Estimating the Growth Rate and Doubling Time for Short-Term Prediction and Monitoring Trend During the COVID-19 Pandemic with a SAS Macro

Stanley Xu[1,2], Christina Clarke[1], Susan Shetterly[1], Komal Narwaney[1]

[1] The Institute for Health Research, Kaiser Permanente Colorado, 2550 S. Parker Road, Aurora, CO 80014

[2] School of Public Health, University of Colorado, 13001 E 17th Pl, Aurora, CO 80045

ABSTRACT Coronavirus disease (COVID-19) has spread around the world causing tremendous stress to the US health care system. Knowing the trend of the COVID-19 pandemic is critical for the federal and local governments and health care system to prepare plans. Our aim was to develop an approach and create a SAS macro to estimate the growth rate and doubling time in days if growth rate is positive or half time in days if growth rate is negative. We fit a series of growth curves using a rolling approach. This approach was applied to the hospitalization data of Colorado State during March 13th and April 13th. The growth rate was 0.18 (95% CI=(0.11, 0.24)) and the doubling time was 5 days (95% CI= (4, 7)) for the period of March 13th-March 19th; the growth rate reached to the minimum -0.19 (95% CI= (-0.29, -0.10)) and the half time was 4 days (95% CI= (2, 6)) for the period of April 2nd – April 8th. This approach can be used for regional short-term prediction and monitoring the regional trend of the COVID-19 pandemic.

1. BACKGROUND

In December 2019, an outbreak of coronavirus disease (COVID-19) caused by the novel coronavirus (SARS-CoV-2) began in Wuhan, China and has now spread across the world [1,2]. In the United States, the cumulative number of identified COVID-19 cases was 186,101 as of March 31st, 2020; among the identified cases, 3603 died [3]. To slow the spread of COVID-19, federal and local governments have issued mitigation measures such as case isolation, quarantine, school closures and closing non-essential businesses. The COVID-19 pandemic imposes tremendous challenges to the US health care system, particularly given concerns that the need for hospital beds and ICU beds could exceed capacity [4-6]. Predicting the future numbers of COVID-19 cases and healthcare utilization is critical for governments and health care systems preparation plans [4,6,7]. Two useful and critical quantities for prediction are the growth rate [8] and the doubling time of number of events [9]. The growth rate is the percent change of daily events (e.g, COVID-19 cases, number of patients hospitalized or number of deaths). The doubling time is the length of time required to double the number of daily events.

Our goal was to develop an approach and create a SAS macro using observed data to estimate the growth rate and doubling time in days for short-term prediction.

2. METHODS

2.1 A rolling growth curve approach (RGCA)

In the United States, there are several barriers for testing people for COVID-19 such as shortages of swabs and testing kits and restrictions on who should get tested. Therefore, the number of COVID-19 cases is often under-identified and under-reported. However, the number of hospitalized COVID-19 patients (hospitalizations) and number of deaths due to COVID-19 are more reliable than the reported number of COVID-19 cases [10]. In this paper, we used the number of daily hospitalized COVID-19 patients to calculate the growth rate and doubling time in days.

We assumed a growth curve of daily hospitalizations over a period of n days from day t (start day) to day $(t+n-1)$. Let $y_{(t+j-1)}$ denote the daily hospitalizations at day $(t+j-1)$, $1 \leq j \leq n$. Based on the growth model, we have

$$y_{(t+j-1)} = y_t(1+r)^{j-1} \qquad (1)$$

where y_t is the number of hospitalizations at the start day t; r is the growth rate. When the growth rate $r > 0$, the number of daily hospitalizations increases. For example, if r =0.4, the growth rate of hospitalizations is 40% more for each day. When growth rate $r = 0$, the number of daily hospitalizations has no change. When growth rate $r < 0$, the number of daily hospitalizations declines. When the number of hospitalizations doubles at $j = D$, that is $y_{(t+D-1)} = 2y_t$, we have,

$$2y_t = y_t(1+r)^{D-1}$$

Further, it can be shown that

$$r = e^{\left(\frac{\ln(2)}{D-1}\right)} - 1 \qquad (2)$$

We fit two models: a) using equation (1) which estimates the growth rate r; b) using equation (1) with r substituted with $e^{\left(\frac{\ln(2)}{D-1}\right)} - 1$ from equation (2). The second model estimates the doubling time in days D, meaning that it takes D days from the start day t for the number of daily hospitalizations to double. We used SAS PROC NLIN [11] to fit these two nonlinear models. Note that equation (2) is valid for $r > 0$. When $r < 0$, one can use $r = e^{\left(\frac{\ln(0.5)}{D-1}\right)} - 1$; the estimated D represents the days required to reduce the number of hospitalizations by half (half time).

Because the growth rate and doubling time may change over time, we used a rolling growth curve approach (RGCA). For example, we set the length of the period to be 7 days ($\mathrm{n} = 7$ days). We estimated the growth rate and the doubling time in days for the following periods for hospitalization data from Colorado State from March 13th – April 13th [12]: March 13th-19th,14th-20th, 15th-21st,…, April 7th-April 13th.

2.2 Short-term prediction

The estimated growth rate from the last period of the RGCA approach (e.g., April 7th-April 13th) can be used for future short-term prediction of hospitalizations. Let k denote the last day of the last period, y_k is the number of hospitalizations on this day. For the Colorado hospitalization data in this analysis, k is April 13th, $y_k = 36$. Let m denote the date after date k, then the predicted y_m is

$$\bar{y}_m = y_k(1 + \hat{r}_k)^{m-k}$$

where $\hat{r}_k$ is the estimated growth rate from the last period. As the growth rate changes over time, the prediction is only appropriate for short-term prediction (e.g., within 7 days) and updated growth rates should be used.

3. RESULTS

We estimated a series of growth rates using RGCA with a length of 7 days. The estimated growth rates and 95% CIs were plotted over time using the mid-day of a 7 day period (Figure 1). The growth rate peaked with a value of 56.2% at the mid-day of March 18th for the period March 15th and March 21st. Between March 18th and April 1st, although the growth rate continuously decreased, the daily number of hospitalizations increased because of positive growth rates. We started to observe negative growth rates after April 1st, except for a positive growth rate on April 9th. The growth rate reached its minimum at the mid-day of April 5th (period April 2nd-April 8th) with a value of -19.2%. The growth rate then increased after April 5th. Note that a negative growth rate represents a reduction in number of hospitalizations.

The doubling time (growth rate>0) and half time (growth rate<0) in days over time are displayed in Figure 2. Before April 1st, the y-axis represents the doubling time in days because of positive growth rates. After April 1st, except for a positive growth rate on April 9th, the y-axis represents a half-time because of negative growth rates. On April 1st, the reduction rate was very small (0.5%) which resulted in a high half-time, 128 days with very wide 95% CIs (not shown in the figure). On April 8th, there was a small reduction rate (1.2%) resulting in 55 days of half time.

Using the estimated growth rate from the last period April 7th-April 13th, $\hat{r}_k = 0.027$, the predicted numbers of daily hospitalizations for April 14th and 15th were 35 and 34, respectively. SAS programs are available for conducting these analyses (Appendix A and Appendix B).

4. DISCUSSION

These models can be similarly applied to death data if they are available and not sparse. When COVID-19 testing is widely available to the public and the number of COVID-19 testing is less selective, these models can also be used to directly estimate the growth rate and the doubling time for COVID-19 cases. Due to a lag in reporting hospitalization, it is recommended to exclude the recent 1-2 days' hospitalization data in fitting the growth curves. This paper illustrates that hospitalization data can be used to estimate the growth rate and doubling (or half) time to aid predicting future hospitalizations, deaths and COVID-19 cases. Because a series of growth curves were fit, the RGCA approach can also be used for real-time monitoring of the epidemic trend as shown in Figure 1.

Colorado state issued three social distancing guidelines: a voluntary social distancing on March 17th, closing educational facilities on March 23rd, and closing non-essential services on March 26th (Figure 1). It takes some time (e.g., 2 weeks) for these mitigation measures to have impact. Although the effectiveness of these mitigation measures has not been investigated formally, it is believed that they helped to slow the spread of COVID-19 and reduced number of hospitalizations and death in Colorado and across the United States.

Acknowledgements

This research was supported by the Institute for Health Research, Kaiser Permanente Colorado. Xu was also supported by NIH/NCRR Colorado CTSI Grant Number UL1 RR025780.

REFERENCES

1. CDC. 2019 Novel Coronavirus, Wuhan, China. Available at https://www.cdc.gov/coronavirus/2019-ncov/
2. WHO. WHO Director-General's opening remarks at the media briefing on COVID-19 – 11 March 2020
3. CDC. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html
4. IHME COVID-19 health service utilization forecasting team. Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. MedRxiv. 26 March 2020. doi:10.1101/2020.03.27.20043752.
5. Ferguson NM, Laydon D, Nedjati-Gilani G, et al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand. Imp Coll COVID-19 Response Team. March 2020:20. doi:https://doi.org/10.25561/77482
6. Tsai TC, Jacobson B, Jha AK. American hospital capacity and projected need for COVID-19 patient care. Health Aff (Millwood). March 2020. doi:10.1377/hblog20200317.457910

7. Petropoulos F, Makridakis S (2020) Forecasting the novel coronavirus COVID-19. PLoS ONE 15(3): e0231236. https://doi.org/10.1371/journal.pone.0231236
8. Du Z, Xu X, Wu Y, Wang L, Cowling BJ, Ancel Meyers L. Serial interval of COVID-19 among publicly reported confirmed cases. Emerg Infect Dis. 2020 Jun [date cited]. https://doi.org/10.3201/eid2606.200357
9. Nunes-Vaz, R., 2020. Visualising the doubling time of COVID-19 allows comparison of the success of containment measures. Global Biosecurity, 1(3), p.None. DOI: http://doi.org/10.31646/gbio.61
10. https://covid19.healthdata.org/projections
11. SAS Institute, version 9.4, Cary, NC
12. https://covid19.colorado.gov/case-data (accessed on April 14th 2020)

Figure 1. Estimated growth rate with 95% CIs over time using hospitalization data from Colorado State.

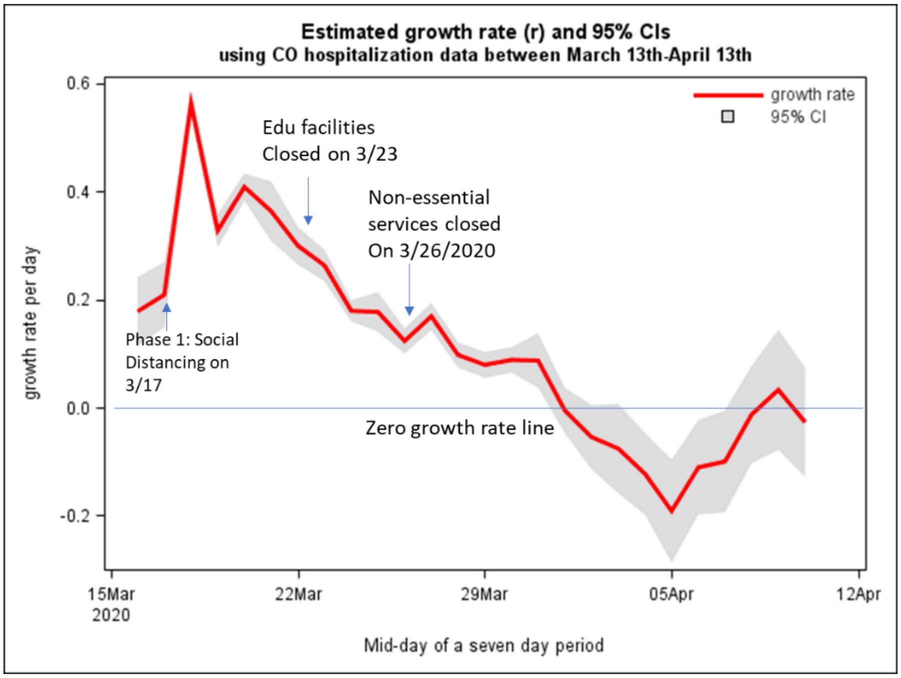
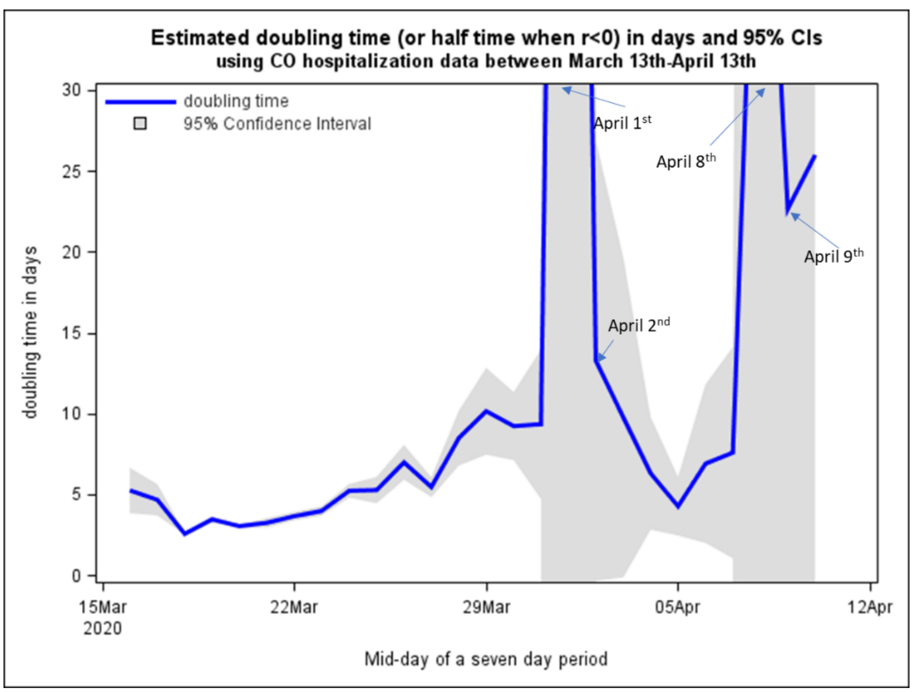


Figure 2. Estimated doubling (or half) time in days with 95% CIs over time using hospitalization data from Colorado State.



APPENDIX A

```
/************************************************************
*  Title:
*  Programmer: Christina Clarke
*              Stanley Xu
*              Institute for Health Research
*              Kaiser Permanente Colorado
*              303-614-1252
*  Date Created: 4/3/2020
*  Description:
*
*  Input:
*
*  Output:
*
*  References:
*
*  Changes to the code log :
*  Date       Programmer          Description
*------------------------------------------------------------
*
*************************************************************/


%let codeanddat =C:\Xu\covid19\macro_paper;

%INCLUDE "&codeanddat.\COVID_growthrate.sas";

/*import the death file*/
PROC IMPORT OUT = events
    DATAFILE = "&codeanddat.\Cumulative_CO_HOSP_4_13_2020.xlsx"
    DBMS = XLSX REPLACE;
RUN;

proc print data=events;
run;


/*Clean up the datasets */
proc datasets library=work nolist;
      delete cgr_dat01
             cgr_dat02
             cgr_dat03
               doubling_time
               doubling_time1
               preddeath:
               r_doubling_time
               r_doubling_time0
               r_estimate1
               r_estimates
               prediction;
run; quit;
%Calc_GrowthRates(indat=events
                        , dateofevent = date
                        , numevents = dailyHosp
                        , int_length = 7);
```

```
title 'Growth rate and doubling time for each interval';
proc print data = r_doubling_time;
      var start_day
          end_day
            mid_day
            r r_lowercl r_uppercl
            doubling_time
            doubling_time_lowercl
            doubling_time_uppercl;

      format start_day end_day mid_day mmddyy10.;
run;


ods listing gpath="&codeanddat";

title1 'Estimated growth rate (r) and 95% Confidence Intervals';
title2 'Using CO Hospitalization data between March 13th-April 13th';

proc sgplot data=r_doubling_time;
      band x = mid_day lower = r_LowerCL upper = r_UpperCL/
            fillattrs = (color = graydd) name='band95' legendlabel='95% CI';
      series x = mid_day y = r/
            lineattrs = (color = red thickness = 3px) name = 'line'
legendlabel = 'growth rate';

      yaxis labelpos = center label = "growth rate per day" thresholdmax=.8;
      xaxis labelpos = center label = "Mid day of a seven day period"
thresholdmax=.8;

      keylegend 'line' 'band95' /
            position = topright location = inside across = 1 opaque noborder
valueattrs=(color = gray33);
run;



title 'Estimated doubling time (or half time when r<0) in days (D) and 95%
CIs';
title2 'Using CO Hospitalization data between March 13th-April 13th';
proc sgplot data=r_doubling_time;
      band x = mid_day lower = doubling_time_LowerCL upper =
doubling_time_UpperCL/
            fillattrs = (color = graydd) name = 'band95' legendlabel = '95%
Confidence Interval';
      series x = mid_day y = doubling_time/
            lineattrs = (color = blue thickness = 3px) name = 'line'
legendlabel = 'doubling time';

      yaxis labelpos = center label = "doubling time in days" thresholdmax =
.8 min=0 max=30;
      xaxis labelpos = center label = "Mid day of a seven day period"
thresholdmax = .8;

      keylegend 'line' 'band95' /
```

```
            position = topleft location = inside across = 1 opaque noborder
valueattrs = (color = gray33);
run;
```

APPENDIX B

```
/***************************************************************
*  Title:
*  Programmer: Stanley Xu and Christina Clarke
               Institute for Health Research
*              Kaiser Permanente Colorado
*
*  Date Created: 4/3/2020
*  Description: This macro is designed to calculate a predicted
*    growth and doubling time of a disease given observed
*    data. In particular, these models were based on observed
*    deaths since the true denominator is often unknown given
*    testing may not be done on all symptomatic or asymtomatic
*    individuals. Further, hospitalizations could be used if they
*    are known.
*
*  Input: indat = input dataset with the number of deaths and date of those
deaths during a date range
*                 that is to be modeled.
*           dateofevent = variable name of te date the deaths occurred from
the indat dataset
*         numevents = variable name that has the number of deaths that
occurred on each date of death
*                                from the indat dataset
*         int_length - number of days in each interval - our
*                     example examined 7 day intervals to create piece-wise
growth intervals
*
*
*  Output:
*
*  References:
*
*  Changes to the code log :
*  Date       Programmer         Description
*-------------------------------------------------------------
*  4/3/2020   cclarke                    CH001 remove the state variable
option
***************************************************************/


%macro Calc_GrowthRates(indat, dateofevent, numevents, int_length);
*First, we need to get the start and end dates from the input dataset.;
proc sql noprint;
      select distinct min(&dateofevent)
                           ,max(&dateofevent)
```

```
                   into :fup_start
                                    ,:fup_end
      from &indat
;
quit;

/*For QA - Prints the first and last date found in the input data file
 which will appear in the log*/
data _null_;
         start  = &fup_start;
         endloop = &fup_end-&int_length+1;

         format start endloop date9.;

         put "Looping through the starting date and the last date -
&int_length +1 days:";
         put start =
               endloop =;
run;

%do start_day = &fup_start %to (&fup_end-&int_length+1);

/*Hold onto the current start day to append to some of the final datasets
so they do not get overwritten*/
%let stdyfmt = %sysfunc(putn(&start_day,date9.));

/*Using the input dataset, calculate the last day that will be considered
  in these calculations for each interval.
  The date of death needs to be between the start and ending
 day*/
data CGR_dat01;
 set &indat;

 end_day = &start_day + &int_length - 1;
 if &start_day <= &dateofevent <= end_day;

 format end_day date9.;
proc sort;
      by &dateofevent;
run;

/*This step will retain the number of deaths from the first date of the
   current interval through each date deaths were reported.*/
data CGR_dat02;
 set CGR_dat01;
      by &dateofevent;

      retain start_new_event;
      if _N_=1 then start_new_event = &numevents;
run;

/*This model will estimate r, the growth rate. Create one dataset
for each iteration of the start day.*/
proc nlin data=CGR_dat02 list noitprint;
   parms r 0.75;
   model &numevents = start_new_event*((1+r)**(&dateofevent - &start_day));
   output out       = preddeath_start_&stdyfmt
```

```
            predicted = Pred
          lclm      = Lower95
          uclm      = Upper95;
   ods output ParameterEstimates = r_Estimates;
run; quit;

data r_macro_var;
 set r_Estimates;
 call symput('r_macro_var',estimate);
 run;

/*Model the doubling time, D*/

data CGR_dat03;
 set CGR_dat02;
 r=&r_macro_var;
 run;

proc nlin data = CGR_dat03 list noitprint;
   parms d 3.0;
   if r>=0 then r_D=(exp(log(2)/(d-1))-1);
   else if r<0 then r_D=(exp(log(0.5)/(d-1))-1);

   model &numevents = start_new_event*((1+ r_D )**(&dateofevent -
&start_day));
   ods output ParameterEstimates = doubling_time;
run; quit;

data r_estimate1;
 set r_Estimates;

  r          = Estimate;
  r_LowerCL = LowerCL;
  r_UpperCL = UpperCL;
  start_day = &start_day;
  end_day   = &start_day + &int_length - 1;

  format start_day end_day date9.;
  keep start_day
         end_day
       r
       r_LowerCL
       r_UpperCL;
run;

data doubling_time1;
  set doubling_time;

  doubling_time          = Estimate;
  doubling_time_LowerCL = LowerCL;
  doubling_time_UpperCL = UpperCL;
  start_day              = &start_day;
  end_day                = &start_day + &int_length - 1;

  format start_day end_day date9.;

  keep start_day
```

```
        end_day
        doubling_time
        doubling_time_LowerCL
        doubling_time_UpperCL;
run;


proc sort data = doubling_time1;
      by start_day end_day;
run;

proc sort data = r_estimate1;
      by start_day end_day;
run;

/*Merge the doubling time and growth rate together*/
data r_doubling_time0;
      merge r_estimate1
              doubling_time1;
      by start_day
         end_day;

      mid_day = start_day + round(&int_length/2)-1;  ***add on 4/6/2020;
      format mid_day date9.;
run;

/*Add the current doubling time and growth rate to an overall dataset to
print
  charts*/
proc append base = r_doubling_time
            new  = r_doubling_time0
            force;
run;  quit;


%end;

/*Print the resulting data*/
%do start_day = &fup_start %to (&fup_end-&int_length+1);

%let stdyfmt = %sysfunc(putn(&start_day,date9.));

TITLE "Observed and predicted events and 95% CI for the interval beginning on
&stdyfmt";
proc print data=preddeath_start_&stdyfmt  noobs;
      var &dateofevent
          &numevents
            pred
            lower95
            upper95;

      format &dateofevent mmddyy10.;
run;
%end;

*Now look at the short-term future predictions based on the
 last date of deaths from the incoming dataset. The
```

```
corresponding number of deaths will also be used.;

/*First, get the number of deaths/events on the last day (k) of
the last period for estimating the growth rate*/
proc sql noprint;
      select distinct &numevents into :y_k
      from &indat
      where &dateofevent = &fup_end
;
quit;

/*Next, get r_k from the doubling time, which is the estimated growth rate
 from the last period*/
proc sql noprint;
      select distinct r
                            ,r_lowerCL
                              ,r_upperCL
                              into :r_k,
                         :r_k_lower,
                         :r_k_upper
      from r_doubling_time
      where end_day = &fup_end
;
quit;

%put &fup_end &y_k &r_k &r_k_lower &r_k_upper;

data prediction;

      k=&fup_end;
      do i=1 to &int_length;
        m          = k + i;
        y_m        = round(&y_k*((1+&r_k))**(m-k));
        y_m_lowerCL = round(&y_k*((1+&r_k_lower))**(m-k));
        y_m_upperCL = round(&y_k*((1+&r_k_upper))**(m-k));
        output;
       end;
       format m date9.;

       keep m
          y_m
          y_m_lowerCL
          y_m_upperCL;
run;

title "Predicted number of deaths for the next &int_length days";
proc print data=prediction noobs;
run;

%mend Calc_GrowthRates;
```