

A Logistic Growth Model with Logistically Varying Carrying Capacity for Covid-19 Deaths Using Data from Ontario, Canada



Georges Bucyibaruta, C. B. Dean, and E. M. Renouf

1 Introduction

Covid-19 infections have increased rapidly worldwide for the last several months. By mid-June, 2020, worldwide confirmed cases of Covid-19 were setting records for daily growth. This growth has continued into late July, with newly reported infections surpassing one quarter of a million daily, driven largely by new infections in the United States, Brazil, India, and South Africa. By July 30, 2020, worldwide recorded infections have surpassed 17 million. Canada announced its first case of Covid-19 in Toronto, Ontario on January 25, 2020 in an international traveller from Wuhan. By July 2020, Canada surpassed 115,000 cases of recorded infections, and almost nine thousand deaths. Within Canada, most cases have occurred within the provinces of Ontario and Quebec. Both provinces adopted more stringent public health measures to reduce infections after major issues arose within long term care facilities. Within Ontario, the number of cases grew rapidly by the end of March, at more than 400 daily new cases, reaching a peak of over 600 daily new cases in the third week of April. By the third week of May, Ontario saw a decline in daily deaths. The Ontario provincial database available through the provincial data portal records onset dates as early as January 01, 2020 in both travellers and non-travellers. While no death dates are recorded in this publicly available database, the earliest deaths recorded show 17 cases with onset dates on or before March 10, 2020 over a geographically dispersed set of health units (Windsor, Chatham, London, Niagara, Waterloo, Peel, Haliburton, Simcoe, Durham, Toronto, York and more).

While Ontario may be past the first peak of the epidemic, the lifting of public health restrictions and social distancing measures in a series of three planned stages

G. Bucyibaruta · C. B. Dean (✉) · E. M. Renouf
University of Waterloo, Waterloo, ON, Canada
e-mail: gbucyiba@uwaterloo.ca; vpri@uwaterloo.ca

may result in a rise in cases and hospitalizations due to Covid-19. Approximately 2 months after the initial lockdown in March 2020, the province of Ontario unveiled its three stage re-opening plan. Stage 1 of re-opening began on May 19, 2020, with a limited set of businesses allowed to re-open in accordance with strict public health guidelines. Two weeks after the first stage of re-opening, Ontario cases remained stubbornly flat with new cases persisting between 300 and 450 daily. Approximately two-thirds of the health regions in Ontario were allowed to progress to Stage 2 re-opening on June 12, 2020, where businesses such as hair and nail salons, restaurants with outdoor patios, campgrounds and community pools were allowed to re-open. At that time, new recorded cases had fallen to just below 200 per day in Ontario. All remaining health regions progressed to Stage 2 by July 7. On July 17, 2020, 24 of Ontario's 34 health regions, who were among the first health units to progress to Stage 2, were allowed to progress to Stage 3 of re-opening, which allows most businesses and workplaces to re-open but with limits on capacity and with measures in place such as wearing of masks indoors. By the final week of July, reported new infections hovered between approximately 100 and 170 per day, with some of the higher numbers driven by large outbreaks in agricultural farm workers. Meanwhile, health officials across the province remain alert to warning signs of resurgence.

Our approach here uses a logistic growth model for the cumulative number of deaths from Covid-19. The logistic growth model includes a carrying capacity parameter that is meant to reflect an upper limit in the number of deaths. Here we consider, conceptually, that deaths would be limited by the number of (true) cases. Since the number of cases is changing over time, so should the carrying capacity parameter for death in the logistic model. The logistic growth model we employ therefore allows the carrying capacity parameter to change over time by incorporating a logistic growth function for the carrying capacity parameter.

Throughout the course of the Covid-19 pandemic, Ontario and Canadian data have been modelled in a series of important papers using a broad set of epidemiological methods [4, 13, 14]. One of the first papers using Canadian data fit an exponential curve to the number of daily cases and estimated growth rates by fitting a linear regression model to the logarithm of the data [11]. Growth rates were estimated for two separate windows of time, to allow for changes in trend as a result of a significant public health intervention.

In terms of analysis of Ontario data, simple exponential or logistic growth models have been considered in a report published by a resource management group in collaboration with researchers at academic institutions across Canada. Growth rate curves of time series data of Covid-19 deaths for Canada are modeled to incorporate changes in rates under different public health interventions [5]. From the time series of daily deaths, the authors back-calculate the cumulative number of infections. Using a time series growth model the number of future infections are predicted. The model utilizes either a simple exponential growth model or a logistic growth model. In a report published May 17, 2020, the authors demonstrate how to use their modeling framework and provide software to generate forecasts of cases and deaths for Ontario and Canada under the assumption of continued health policy measures as they currently stand.

There have been other studies of Ontario data beyond growth curve modeling. For example, in a recently published article, Wu and colleagues [15] use a generalized Susceptible Exposed Infected Recovered (SEIR) model that allows incorporation of asymptomatic infectious, quarantined susceptible and isolated exposed model compartments. They evaluated trends in transmission and the effect of social distancing measures based on data up to March 29, 2020, showing an increasing effectiveness of public health interventions in lowering the reproductive number of Covid-19.

Analyses utilizing logistic growth curve models to forecast epidemic growth of infections have been published using data outside of Canada as well. These models use the standard form of the logistic growth curve to model infections. To model early growth of the epidemic in Hubei, China, [10] applied the generalized logistic growth model to produce short-term forecasts of cases using data up to February 29, 2020. [16] also use a generalized logistic growth model to forecast cases in mainland China excluding Hubei province. They compare the classical logistic growth model, a generalized model, and a generalized Richards model using data up to March 10, 2020. [3] uses a logistic growth model as well as an SEIR model to estimate final epidemic size worldwide.

A recent paper using publicly available data from Nigeria applied a logistic growth curve as one element of a larger model forecasting epidemic growth [1]. Using the daily number of new cases of Covid-19 in Nigeria, the authors implemented an ensemble of forecast models. One of these models included as a component a logistic growth model with time-varying carrying capacity. This implementation allows the carrying capacity to vary as a function of time, rather than a logistically varying carrying capacity as in our model.

The logistic growth model has been used extensively, either in traditional or extended form, to model new infections for Covid-19. A model accommodating a varying carrying capacity parameter as considered in this paper has not been used in logistic growth models for Covid-19. In our application which follows, we demonstrate the utility of the logistic growth model to model deaths, with the modification that the parameter identified with carrying capacity varies over time. We will also provide confidence bands for potential indicators through Monte Carlo methods.

2 Data Description

The aggregate data used here were obtained from the daily epidemiological summaries released by Public Health Ontario on their website. In a table under the heading ‘Severity’, these daily reports show the reported cumulative deaths and the daily change from the previous report, with the caveats that only deaths for lab-confirmed cases are included here, and also that there is a reporting delay for deaths. Figure 1 shows deaths in Ontario as reported by Public Health Ontario.

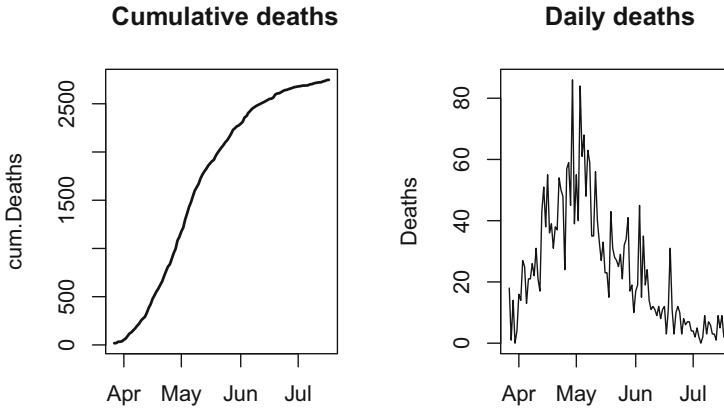


Fig. 1 Cumulative and daily number of deaths in Ontario from 27/03/2020 until 17/07/2020 (data from Public Health Ontario daily epidemiological summaries)

3 Methods

Our focus is on modeling deaths in the province of Ontario. This approach uses a growth model for the cumulative number of deaths from Covid-19, and our growth model accommodates, conceptually, that deaths have a carrying capacity that would be limited by, for example, the number of cases or hospitalizations which change over time. This is not a process model, but an empirical model. A series of publications, [2, 8, 12], portray an array of logistically developing and diffusing social mechanisms. They compare technological innovations as a social epidemic by arguing that the former do not usually distribute themselves evenly through time. They consider a model where the carrying capacity of the system increases dynamically, but in a distinct pulse. Conceptually, we adapt this approach and we allow the carrying capacity of our logistic growth model to vary as a logistic growth curve.

3.1 Logistic Growth Model for the Mean

As the carrying capacity is meant to conceptually reflect the number of cases or hospitalizations, and since these values are changing over time, so should the carrying capacity for death in the logistic model. Let $N(t)$ be the cumulative number of deaths at time t , where $t = 0$ is the recorded date of the first death in Ontario. The logistic growth curve model can be represented in the following way:

$$\frac{dN(t)}{dt} = rN(t) \left[1 - \frac{N(t)}{K(t)} \right] \quad (1)$$

where

r is the growth rate and

$K(t)$ is the carrying capacity parameter for modeling $N(t)$ at time t ,

$N(t = 0) = N_0$ is the cumulative number of deaths at the initial time.

Time here is recorded as rounded to days, so N_0 is the number of deaths on the first day of recorded deaths. The general solution of (1) is

$$N(t) = \frac{N_0 \exp(rt)}{1 + rN_0 \int_0^t \frac{\exp(rx)}{K(x)} dx} \tag{2}$$

with $K(t)$ also modeled as a logistic growth curve:

$$\frac{dK(t)}{dt} = \alpha K(t) \left[1 - \frac{K(t)}{G} \right] \tag{3}$$

with α being the growth rate for $K(t)$, and G being the carrying capacity parameter for modeling $K(t)$. The analytical solution of $K(t)$ follows as

$$K(t) = \frac{G}{1 + \left(\frac{G}{G_0} - 1\right) \exp(-\alpha t)} \tag{4}$$

where G_0 is the initial value of the carrying capacity for modeling $K(t)$. Substituting (4) in (2) yields the solution:

$$N(t) = \frac{G}{1 + (A_1 \exp(-\alpha t)) + (A_2 \exp(-rt))} \tag{5}$$

where

$$A_1 = \left(\frac{G}{G_0} - 1\right) \left(\frac{r}{r - \alpha}\right) \tag{6}$$

and

$$A_2 = \left(\frac{G}{N_0} - 1\right) - \left(\frac{G}{G_0} - 1\right) \left(\frac{r}{r - \alpha}\right). \tag{7}$$

3.2 Non-linear Least Squares Estimation

We employ non-linear least squares estimation for the parameters. The function $N(t)$ is known up to a set of $p = 4$ unknown parameters $\theta = (\theta_1, \dots, \theta_p) = (G, G_0, r, \alpha)$ which also must be estimated. Under the assumption that both the

predictor and the response are observed without error, the relationship in (5) will hold to define cumulative counts of deaths over time if the model is correct.

In reality, measurement errors will arise. Non-linear least squares estimation proceeds by finding $\hat{\theta}$ that minimizes

$$RSS(\theta) = \sum_{i=1}^n (y_i - N(t_i))^2 \quad (8)$$

where $y_i = N(t_i) + \epsilon_i$, and $\epsilon_i \sim N(0, \sigma^2)$.

An estimate of the measurement error is obtained as

$$\hat{\sigma}^2 = \frac{RSS(\hat{\theta})}{n}$$

where $RSS(\hat{\theta})$ is the residual sum of squares.

4 Introducing Stochasticity into the Daily Counts

The logistic growth model provides an estimate of the future mean cumulative deaths. We also consider stochasticity in the daily counts, required for short-term analyses of the behaviour of the disease progression and assume that the daily number of deaths follow a negative binomial distribution with mean derived from the fitted values of the logistic growth model. We estimate the dispersion parameter $\hat{\kappa}$ using maximum likelihood estimation and this allows us to incorporate stochasticity in the daily counts. We then utilize a Monte Carlo approach for obtaining future daily predictions by generating future daily data using a negative binomial distribution with mean derived from the fitted logistic model for cumulative deaths and dispersion $\hat{\kappa}$. Using $B = 1000$ simulations, under the negative binomial distribution, we predict future cumulative deaths. In a single peak epidemic wave, an indicator of lack of control could be based on the cumulative number of deaths $N(t)$, and the rate of change of deaths $\frac{dN(t)}{dt}$, as described in the next section.

5 Short-Term Predictions and Beyond

We discuss here potential tools that could be utilized as short-term predictors, as applicable more broadly for pandemic monitoring in various settings. For example, we may calculate the probability that the total number of deaths observed in the next l days, after a reference point t_0 , indicating current time, will exceed that observed in the past l days, where l monitors short-term activity, for example, $l = 3$ or $l = 5$. We also examine the probability that the current growth rate of deaths exceeds that

seen during the beginning of the first phase of re-opening when the pandemic was seen to be sufficiently under control.

Looking beyond the short-term predictors, a few indicators of a second wave, or resurgence, have been proposed however most are ad hoc apart from the risk ratio recently proposed by Noorbhai [9], who offers a model based on the ratio of total recoveries to cases. This is problematic for Ontario where recoveries are tracked with different methodologies within each of the 34 health units. Freitag et al. [6] has proposed a spatio-temporal model of mobility levels, taking into account population density, as an indicator of resurgence. An immediate issue for Ontario, however, is that indicators for resurgence cannot be based upon growth models, such as the one used here, which are meant for modeling an epidemic with a single peak.

6 Results

The logistic growth model with a logistically varying carrying capacity parameter was fitted to cumulative deaths. Figure 2 shows the results of the analysis with both fitted and observed values displayed. The time-varying carrying capacity curve is shown in blue. The plot illustrates how the logistic curve approaches its upper limit over time. As the logistic growth model reaches the asymptote, the difference between the carrying capacity and the logistic growth curve diminishes. Figure 3 displays the confidence bands associated with the fit are also provided. The model

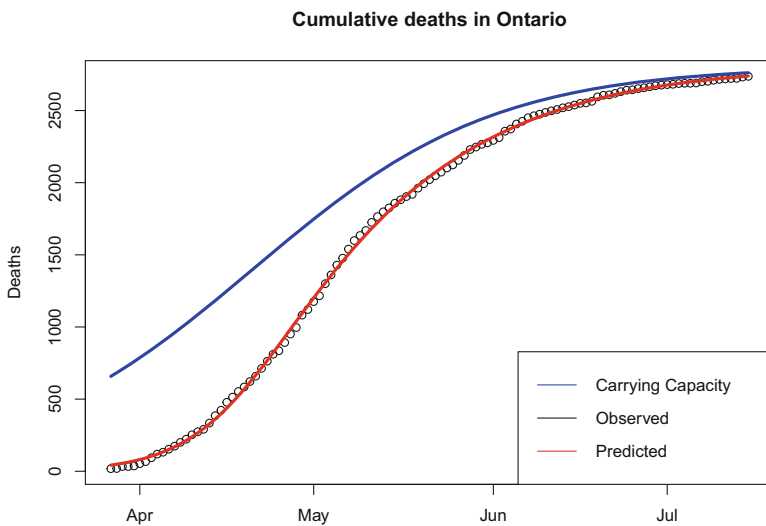


Fig. 2 Fitted logistic growth model with logistically varying carrying capacity from 27/03/2020 to 17/07/2020

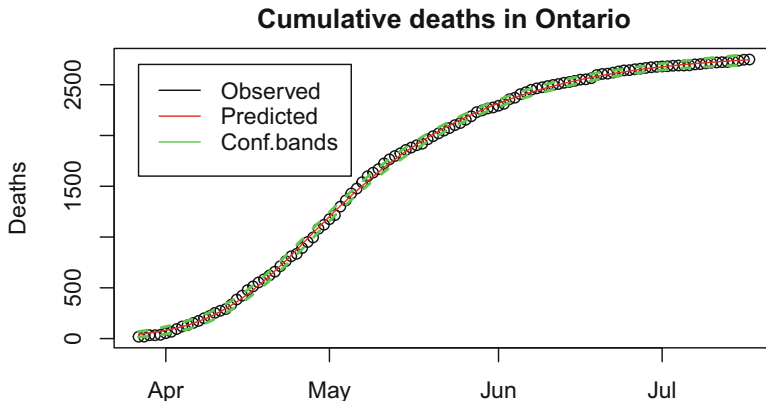


Fig. 3 Fitted Logistic growth model with logistically varying carrying capacity from 27/03/2020 to 17/07/2020, showing confidence bands in green

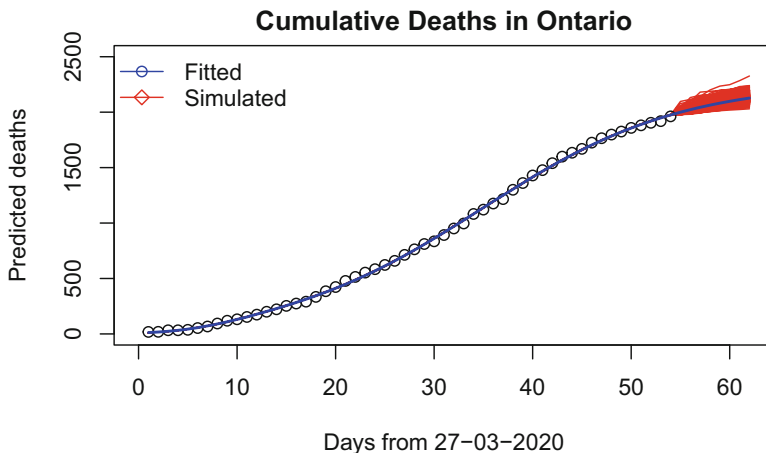


Fig. 4 Total number of Deaths predicted under logistic growth curve (blue) with 95% confidence bands (green) and simulated curves under negative binomial distribution (red)

show no obvious lack of fit. Figure 4 displays the logistic growth curve with 95% confidence bands and simulated curves under negative binomial distribution.

In order to assess our model we present some comparisons between the give day ahead forecasted number of deaths for various dates versus what was actually observed, in Table 1.

To examine the future trajectory of predicted deaths, we consider the probability that the total number of deaths observed in the next l days after day $t_0 = 54$ in our dataset (19 May, 2020), exceeds that observed in the last l days, where $l = 3$

Table 1 Observed cumulative deaths at time t_0 (second column), predicted cumulative deaths after 5 days (3rd column) and observed cumulative deaths after 5 days (last column)

Date (t_0)	$N(t_0)$	Predicted $N_{(t_0+5)}$	Observed $N_{(t_0+5)}$
2020-05-10	1669	1952	1858
2020-05-19	1962	2081	2012
2020-06-04	2372	2447	2475
2020-06-12	2507	2599	2553
2020-07-04	2689	2731	2710
2020-07-17	2748	2758	2755

or 5. Using the Monte Carlo approach described in Sect. 4, these probabilities are estimated as,

$$\hat{P}(N(t_0 + 3) > N(t_0) + \{N(t_0) - N(t_0 - 3)\}) = .29$$

$$\hat{P}(N(t_0 + 5) > N(t_0) + \{N(t_0) - N(t_0 - 5)\}) = .14$$

It is useful to note that the probability declines as l increases. In reality, starting at the reference time, the number of deaths in the next l days slightly exceeded the number of deaths observed during the previous l days, for both $l = 3$ and 5, by 5 deaths and 3 deaths respectively. As another example, when $t_0 = 84$, corresponding to 18 June, 2020, we have,

$$\hat{P}(N(t_0 + 5) > N(t_0) + \{N(t_0) - N(t_0 - 5)\}) = 0.72$$

Starting at the reference time, the number of deaths in the subsequent 5 days exceeded the number of deaths observed during the previous 5 days by 22 deaths. The probability values themselves give some indication of the strength of evidence concerning the prediction, yet a threshold is required to form an alarm system. This could be developed through a receiver operating characteristic curve analysis.

Figure 5 shows the receiver operating characteristic curve for the model. This curve was developed through a Monte Carlo simulation, similar to the method described in Sect. 4, where we calculated the probability of exceeding the l -day-ahead target for deaths, the target being the total number of deaths in the previous l days, for different t_0 's. Then we assessed whether the observed number of deaths actually exceeded the target: for each t_0 an *outcome* of 1 was assigned if the number of deaths in the next l days exceeded the number of deaths observed during the previous l days, otherwise an outcome of 0 was assigned. By comparing the estimated probabilities of exceeding the l -day-ahead target to a series of thresholds between 0 and 1, we obtain a *prediction* of whether the number of deaths l days ahead will exceed the target. For each of these probability thresholds, we are then able to compare *outcomes* with *predictions* to calculate the true positive rate (TPR) and false positive rate (FPR) associated with each of these thresholds.

On the ROC curve, the point at the top left corner of the curve is identified as providing the level of best performance. This occurs at a TPR of 0.9 and FPR of 0.3, as shown in Fig. 5. This, in turn, corresponds to a threshold probability of 0.6. At

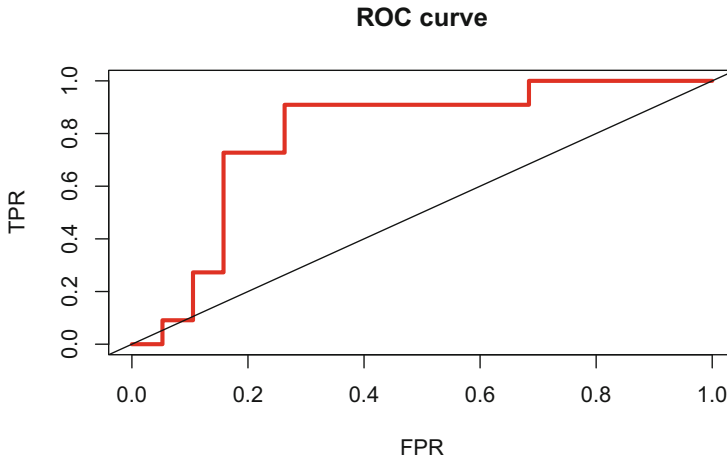


Fig. 5 Optimal threshold for predicting an increase in deaths over the next l days, compared to the previous l days, where $l = 5$

this point, we therefore expect 90% of true positives will be well classified, and 30% of false positives will be misclassified. Therefore, when considering predictions of deaths 5 days ahead, we would sound an alarm indicating an expected increase in deaths when the probability described above is greater than the threshold of 0.6. A probability of 0.72, as obtained in our most recent example, would be considered high enough to sound an alarm, according to the threshold determined by means of our ROC curve analysis.

We can also develop an assessment of risk level given by $dN(t)$, the rate of change. Since $dN(t)$ is the derivative of $N(t)$, these indicators provide the same probability measure, however $dN(t)$ gives us the ability to compare to earlier phases of the pandemic, with a straightforward visual representation. Figure 6 gives a visual representation of the values of $dN(t)$ over time since late March, 2020.

We present in Table 2 a scale of low, medium and high risk values for the predicted rate of change as measured at time t with the intention being that rates of change should be decreasing over time, and where they are not, could support concerns about resurgence.

If we use the date of June 18, 2020 (day 84 in our dataset) given in our previous example, we calculate the predicted average growth rate over the next 5 days to be 11.4. We can compare this value to the start of Phase 2 on June 12, shown in Fig. 6, where the instantaneous growth rate was 13, and we see therefore, that growth rate is declining.

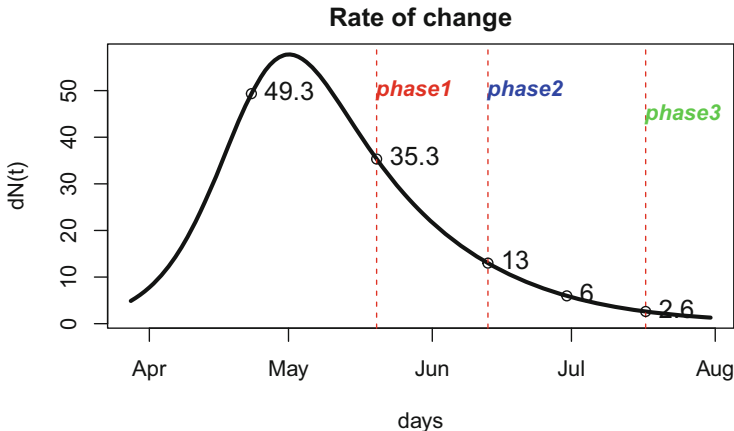


Fig. 6 An indicator based on $dN(t)$, the rate of change of deaths

Table 2 Risk levels based on the rate of change, $dN(t)$

$dN(t)$	Risk level
>35	Very high
13–34	High
7–12	Medium
2–6	Low
<2	Very low

7 Discussion

In this paper we have used a simple model to predict future deaths in the short term. Our model fits reasonably well, with narrow confidence bands. We consider an extension of the classic logistic growth model which allows the carrying capacity for deaths to change logistically over time. This model does not take into account the mechanisms of transmission of Covid-19, such as health interventions and human behavior. As the situation evolves, anomalous values or rapidly changing trends could upend any prediction efforts. Worse still, sudden shocks that permanently affect a time series could also render all past data as irrelevant. We also note that phenomenological growth models such as the logistic growth curve model presented here are meant for predicting growth trajectories during a single peak epidemic. However, multiple peak epidemic trajectories caused by factors such as increasing contacts and releasing of public health interventions are much more challenging to model.

Our study has some important limitations to acknowledge with respect to the data used in our model. This pandemic in particular has highlighted challenges in data collection and management in Ontario. The deaths, as reported, do have a lag from their actual death date as can be gleaned from graphs provided in the Daily Epidemiological Summary published by Public Health Ontario. While deaths are

a late indicator, they are still a valuable marker to track as part of a composite surveillance plan. In addition to monitoring the growth rate of deaths as described in this paper, policy makers would consider a broad context, examining other metrics such as hospitalization numbers, test percent positivity, current reproduction number, and number of new cases per 100,000 population. These indicators for the future number of deaths which we present here are meant as an additional layer of insight to combine with other important metrics.

In future work we will explore modeling the carrying capacity parameter as a function of hospitalizations. Hospitalizations may indeed provide conceptually a better upper limit for modeling deaths as many cases may be mild whereas hospitalizations would be strongly linked to the more severe cases that progress to death. In future work we hope to address the reporting lag for deaths and to incorporate a method to adjust for these lags in our predictions based on hospitalization rates.

Provincial analyses could be supplemented by regional analyses in order to detect regional trends. As noted in [7], using aggregate level data as an indicator for all of Ontario can obscure what might be happening at the local level. The indicators in this paper could easily be extended for use at the regional level and we intend to model this in future work.

At the time of completing this study, Ontario has allowed all health regions to progress to Stage 3 of re-opening, where bars and restaurants with indoor seating have re-opened, as well as gyms, personal care services such as hair salons, and also places of worship. Most remaining workplaces and businesses are allowed to re-open with some precautions in place. On September 8, 2020, schools in the province will re-open with some additional public health measures such as mandatory masks for grade 4 and up, but for the most part will proceed as usual except in 24 out of 76 school boards, where secondary school class sizes will be limited to 15 students who will attend on a rotating schedule. Physical distancing requirements are set at only 1 m by the Ministry of Education, even in classes where masks are not mandatory.

This return to school will result in increased contact for the population of Ontario, and in conjunction with the upcoming influenza season, the fall season may be a period of increased risk for Covid-19. Formal indicators of public health interventions may be useful for managing risk, and we intend to investigate extensions of this type of model that will allow modeling of future epidemic peaks in Ontario. In this paper we presented a model to predict future deaths in the short-term with an appropriate measure of uncertainty.

8 Declaration of Competing Interest

We have no conflict of interest.

Acknowledgments The authors are supported in part by a CIHR grant for the Canadian 2019 Novel Coronavirus (COVID-19) Rapid Research Funding Opportunity.

References

1. Abdulmajeed K, Adeleke M, Popoola L (2020) Online forecasting of Covid-19 cases in Nigeria using limited data. Data in Brief p 105683
2. Banks R (2013) Growth and diffusion phenomena: Mathematical frameworks and applications. Springer Science and Business Media 14
3. Batista M (2020) Estimation of the final size of the Covid-19 epidemic. MedRxiv <https://doi.org/10.1101/2020.02.16.20023606>
4. Bignami S, Ghio D, Van Assche A (2020) Estimates of Covid-19 case-fatality risk from individual-level data. medRxiv
5. Daniel C, Frid L, Rayfield B, Embrey A (2020) A scenario-based forecasting framework for Covid-19: Case study application for Canada <https://apexrms.com/wp-content/uploads/download3iles/covid-19-canada-framework-whitepaper.pdf>
6. Freitag MO, Schmude J, Siebensschuh C, Stolovitzky G, Hamann H, Lu S (2020) Critical mobility, a practical criterion and early indicator for regional Covid-19 resurgence. medRxiv
7. Harris JE (2020) Reopening under Covid-19: What to watch for. Tech. rep., National Bureau of Economic Research
8. Meyer PS, Ausubel JH (1999) Technological forecasting and social change. Technological forecasting and social change 61(3)
9. Noorbhai H (2020) A mathematical model to guide the re-opening of economies during the Covid-19 pandemic. Annals of Medicine and Surgery 57:5
10. Roosa K, Lee Y, Luo R, Kirpich A, Rothenberg R, Hyman J, Yan P, Chowell G (2020) Real-time forecasts of the Covid-19 epidemic in China from February 5th to February 24th, 2020. Infectious Disease Modelling 5:256–263
11. Scarabel F, Pellis L, Bragazzi NL, Wu J (2020) Canada needs to rapidly escalate public health interventions for its Covid-19 mitigation strategies. Infectious Disease Modelling 5:316–322
12. Tsoularis A, Wallace J (2002) Analysis of logistic growth models. Mathematical Biosciences 179(1):21–55
13. Tuite AR, Fisman DN, Greer AL (2020) Mathematical modelling of Covid-19 transmission and mitigation strategies in the population of Ontario, Canada. CMAJ 192(19):E497–E505
14. Tuite AR, Greer AL, De Keninck S, Fisman DN (2020) Risk for Covid-19 resurgence related to duration and effectiveness of physical distancing in Ontario, Canada. Annals of Internal Medicine 173(8):675–678
15. Wu J, Tang B, Bragazzi NL, Nah K, McCarthy Z (2020) Quantifying the role of social distancing, personal protection and case detection in mitigating Covid-19 outbreak in Ontario, Canada. Journal of mathematics in industry, 10, 1–12.
16. Wu K, Darcet D, Wang Q, Sornette D (2020) Generalized logistic growth modeling of the Covid-19 outbreak in 29 provinces in China and in the rest of the world. arXiv preprint arXiv:200305681