medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 and it causes 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. We fit a series of growth curves using a rolling approach to estimate the growth rates and the doubling times. This approach was applied to the death data of New York State during March 14th and 31st. The growth rate was 0.48 (95% CI, 0.39-0.57) and the doubling time was 2.77 days (95% CI, 2.49-3.04) for the period of March 14th-March 20th; the growth rate decreased to 0.25 (95% CI, 0.22-0.28) and the doubling time increased to 4.09 days (95% CI, 3.73-4.44) for the period of March 25th-March 31st. This approach can be used for short-term prediction and monitoring the 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 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.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2. METHODS

2.1 A rolling growth curve approach (RGCA)

In the United States, there were 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 was often under-identified and under-reported. However, the number of hospitalized COVID-19 patients and number of deaths due to COVID-19 were more reliable than the reported number of COVID-19 cases [10]. In this paper, we used the number of daily deaths to calculate the growth rate and doubling time in days.

We assumed a growth curve of daily deaths over a period of n days from day t (start day) to day $(t+n-1)$. Let $y_{(t+j-1)}$ denote the daily deaths 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 deaths at the start day t; r is the growth rate. When the growth rate $r > 0$, the number of daily deaths increases. For example, if r =0.4, the growth rate of deaths is 40% more for each day. When growth rate $r = 0$, the number of daily deaths has no change. When growth rate $r < 0$, the number of daily deaths declines. When the number of deaths 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 deaths 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 deaths by half.

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 death data from New York State from March 14th – March 31st [12]: March 14th-20th,15th-21st, 16th-22nd. . ., 25th-31st.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

2.2 Short-term prediction

The estimated growth rate from the last period of the RGCA approach (e.g., March 25th-31st) can be used for future short-term prediction of deaths. Let k denote the last day of the last period, y_k is the number of deaths on this day. For the New York death data in this analysis, k is March 31st, $y_k = 374$. 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

While the growth rate had a spike at March 20th-21th, subsequent days are lower with a decreasing trend for March 22th-28th (Figure 1). Consistently, the doubling time in days had an increasing trend most evident in the later time as well (Figure 2).

While the graphs provide a useful visual tool, the macro supports calculating rates within defined time intervals of interest. At the beginning of this observation period (March 14th-March 20th), the growth rate was 0.48 (95% CI, 0.39-0.57) and the doubling time was 2.77 days (95% CI, 2.49-3.04). At the end of this observation period (March 25th-March 31st), the growth rate decreased to 0.25 (95% CI, 0.22-0.28) and the doubling time increased to 4.09 days (95% CI, 3.73-4.44).

Using the estimated growth rate from the last period March 25th-March 31st, $\hat{r}_k = 0.25$, the predicted numbers of daily deaths for April 1st and 2nd were 468 and 586, respectively. The observed number of deaths in New York State was 498 on April 1st. SAS programs are available for conducting these analyses (Appendix A and Appendix B).

4. DISCUSSION

These models can be similarly applied to hospitalization data if those data are available. 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 death, it is recommended to exclude the recent 1-2 days' death data in fitting the growth curves. This paper illustrates that death data can be used to estimate the growth rate and doubling time to aid predicting future deaths, hospitalizations 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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://coronavirus.1point3acres.com/en

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1. Estimated growth rate with 95% CIs over time using death data from New York State.

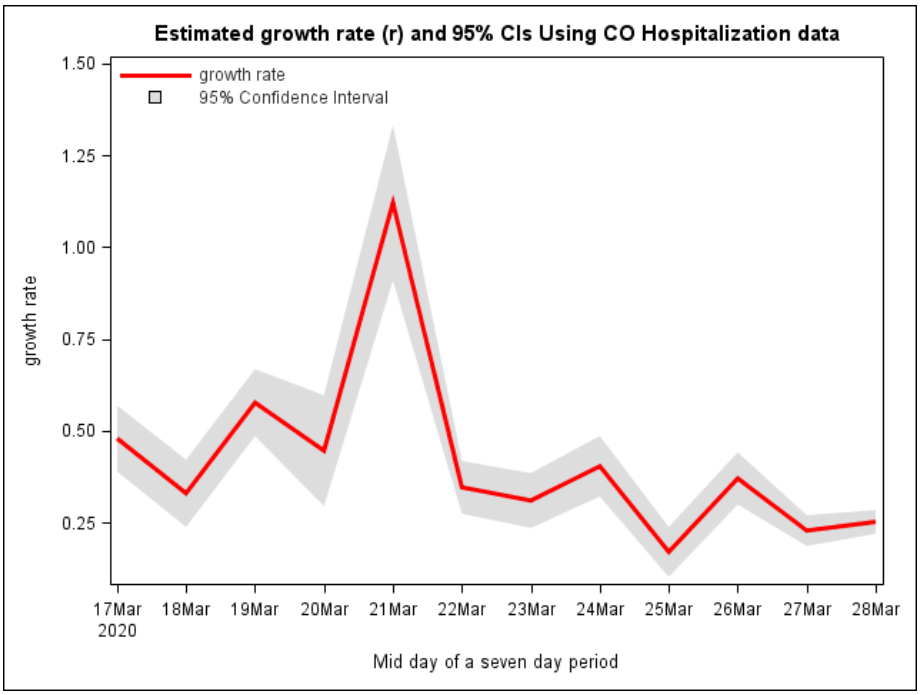
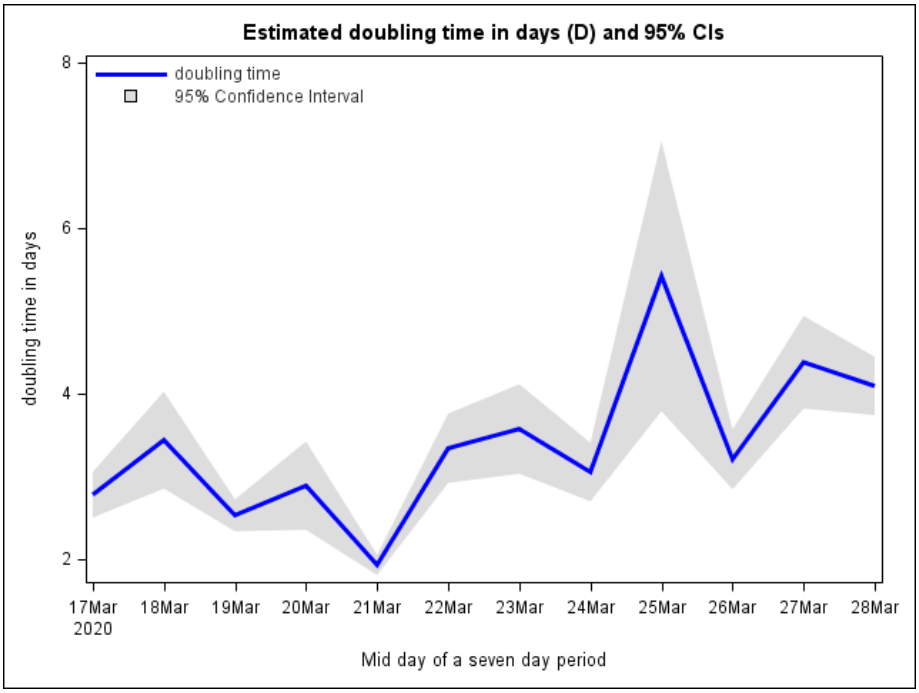


Figure 2. Estimated doubling time in days with 95% CIs over time using death data from New York State.



medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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:\Users\L822155\OneDrive - Kaiser Permanente\Analytic
Projects_OfflineDev\COVID-19\growthratedoublingmacro;

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

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

/*Clean up the datasets */
proc datasets library=work nolist;
      delete cgr_dat01
             cgr_dat02
               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 = numdeaths
                        , int_length = 7);

title 'Growth rate and doubling time for each interval';
proc print data = r_doubling_time;
      var start_day
          end_day
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

```
            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";

title 'Estimated growth rate (r) and 95% CIs Using CO Hospitalization data';
proc sgplot data=r_doubling_time;
      band x = mid_day lower = r_LowerCL upper = r_UpperCL/
            fillattrs = (color = graydd) name='band95' legendlabel='95%
Confidence Interval';
      series x = mid_day y = r/
            lineattrs = (color = red thickness = 3px) name = 'line'
legendlabel = 'growth rate';

      yaxis labelpos = center label = "growth rate" thresholdmax=.8;
      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;


title 'Estimated doubling time in days (D) and 95% CIs';
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;
      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;
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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;
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

```
         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;
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

```
/*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;
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

```
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;
```

medRxiv preprint doi: https://doi.org/10.1101/2020.04.08.20057943; this version posted April 11, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

```
/*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;
```