Forecasting the evolution of COVID-19 from daily data

Oliver Linton
Trinity College and Faculty of Economics, University of Cambridge
with the assistance of Shaoran Li, King's College

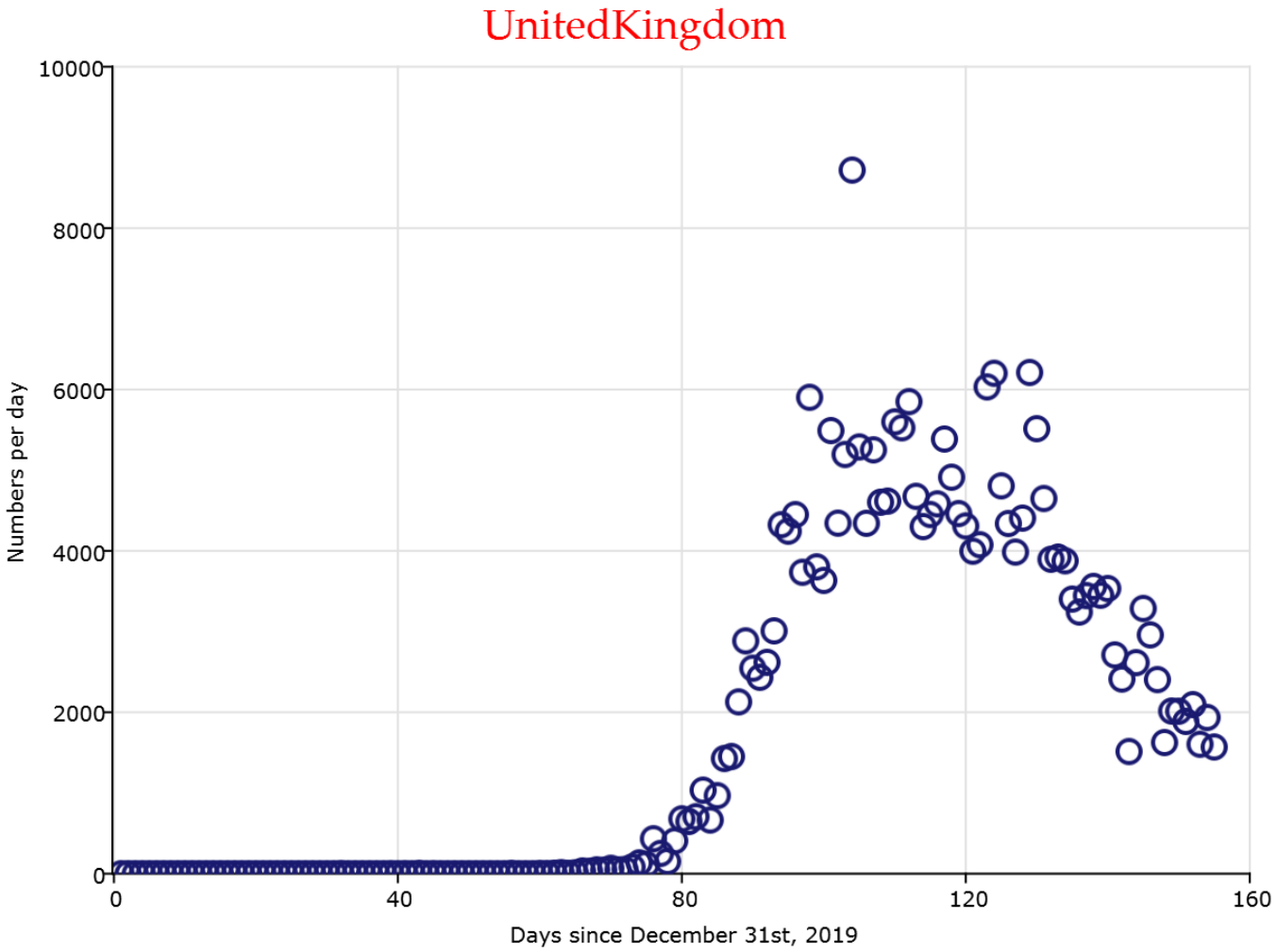
Cambridge University, June 2020

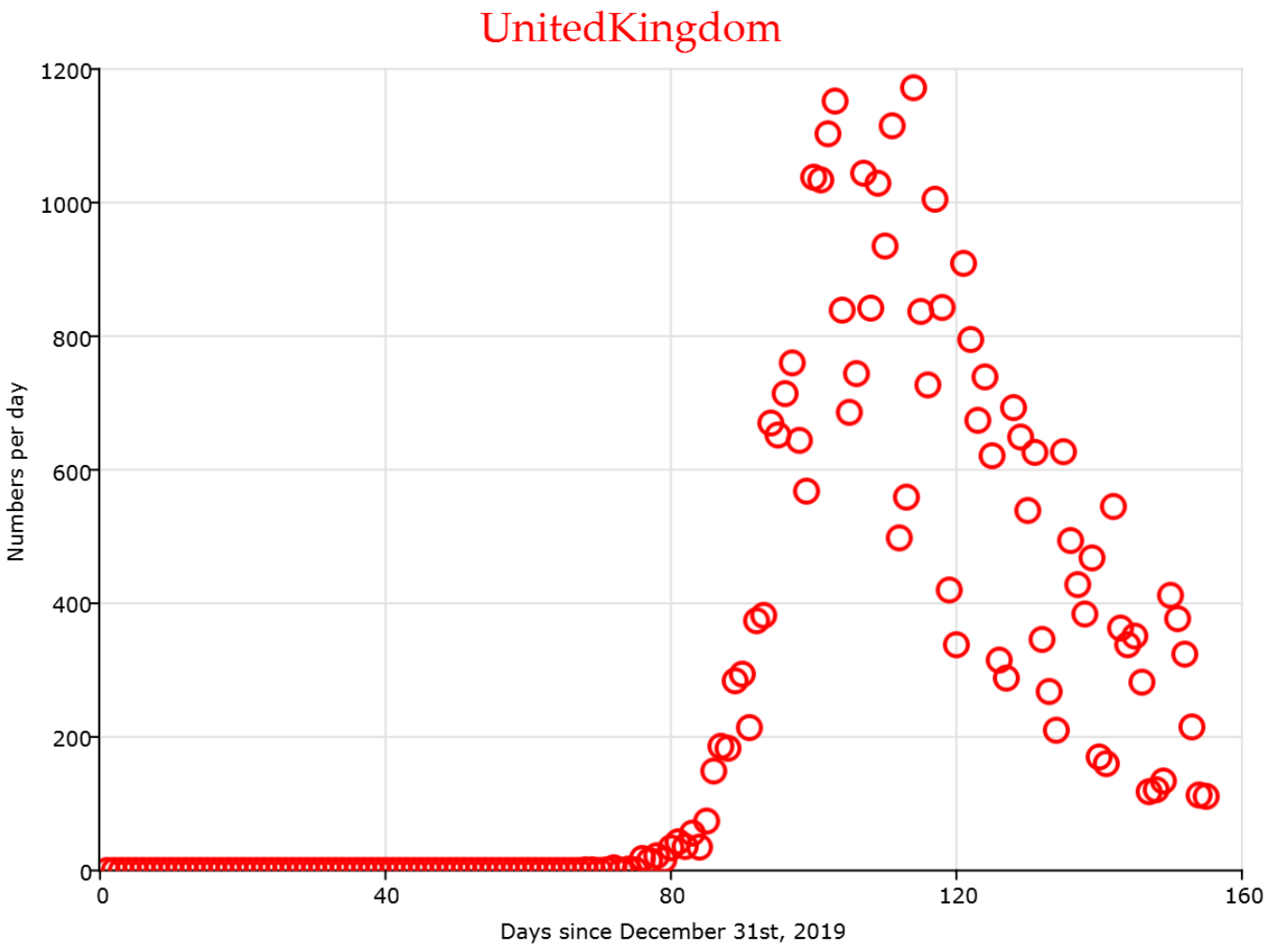
- We model the daily new COVID-19 reported cases and new reported deaths data available from the European Centre for Disease Protection and Control (ECDC) website since December 31st 2019 (Day 1).
- We started this project April 2nd when the number of cases and deaths were growing rapidly in the UK. The initial purpose was to predict the turning point of the epidemic in the UK
- We broadened the project to include all countries and
  - to estimate the total number of cases and deaths expected to transpire
  - to estimate the expected length of time from "peak" to "trough"
  - to provide one day ahead forecasting
  - to evaluate day of the week effects and functional form
- We post new estimates everyday (Thanks to Jake Dyer) at

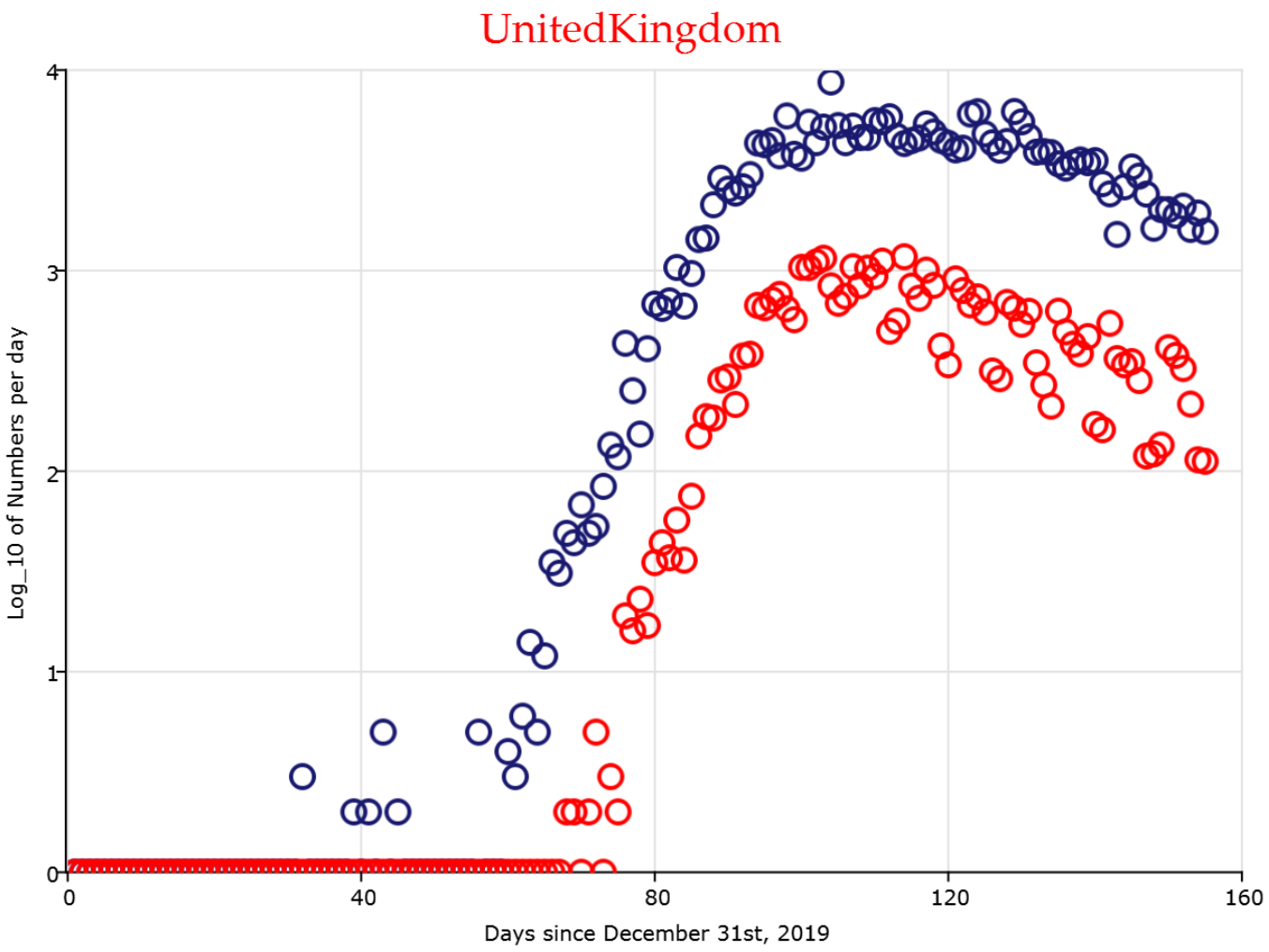
http://covid.econ.cam.ac.uk/linton-uk-covid-cases-predicted-peak

- The trend is the main feature of the data. However, these data are unlike typical economic time series such as GDP and stock market prices that have been relentlessly upward trending. These data have both upward and downward segments.
- We use simple linear in parameters models to capture the trend and provide estimates of the model parameters, which allows us to extrapolate into the future

Blue is cases; Red is deaths




UnitedKingdom
Numbers per day
Days since December 31st, 2019
0
200
400
600
800
1000
1200
0
40
80
120
160

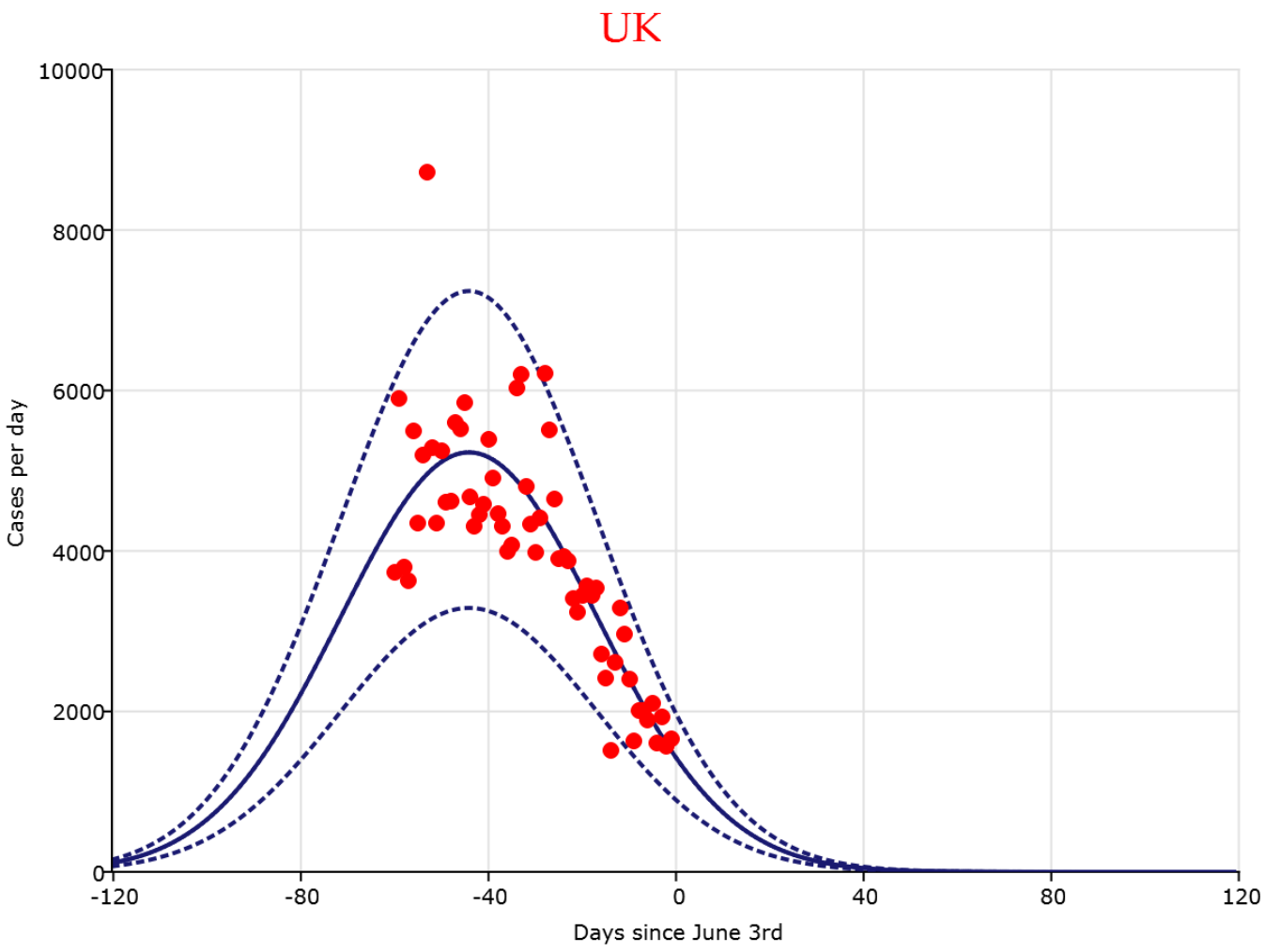


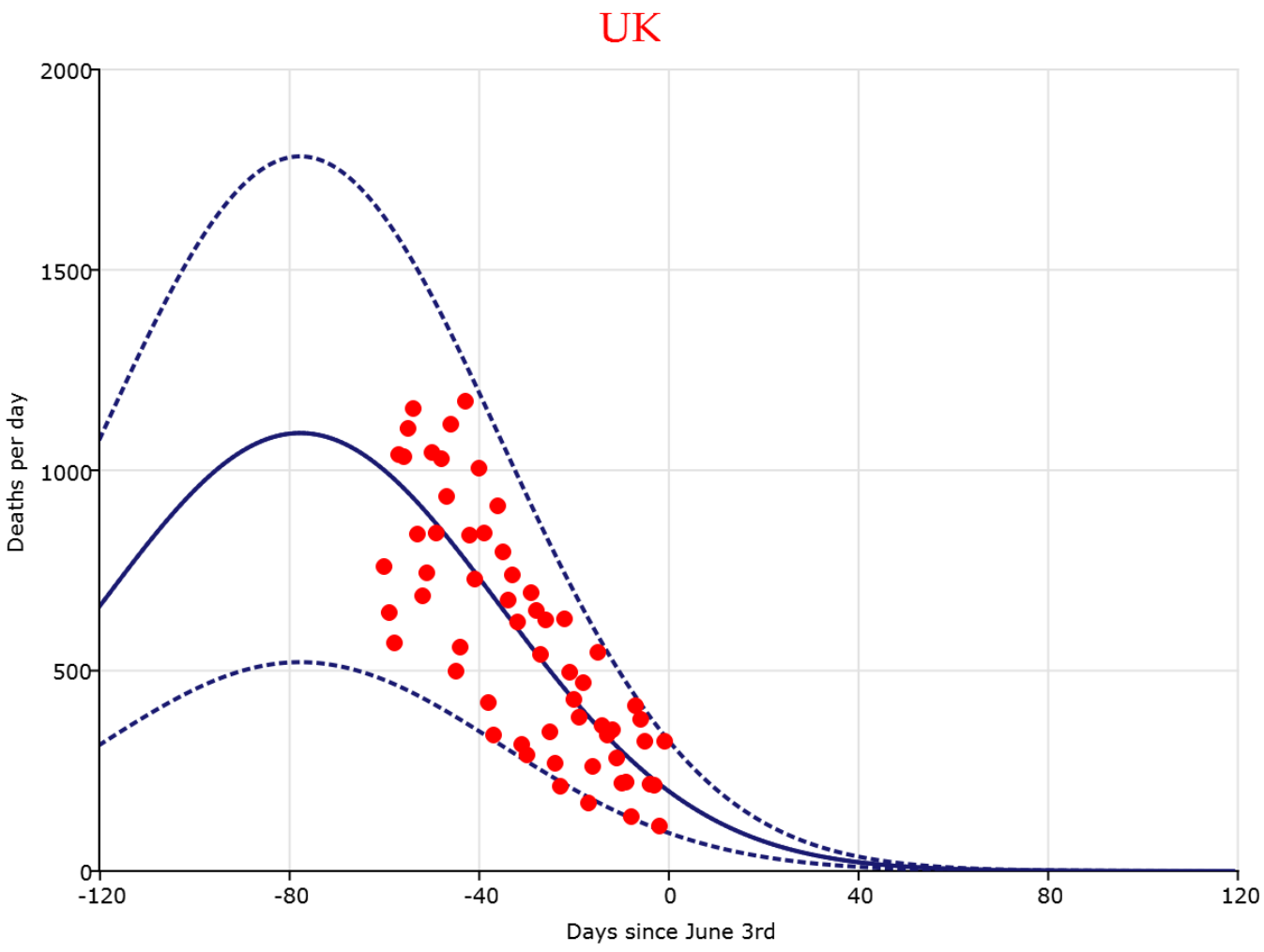
Schools closed 20th March = day 80

Methodology

- Using the most recent data we fit a number of different parametric trend regression models for the logarithm of cases and deaths that allow for a turning point such as:
  - a quadratic trend (transforming to levels implies a Gaussian/Normal bell curve for cases)
  - a quartic trend (transforming to levels gives a "fat bottomed" curve that has an elongated peak)
  - a linear and log trend (transforming to levels gives a chi-squared or gamma shape that allows for asymmetry, e.g., a slower down-stage than up-stage)
- We provide prediction intervals around the curves that reflect the intrinsic uncertainty about predictions. These assume that the model is correct and that the parameters are known knowns.

Solid line is estimated curve, dotted lines are 95% prediction intervals

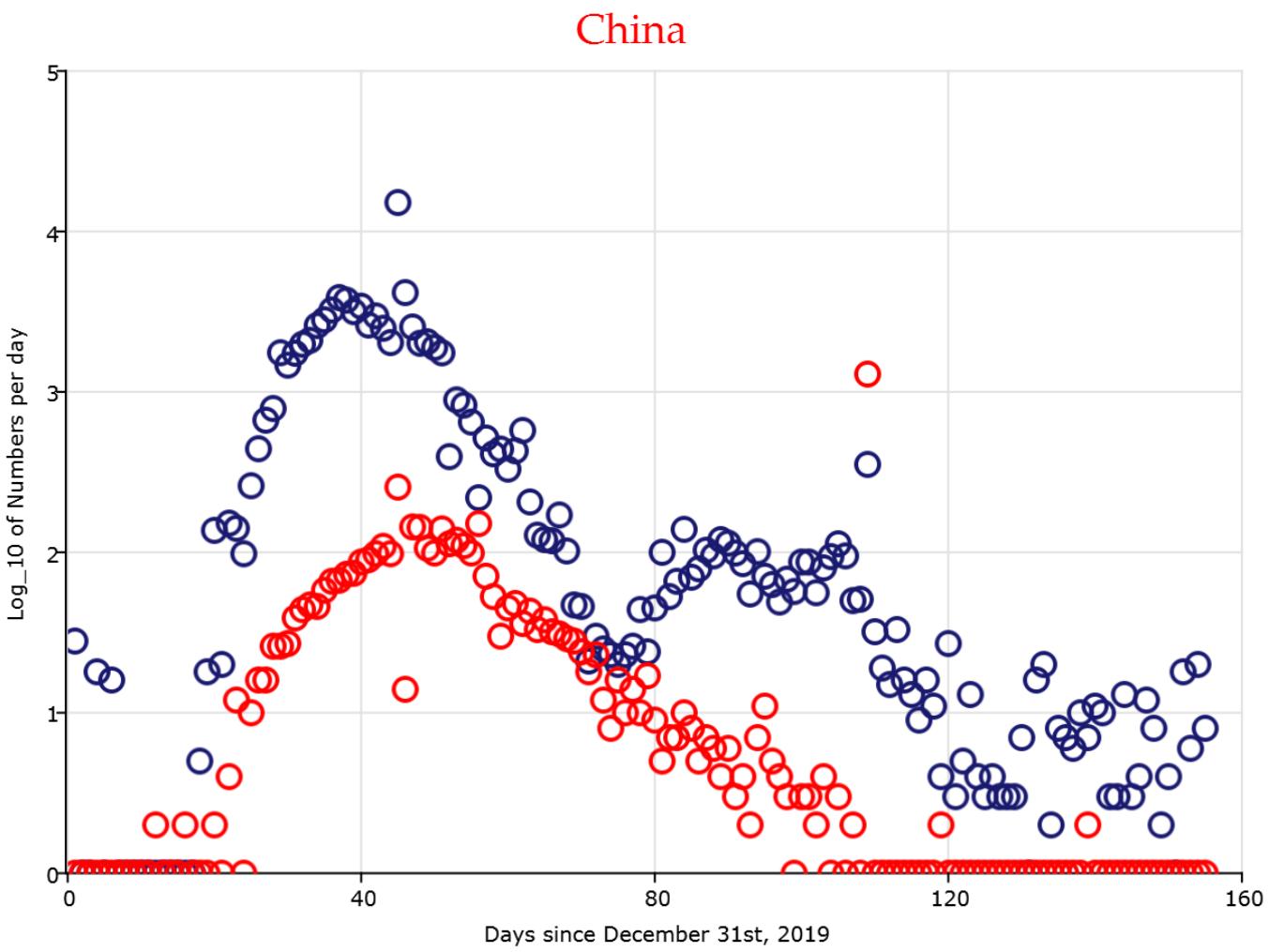


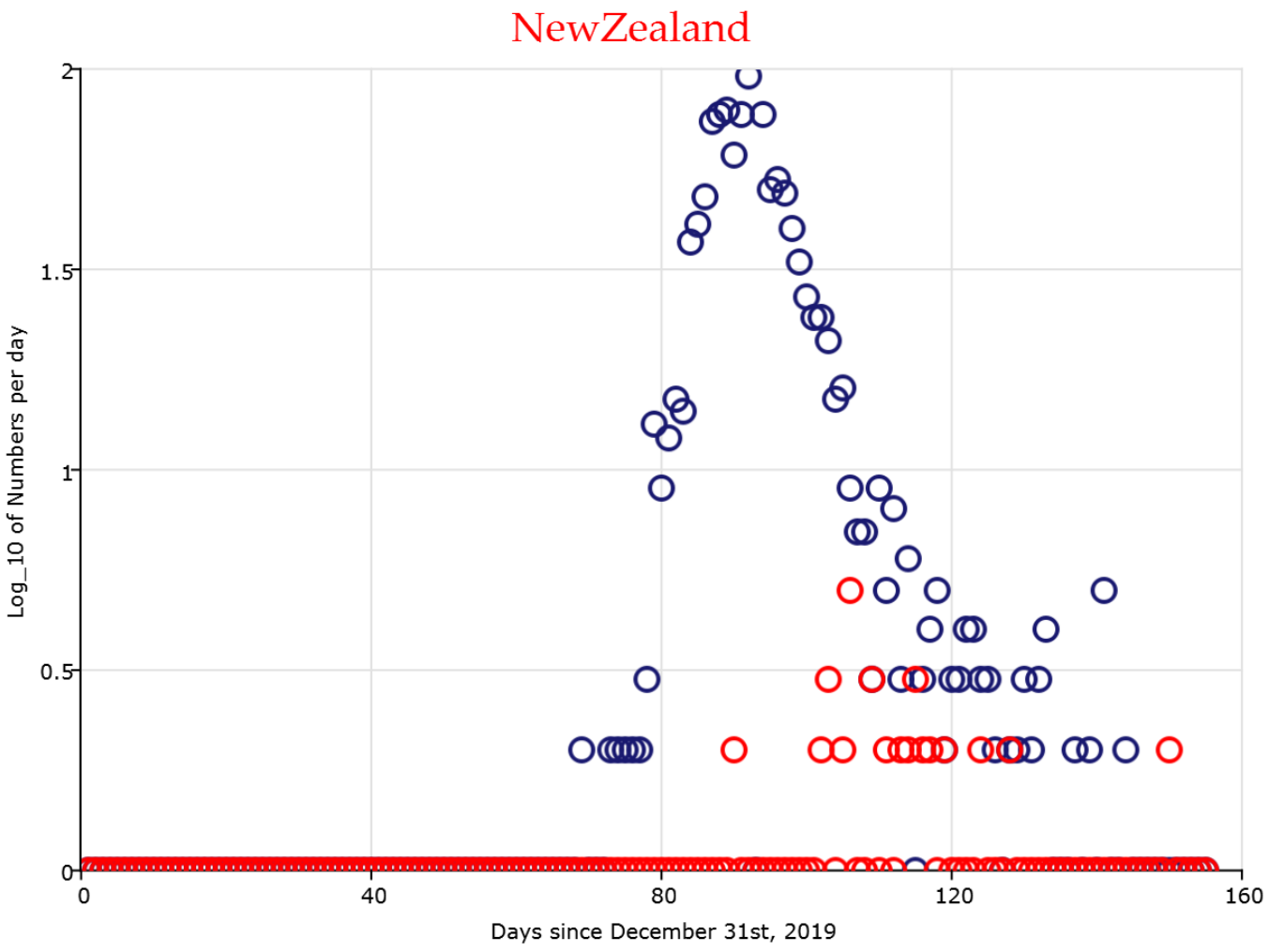

UK
2000
1500
1000
500
0
Deaths per day
-120
-80
-40
0
40
80
120
Days since June 3rd

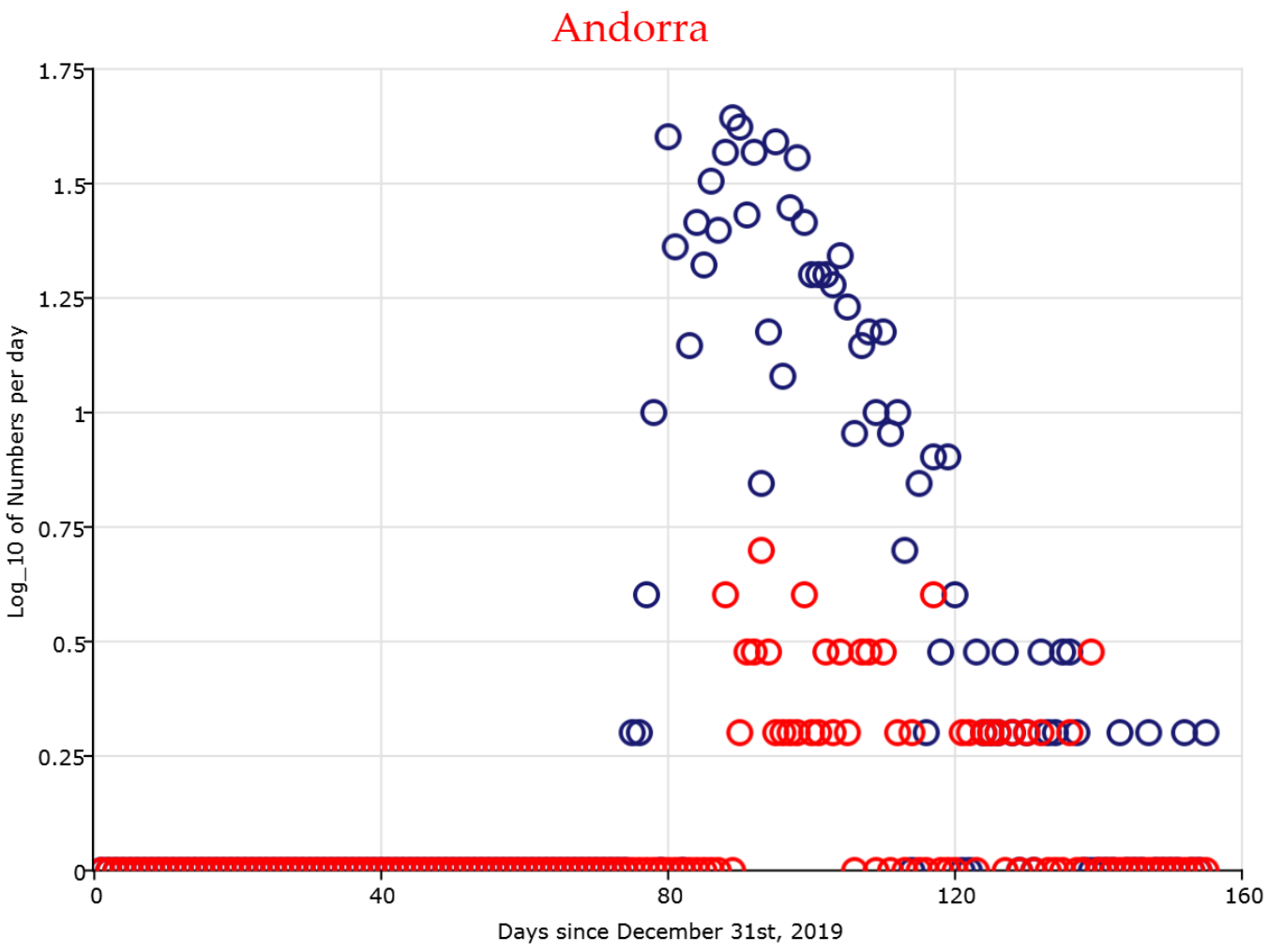
Where the World is now

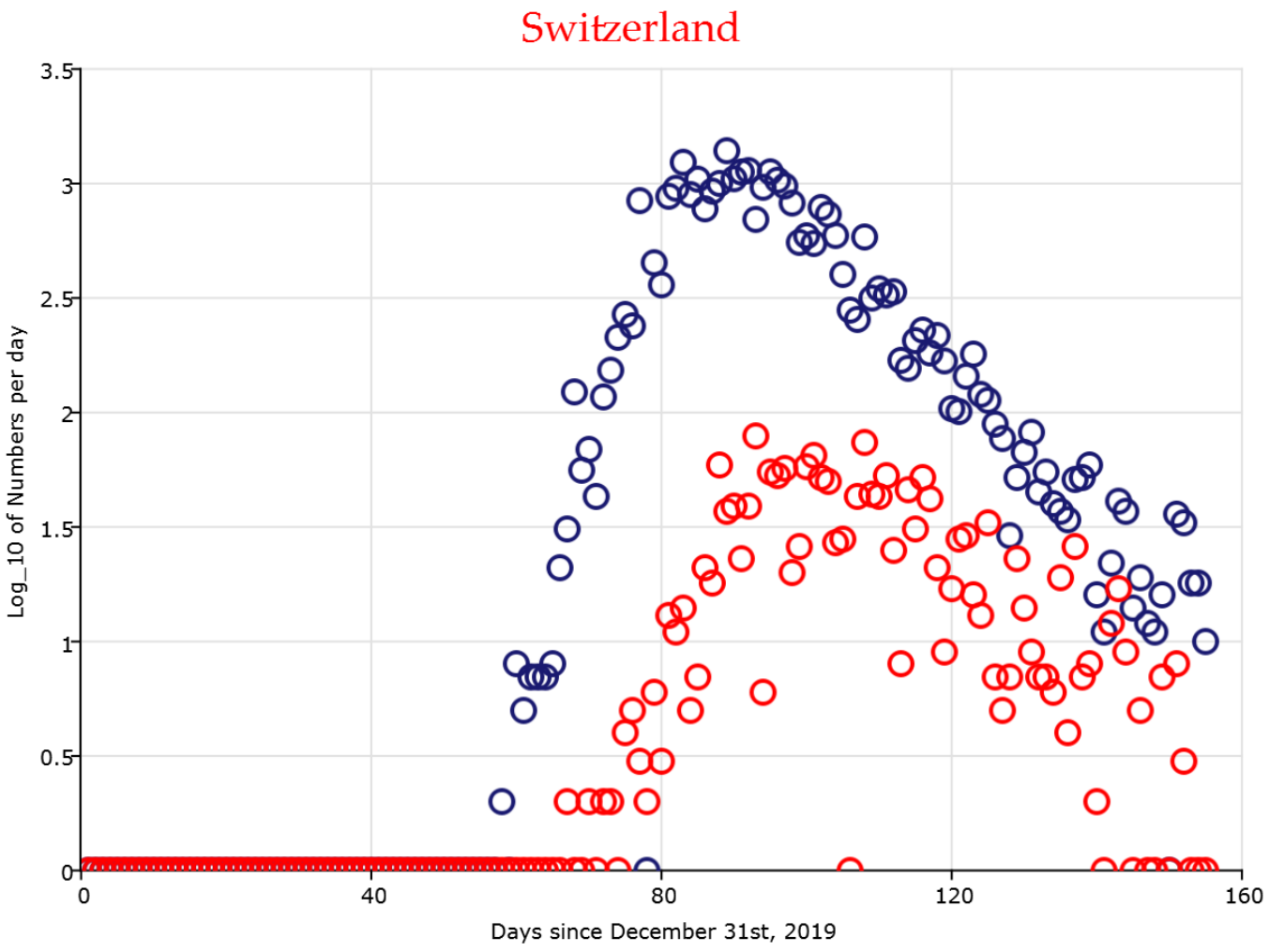
- We categorize Countries or Territories as:
  - Early stages (before peak)
    - ★ Brazil, Chile, Peru, Colombia, Mexico, Egypt, Nigeria, Kenya, South Africa, Kuwait, Qatar, Afghanistan, Iraq, Saudi Arabia, Bangladesh, India, Pakistan, etc.
  - Middle (peak passed)
    - ★ USA, Canada, UK, Sweden, Poland, Romania, Russia, Ukraine, Iran, Turkey, Morocco, Algeria, Chad, Senegal, Ghana, Sudan, Phillipines, Indonesia etc.
  - Endgame ($<$1/10 of peak)
    - ★ Diamond Princess, China, Taiwan, S. Korea, Japan, Thailand, Australia, New Zealand, Ireland, Andorra, Iceland, Germany, Italy, Spain, France, Ireland, Luxembourg, Norway, Slovenia, Isle of Man, Jersey, Guernsey, Falkland Islands, Greece, Austria, Croatia, Cyprus, Cuba, Israel, Niger etc.

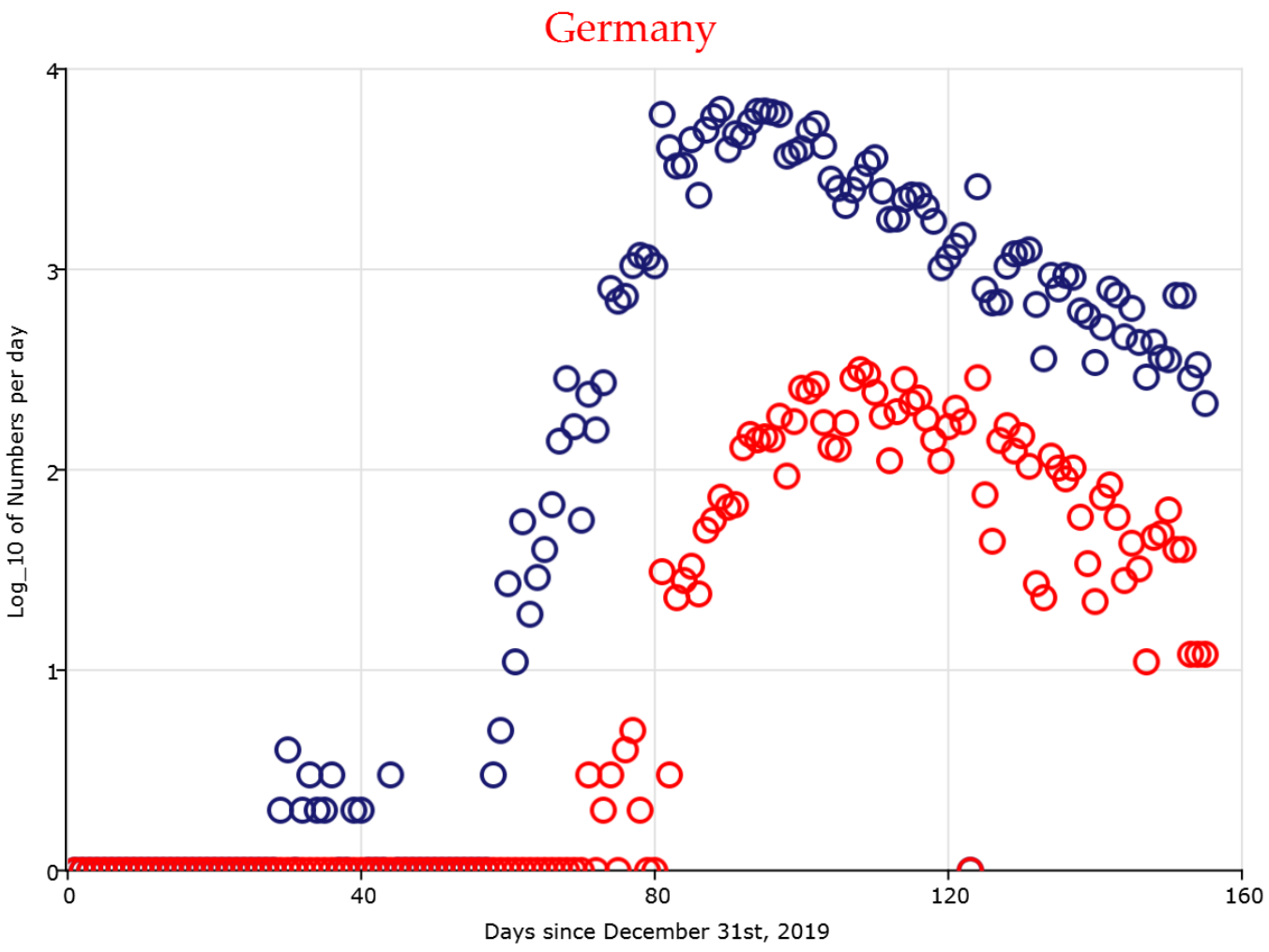
End Game

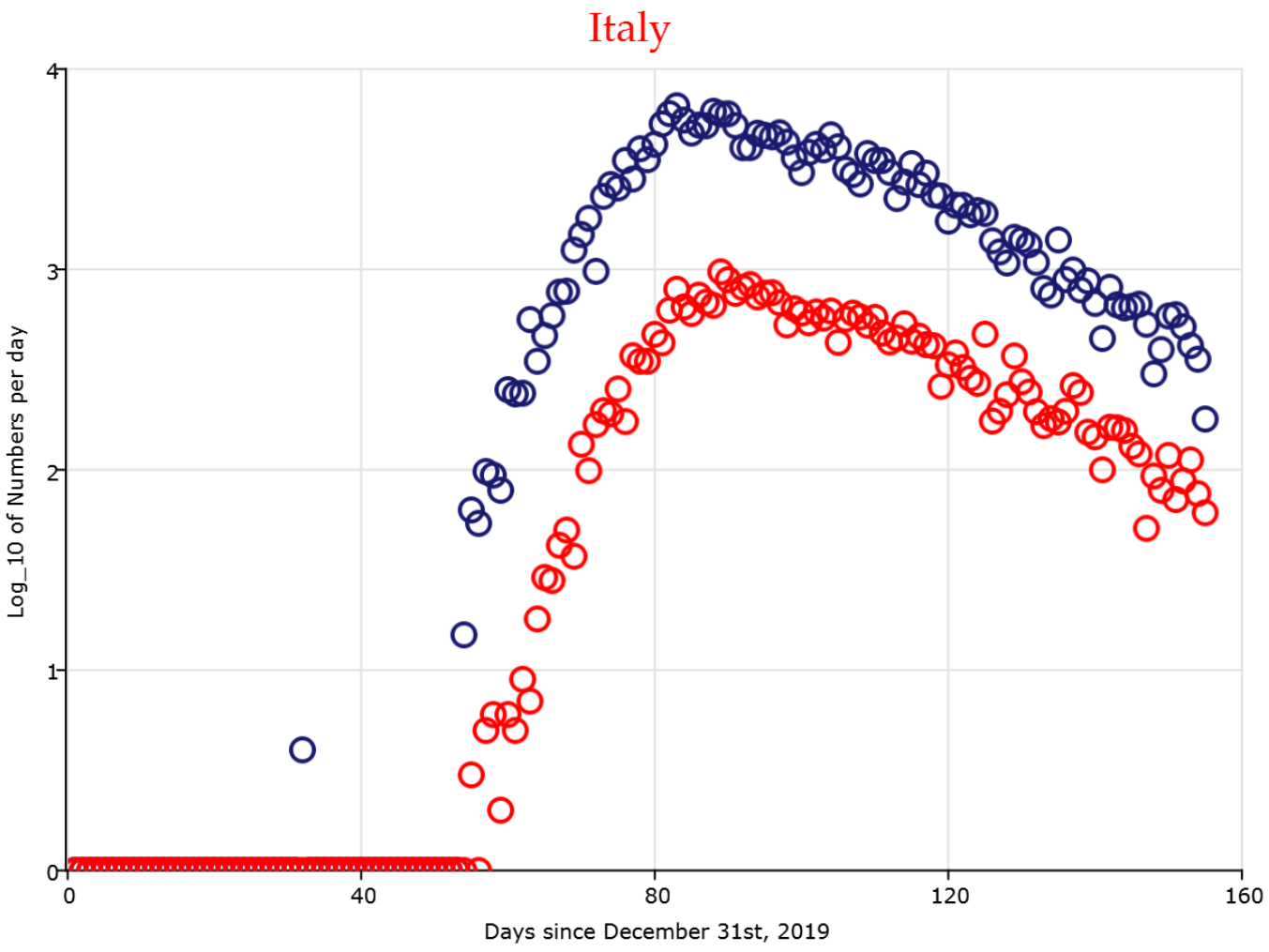

China
Log_10 of Numbers per day
Days since December 31st, 2019
0
1
2
3
4
5
0
40
80
120
160

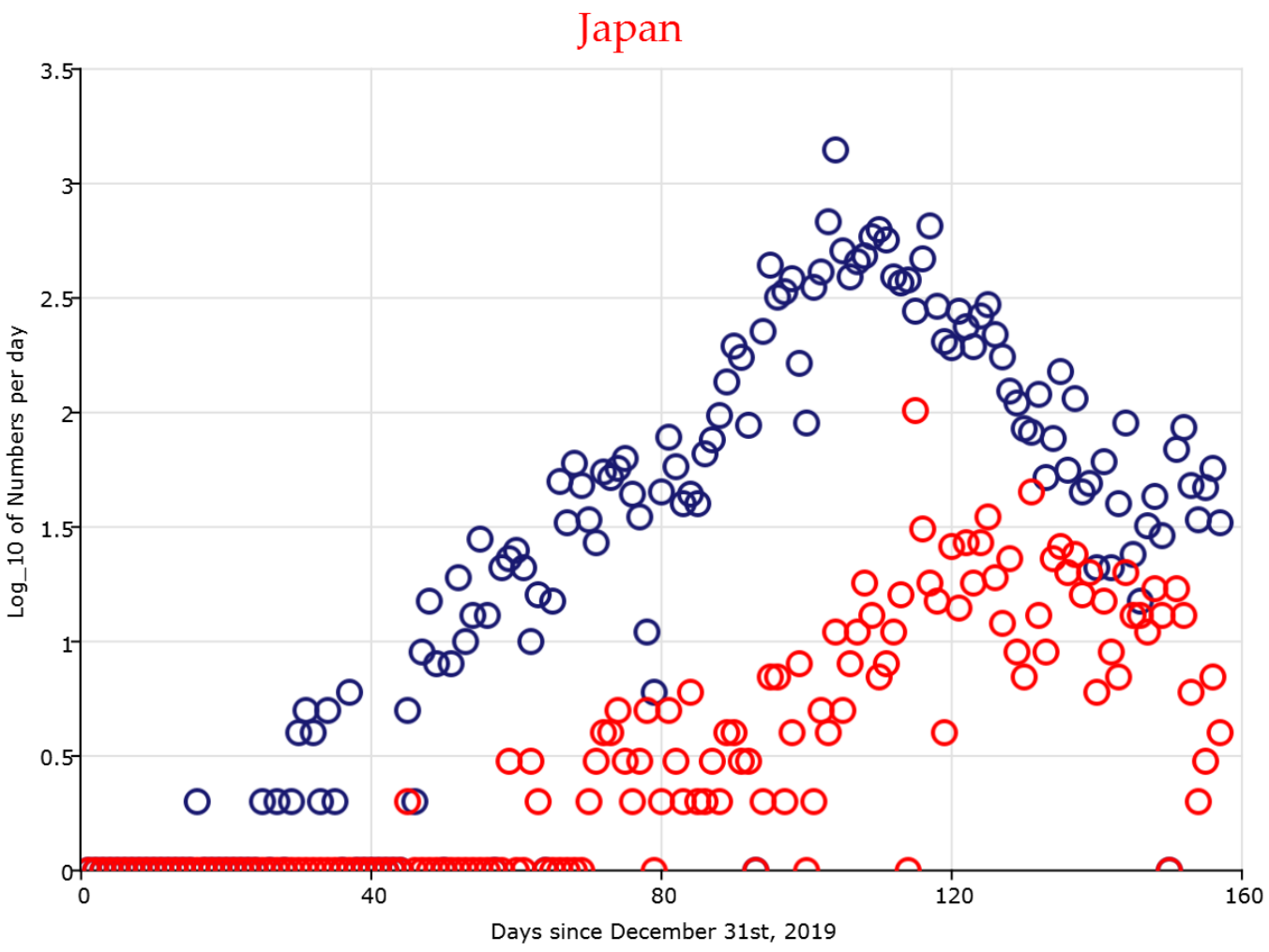

NewZealand
Log_10 of Numbers per day
Days since December 31st, 2019
0
0.5
1
1.5
2
0
40
80
120
160


Andorra
1.75
1.5
1.25
1
0.75
0.5
0.25
0
Log_10 of Numbers per day
0
40
80
120
160
Days since December 31st, 2019

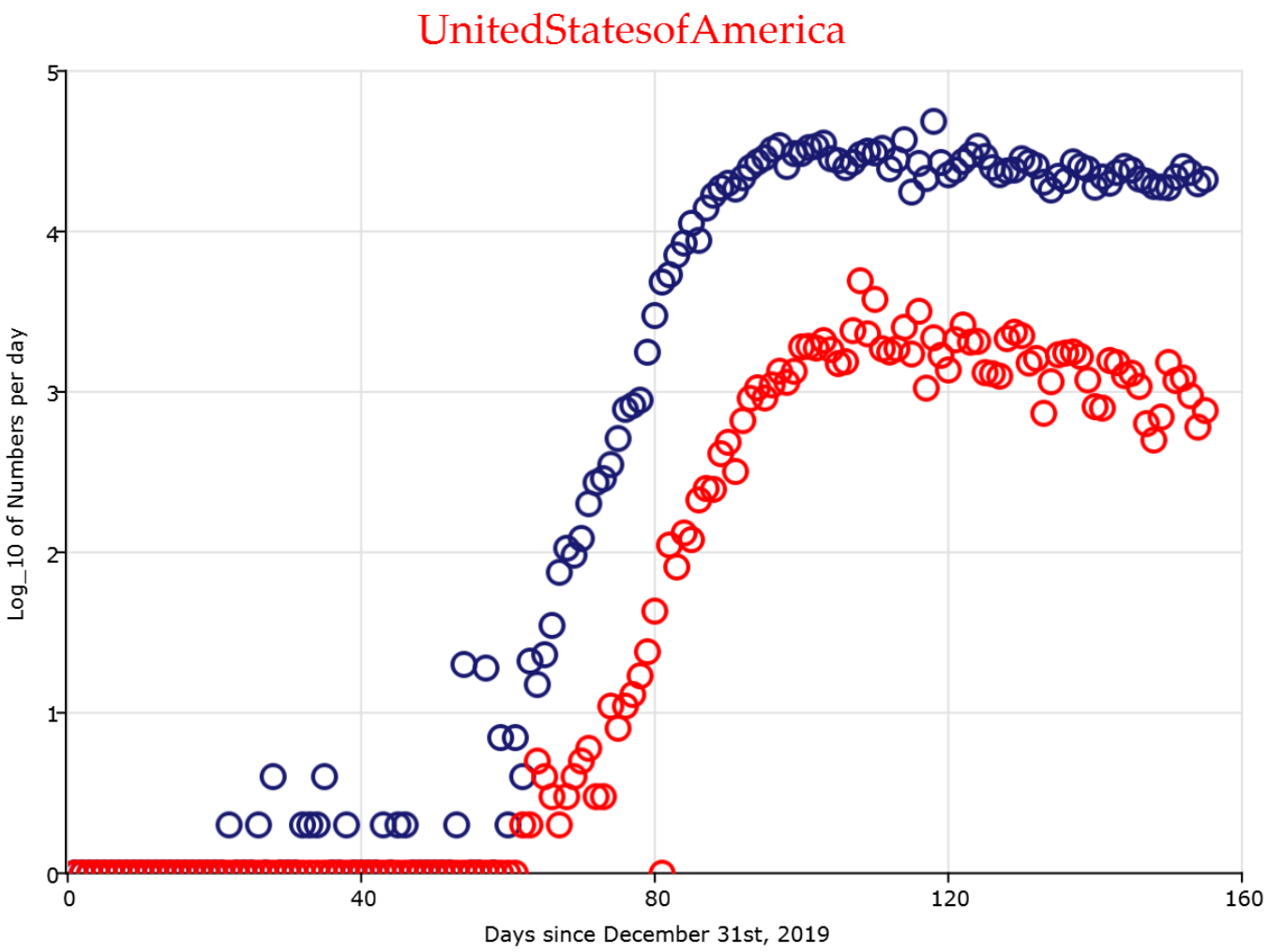

Switzerland
Log_10 of Numbers per day
Days since December 31st, 2019
0
0.5
1
1.5
2
2.5
3
3.5
0
40
80
120
160

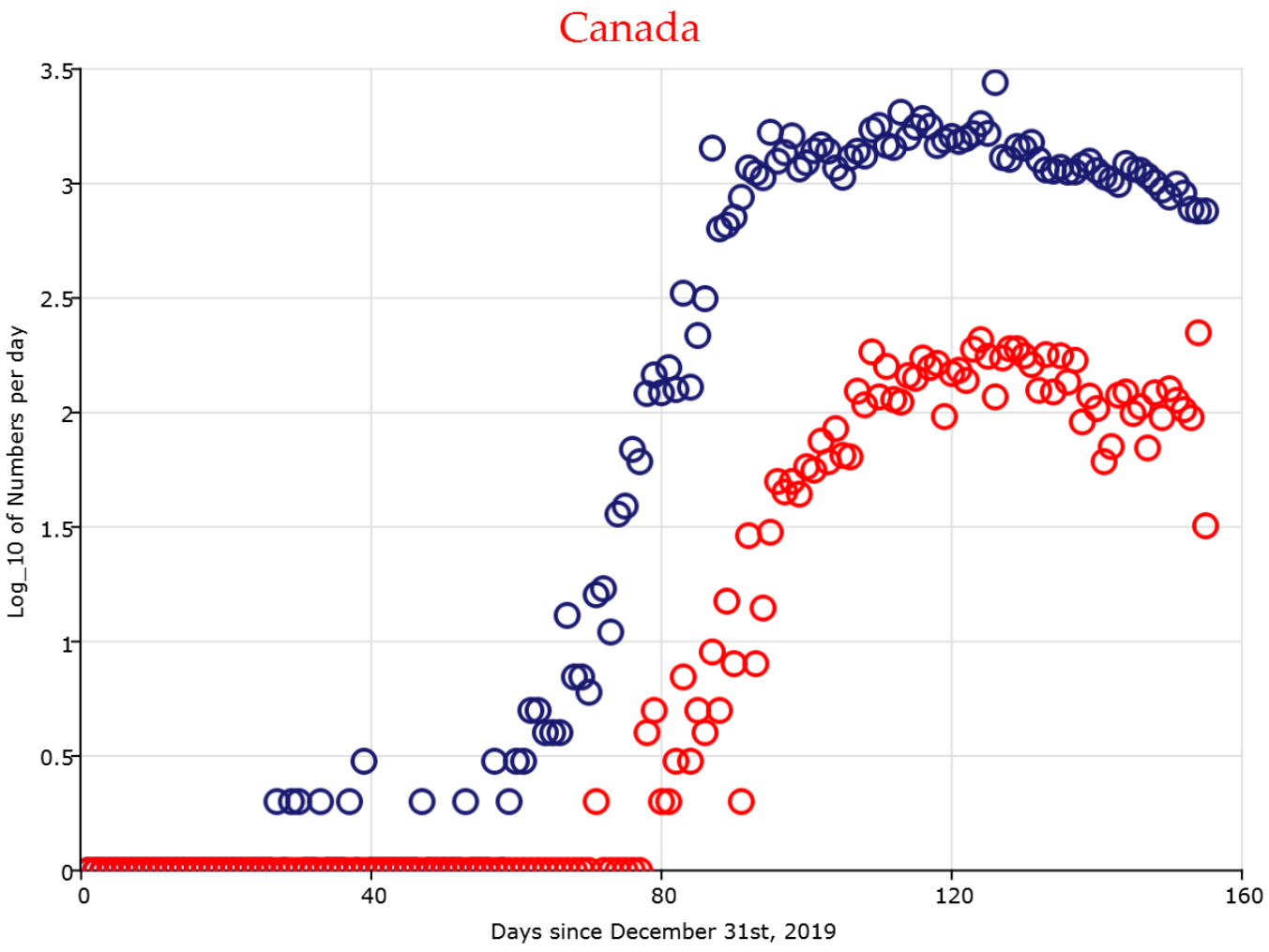

Germany
Log_10 of Numbers per day
Days since December 31st, 2019
0
1
2
3
4
0
40
80
120
160

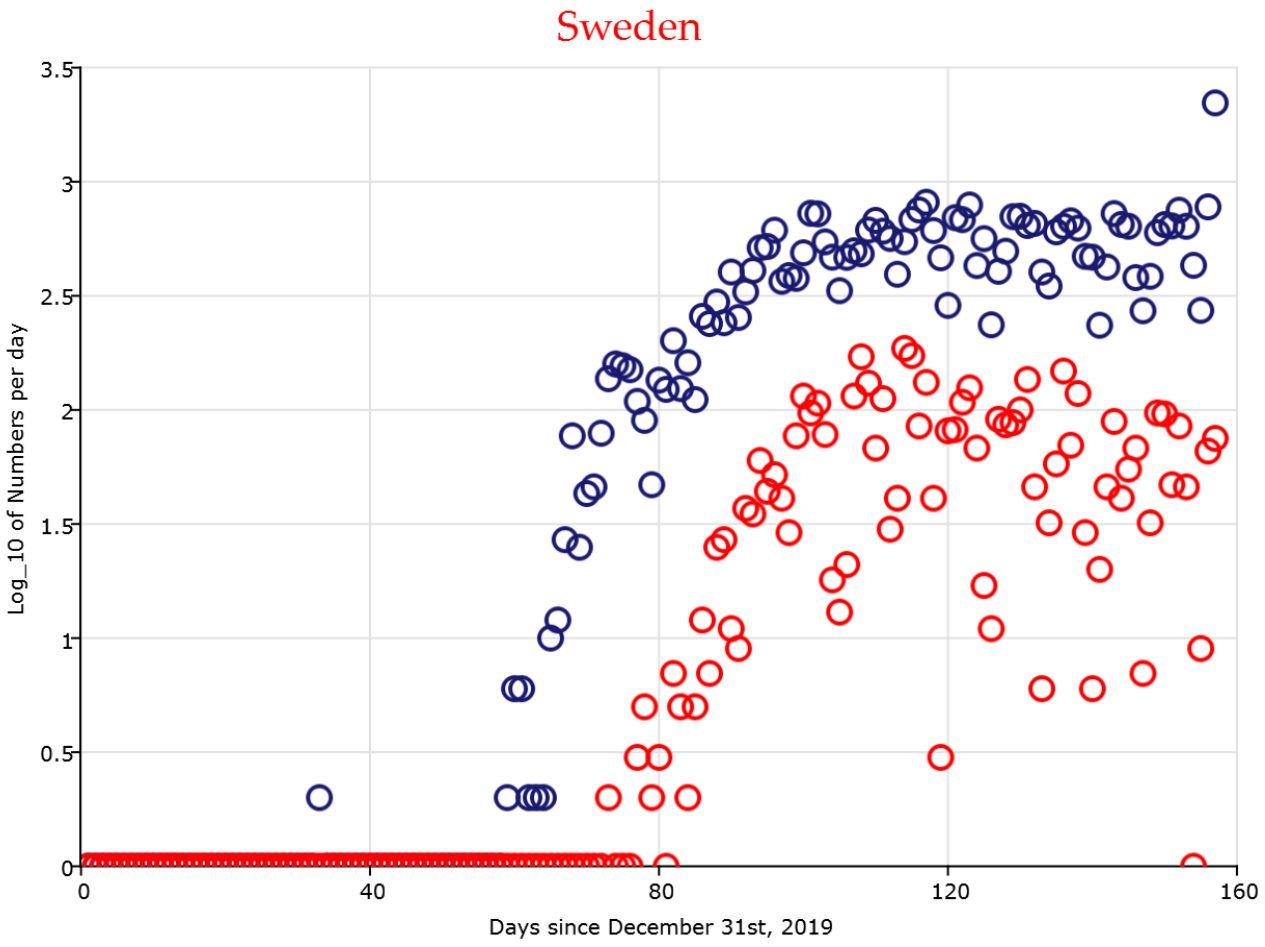

Italy
Log_10 of Numbers per day
Days since December 31st, 2019
0
1
2
3
4
0
40
80
120
160

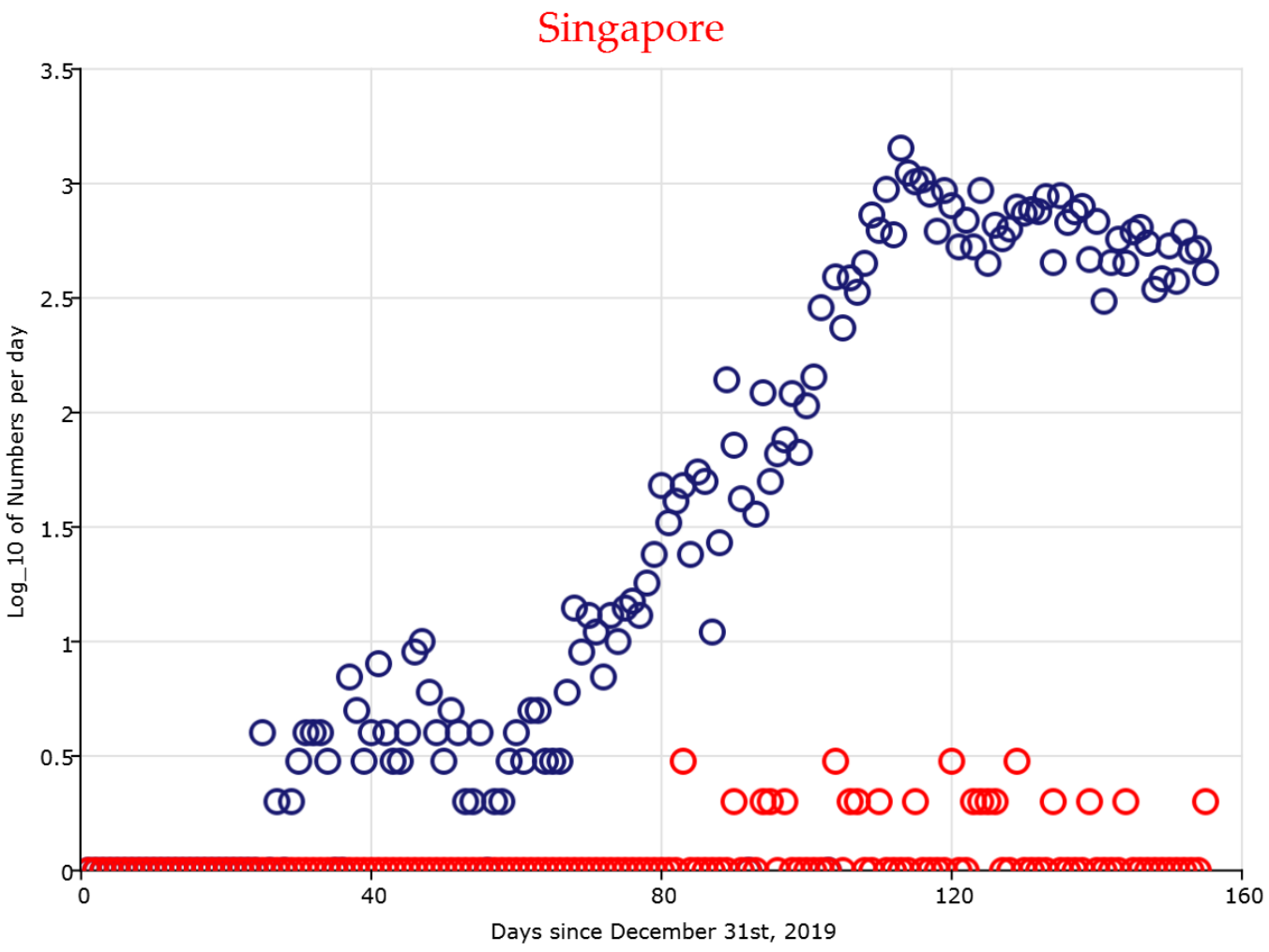

Japan
Log_10 of Numbers per day
Days since December 31st, 2019
0
0.5
1
1.5
2
2.5
3
3.5
0
40
80
120
160

Middle Game

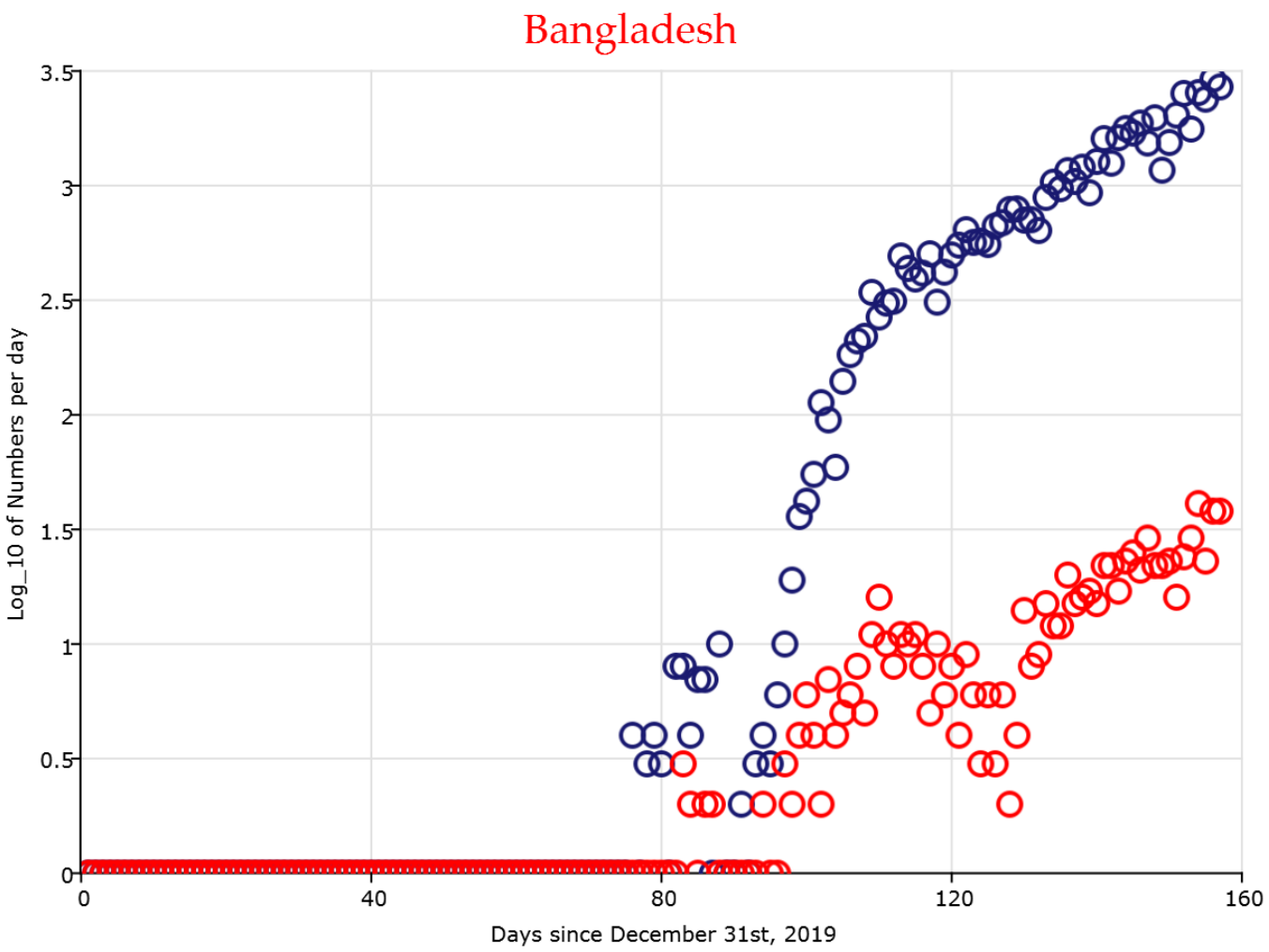

UnitedStatesofAmerica
Log_10 of Numbers per day
Days since December 31st, 2019
0
1
2
3
4
5
0
40
80
120
160

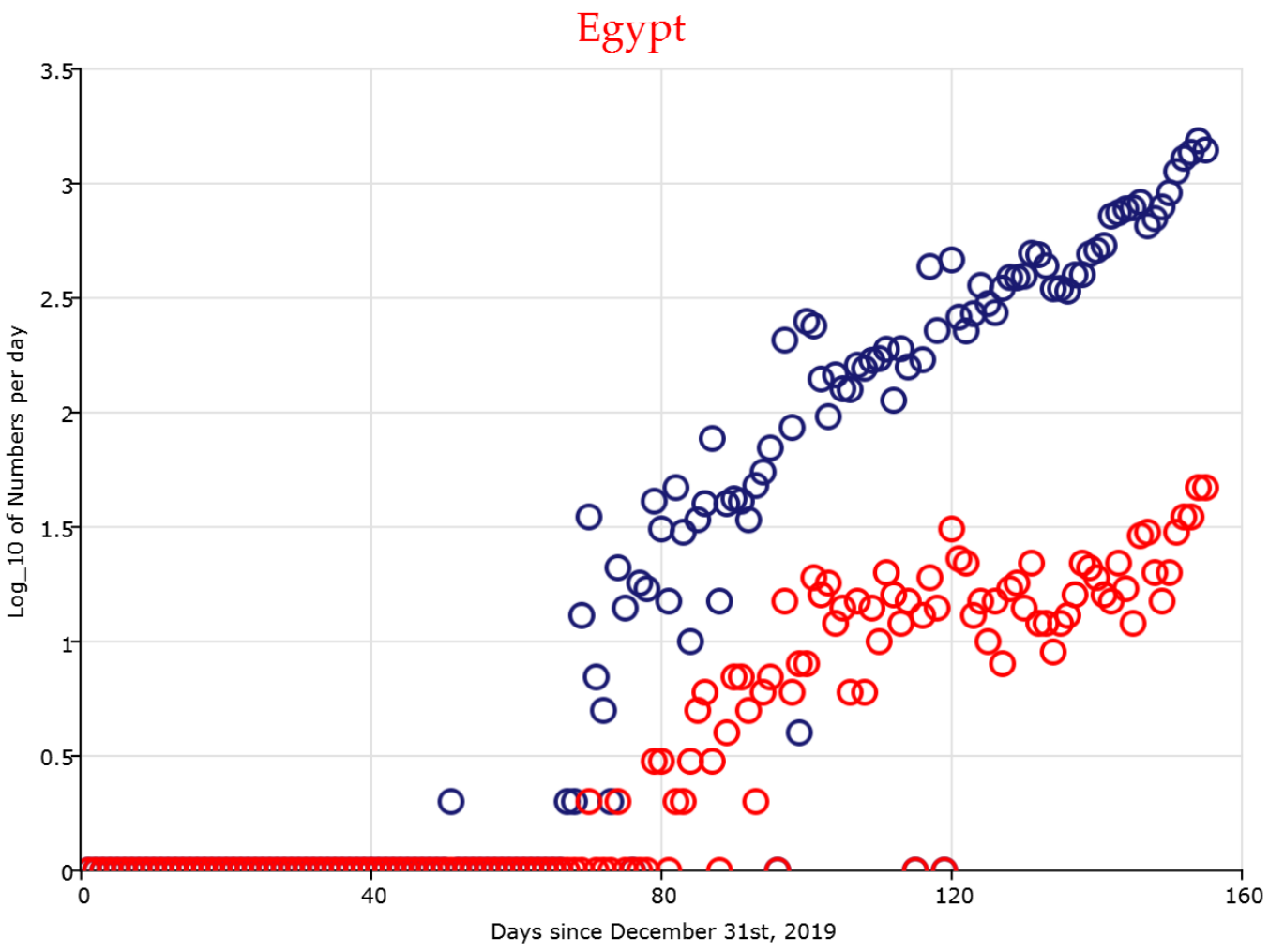

Canada
3.5
3
2.5
2
1.5
1
0.5
0
Log_10 of Numbers per day
0
40
80
120
160
Days since December 31st, 2019

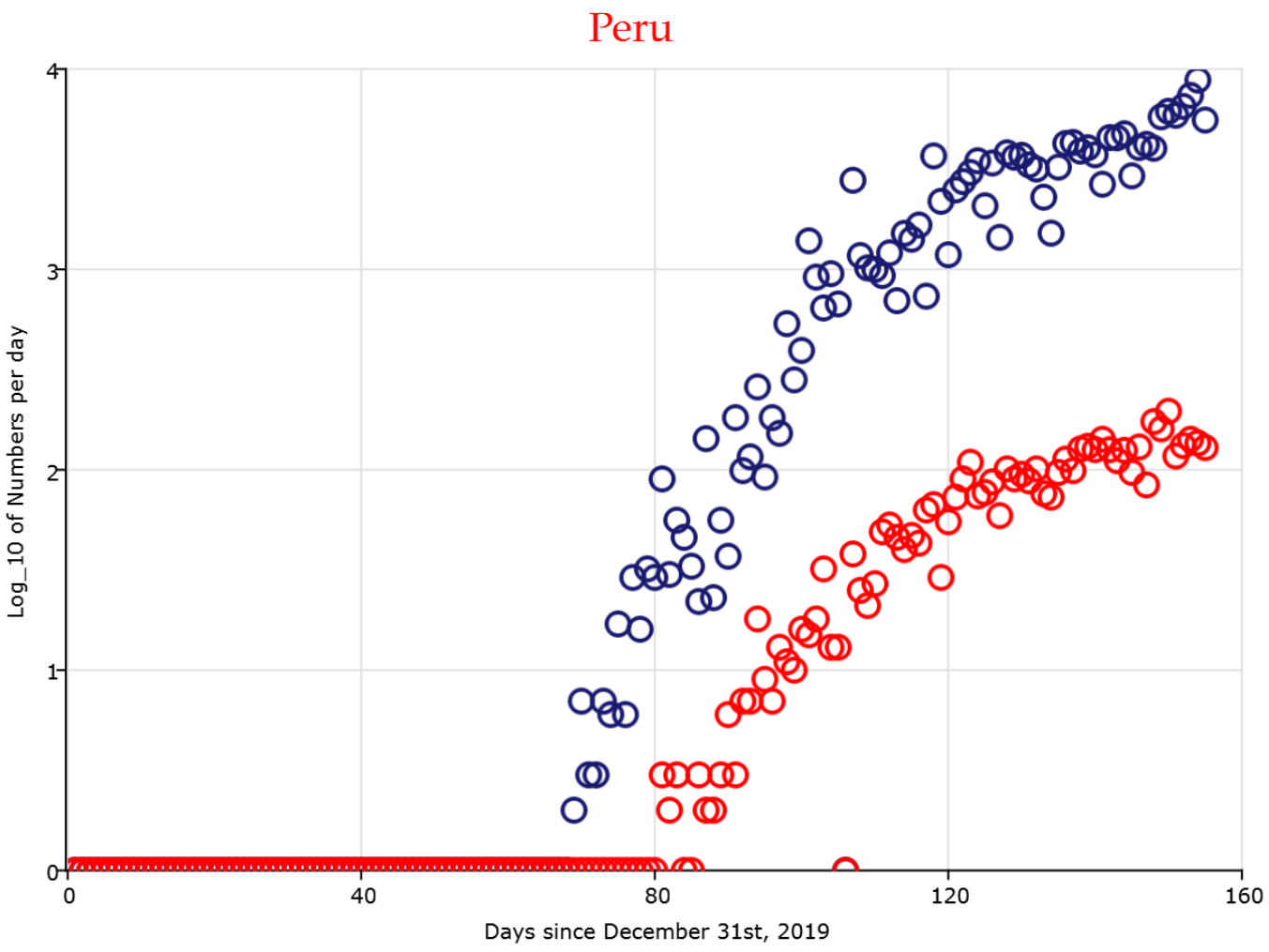

Sweden
Log_10 of Numbers per day
Days since December 31st, 2019
0
0.5
1
1.5
2
2.5
3
3.5
0
40
80
120
160

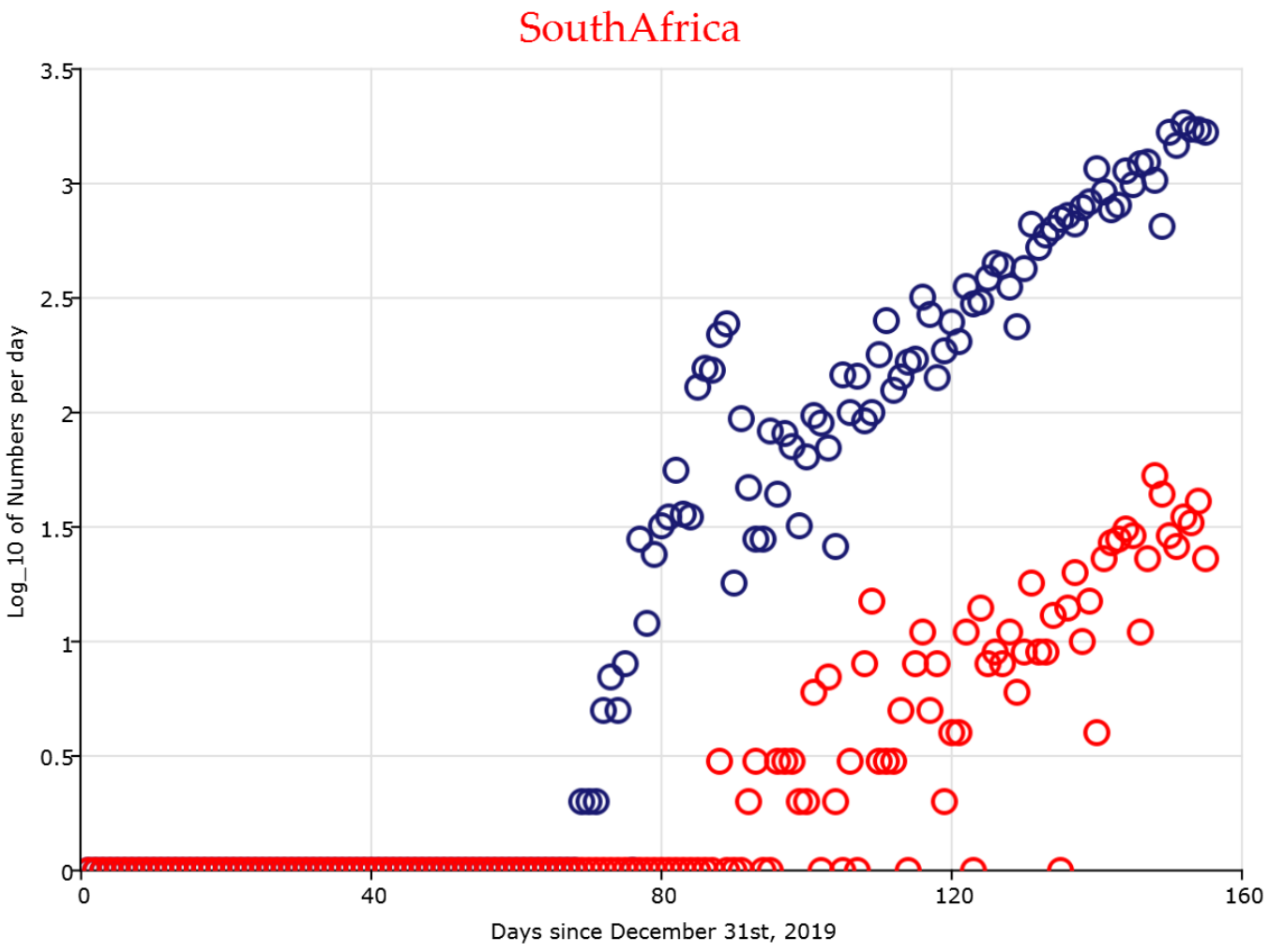

Singapore
Log_10 of Numbers per day
Days since December 31st, 2019
0
0.5
1
1.5
2
2.5
3
3.5
0
40
80
120
160

Early Stages




Egypt
3.5
3
2.5
2
1.5
1
0.5
0
Log_10 of Numbers per day
0
40
80
120
160
Days since December 31st, 2019


Peru
Log_10 of Numbers per day
Days since December 31st, 2019
0
1
2
3
4
0
40
80
120
160


SouthAfrica
Log_10 of Numbers per day
Days since December 31st, 2019
0
0.5
1
1.5
2
2.5
3
3.5
0
40
80
120
160

Conclusions

- The UK and most European countries are past their peaks and close to the endgame. Many countries elsewhere still in early stages and predicted to have a very large number of cases (e.g., Brazil, India)
- One day ahead forecasting record of our model has been quite good, discounting the problems arising from late and incorrect reporting of data
- Longer term forecasting has proved more challenging. Consistently underestimated the likely total number of cases and deaths for the UK and US.

- Some specific findings
  - Shapes of curves appear to be quite different across countries. New Zealand versus USA; Japan versus Italy
  - Timing of peak deaths in some countries precedes the peak of cases, which seems to be against the epidemiological models. This may be because testing capacity has increased a lot and treatment has improved.
  - Characteristics of endgame countries varies a lot. Many small islands, but also Luxembourg, China, Israel, Switzerland
- Differences across countries in terms of
  - When the peaks occurred. Globalization network
  - Ratios of deaths to cases. Part is due to how much testing (USA), death definitions (Russia), but must be more than that (Singapore, Vietnam). Maybe behaviour and social norms.

Future work

- Build a multivariate model that takes account of the network connections between countries and the dynamic relationship between cases and deaths predicted by epidemiology
- Investigate some of the differences between countries in terms of the shape of the curve, the peak, the peak duration, the mortality rate etc in terms of cross-sectional predictors.

We regress total cases per million people until 2020-05-30 on following regressors: population density, GNP per capita and obesity rate, with adjusted $R^2 = 0.3917$.

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-556.9779	336.7298	-1.65	0.1004
popden	0.4673	0.2061	2.27	0.0249
GNP	0.0574	0.0082	6.97	0.0000
obesity	44.1062	17.4130	2.53	0.0124