. https://arxiv.org/abs/2004.01105
9. Temporal variation in transmission during the covid-19 outbreak (2020). https://epiforecasts.io/covid/