Open Access

SCHOLARENA

Journal of Public Health and Disease Prevention

CASE REPORT

ISSN: 2641-8509

Coronavirus (COVID-19) Forecasting and Observed Frequencies

Charles G Minard*

Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, Texas, USA

***Corresponding author:** Charles G Minard, Institute for Clinical and Translational Research, Baylor College of Medicine, Houston, Texas, United States, Tel: 7137982353, E-mail: minard@bcm.edu

Citation: Charles G Minard (2020) Coronavirus (COVID-19) Forecasting and Observed Frequencies. J Public Health Dis Prev 3: 106

Abstract

The coronavirus (COVID-19) was recognized as a global pandemic in January 2020. Governmental and public response to the pandemic has varied by region. An important tool for influencing response is the ability to accurately forecast cases. However, slow response, poor data collection, and lack of testing have inhibited the collection of important information that can help improve model forecasts. Here, the frequency of new and cumulative cases and deaths attributable to the coronavirus are forecast over time. Weekly models, given the available data at the time of the model, are compared with future observed events. Forecasts are primarily made for the US, but comparisons with similar regions are also evaluated. Early models tend to markedly underestimate the number of cases and deaths. Accurate forecasts are not generally available until April 2020.

Keywords: Coronavirus; COVID-19; Forecast

Introduction

Wuhan, China first reported an outbreak of coronavirus (COVID-19) in December of 2019 and the outbreak was recognized as a global pandemic on January 30, 2020 [1]. The response to the pandemic has differed by region with varying success [2-4]. Forecasting the spread of the disease and impact on public health is important for healthcare planning, business operations, and communities [5-9]. However, poor data collection and delayed response to the pandemic have hindered accurate modeling and forecasts [10-12]. Here, the frequency of new and cumulative cases as well as deaths are forecast over time. Past forecasts are compared with the observed frequencies given the amount of information that was available at prior time points.

Methods

The primary purpose of this analysis is to forecast frequencies of new coronavirus cases and deaths as well as cumulative frequencies for the US and other regions over time. Novel coronavirus (COVID-19) cases data on the number of confirmed cases and deaths was downloaded from the United Nations Office for the Coordination of Humanitarian Affairs Human Data Exchange website on April 13, 2020 [13]. This public resource has collected epidemiologic data on coronavirus cases since January 22, 2020. Variables include region (i.e., country, province, state), observation date, and cumulative numbers of events (i.e., confirmed cases and deaths attributable to active coronavirus infections).

An auto-regressive integrated moving average (ARIMA) model is used to forecast future confirmed cases with 95% confidence intervals by region over time [14]. Estimates of new cases and deaths as well as cumulative frequencies are forecasted by region and time using an ARIMA (1,1,1) model. This model is a differenced first-order autoregressive model with smoothing using a first-order moving-average. It is equivalent to an ARMA (1,1) model. The model forecasts outcomes 30-days in the future for each week from March 8, 2020 through April 13, 2020. Series and line plots present the observed and forecasted estimates for comparison in the US and between regions over time. Models are developed for all regions in the dataset; however, this paper focuses attention on US data and comparisons with similar regions. SAS for Windows version 9.4 was used for all analyses [15].

Results

Figure 1a summaries the observed and forecasted estimates of new, confirmed coronavirus cases for the US by date. Early forecasts have narrow confidence intervals, but are poor estimates of the observed future data. The wide variation and inconsistent data observed through March 29, 2020 for the US results in extremely wide confidence intervals that range from 0 to over 125,000 new cases expected per day by the end of April 2020. April 5 and 12 forecasts begin to converge, but the expected number of new coronavirus cases continues to increase for the US. Figure 1b presents observed and forecasted estimates by time and region. Forecasted estimates for Italy and Sweden tend to agree with the observed frequencies of new cases across all weeks. Forecasts for the US show marked departures from the observed data until March 29, 2020 where excessive variation is evident.

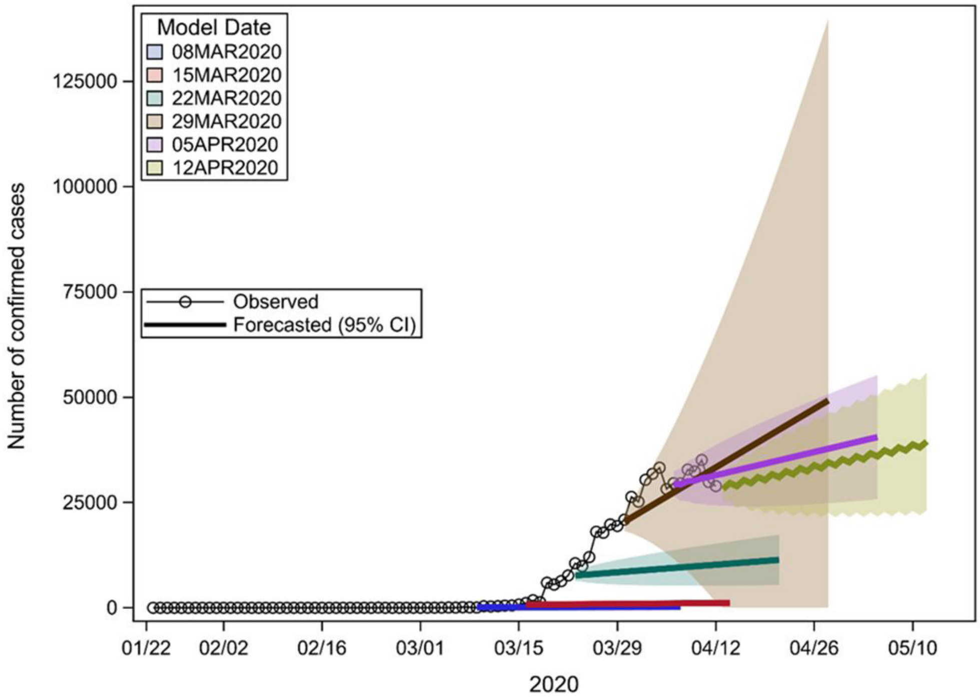

Figure 1a: Number of Confirmed Cases (US)

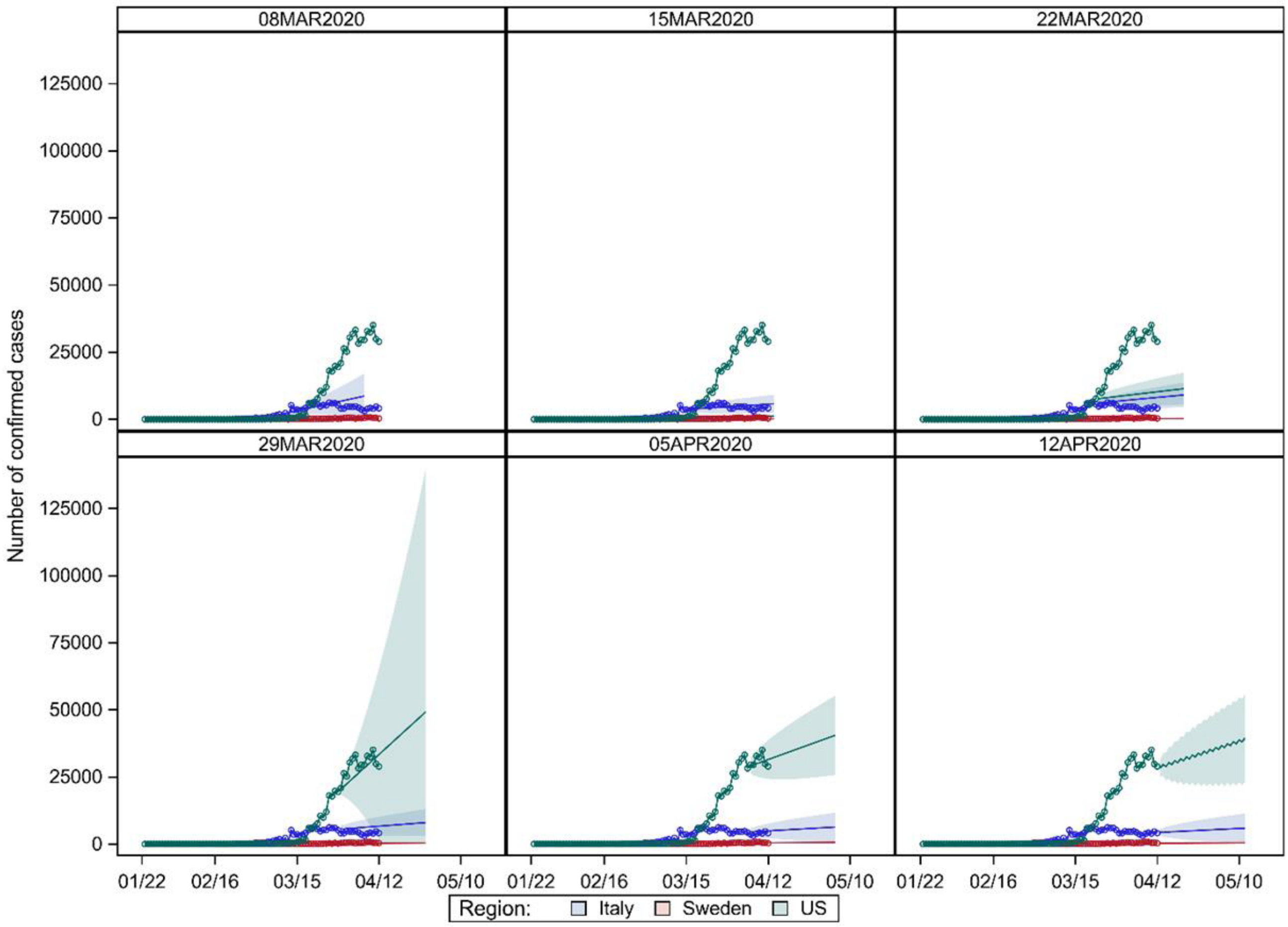

Figure 1b: Number of Confirmed Cases

Figures 2a & 2b present the observed and forecasted cumulative frequencies of confirmed coronavirus cases. Figure 2a shows that the cumulative number of new cases in the US continued to outpace the prediction models until about April 5, 2020. At this point, the observed cumulative frequency begins to fit the forecast model. The forecasts for the cumulative number of cases tends to underestimate the observed frequencies for all three countries in early models (Figure 2b). However, the largest difference observed is for the US. Forecasts for the US and Italy are virtually identical on March 22, 2020. However, Italian cases began to level off on this date while the number of cases continued to increase in the US.

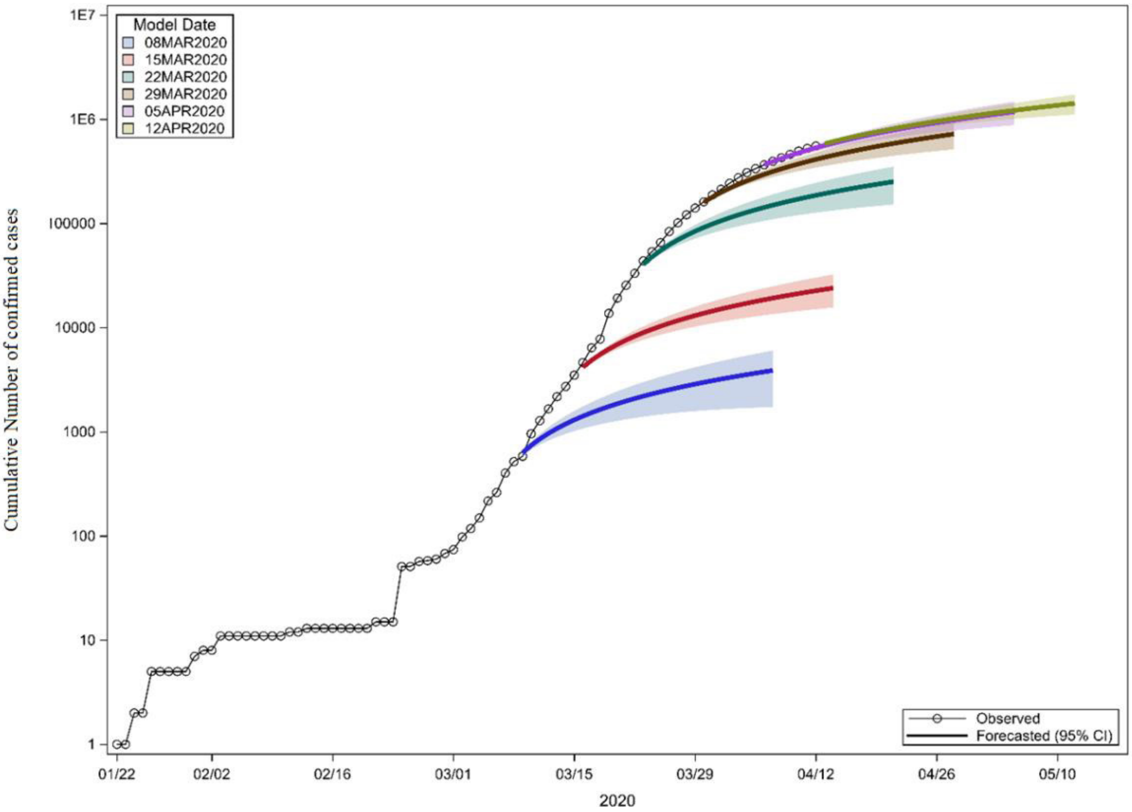

Figure 2a: Cumulative Number of Confirmed Cases (US)

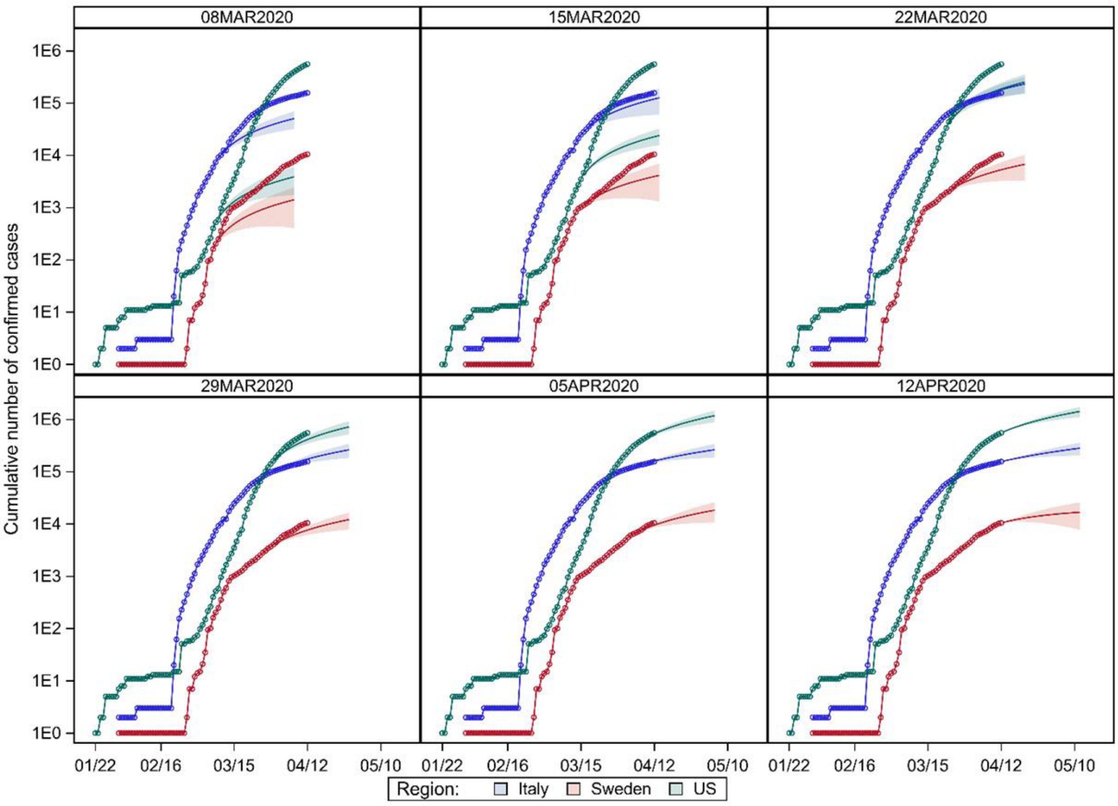

Figure 2b: Cumulative Number of Confirmed Cases

Figures 3a & 3b present observed and forecasted estimates of the frequencies of new coronavirus deaths by time and region. Figure 3a summaries the observed and forecasted estimates for the US by model date. Early forecasts have narrow confidence intervals, but are poor estimates of the observed future data. The wide variation and inconsistent data observed through March 29, 2020 for the US results in extremely wide confidence intervals that range from 0 to over 3,000 new deaths expected per day by the end of April 2020. Sharp differences in the forecasted number of deaths continued in more-recent forecasts, and the expected number of new deaths continues to increase for the US. Forecasts for the US show marked departures from the observed data until April 12, 2020.

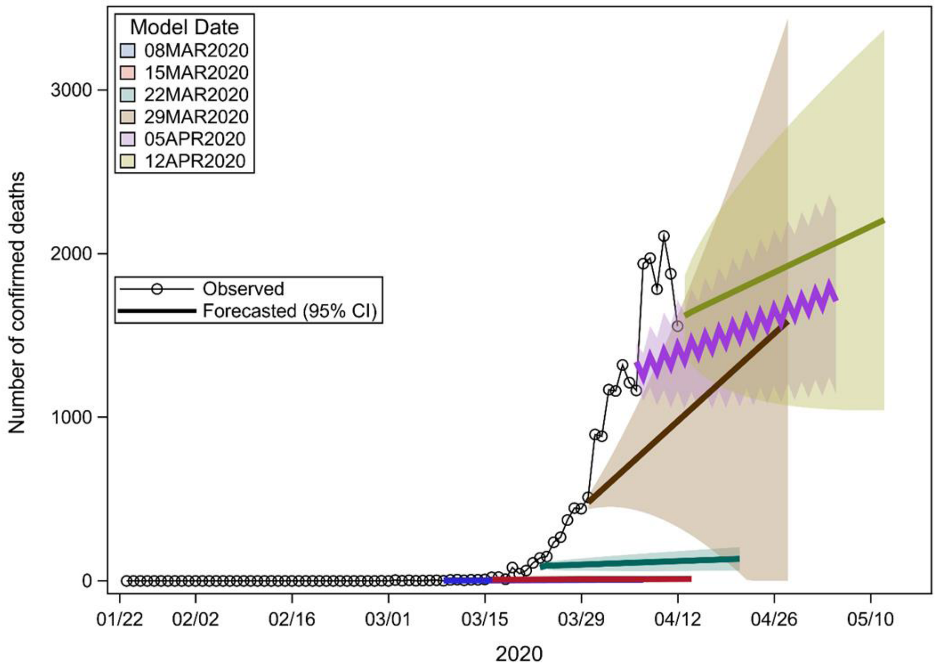

Figure 3a: Number of Confirmed Deaths (US)

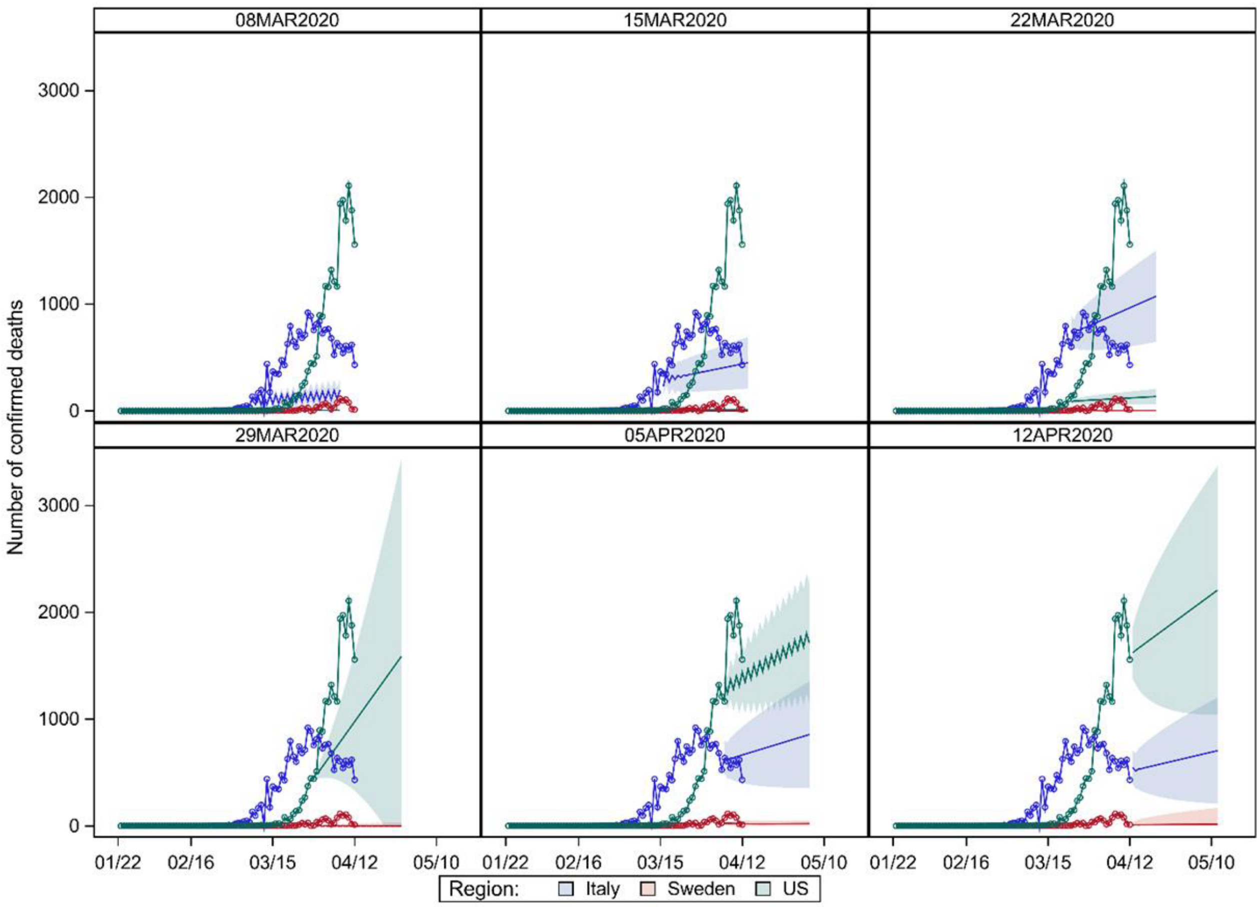

Figure 3b: Number of Confirmed Deaths

Figures 4a and 4b present the observed and forecasted cumulative frequencies of deaths. Cumulative deaths attributable to the coronavirus are markedly underestimated for the US (Figure 4a). Observed cumulative deaths are consistently underestimated through the April 5 model, However, cumulative deaths are poorly estimated for both the US and Sweden across all weeks (Figure 4b). Cumulative deaths in Italy are notably within forecasted estimated as early as March 22, 2020.

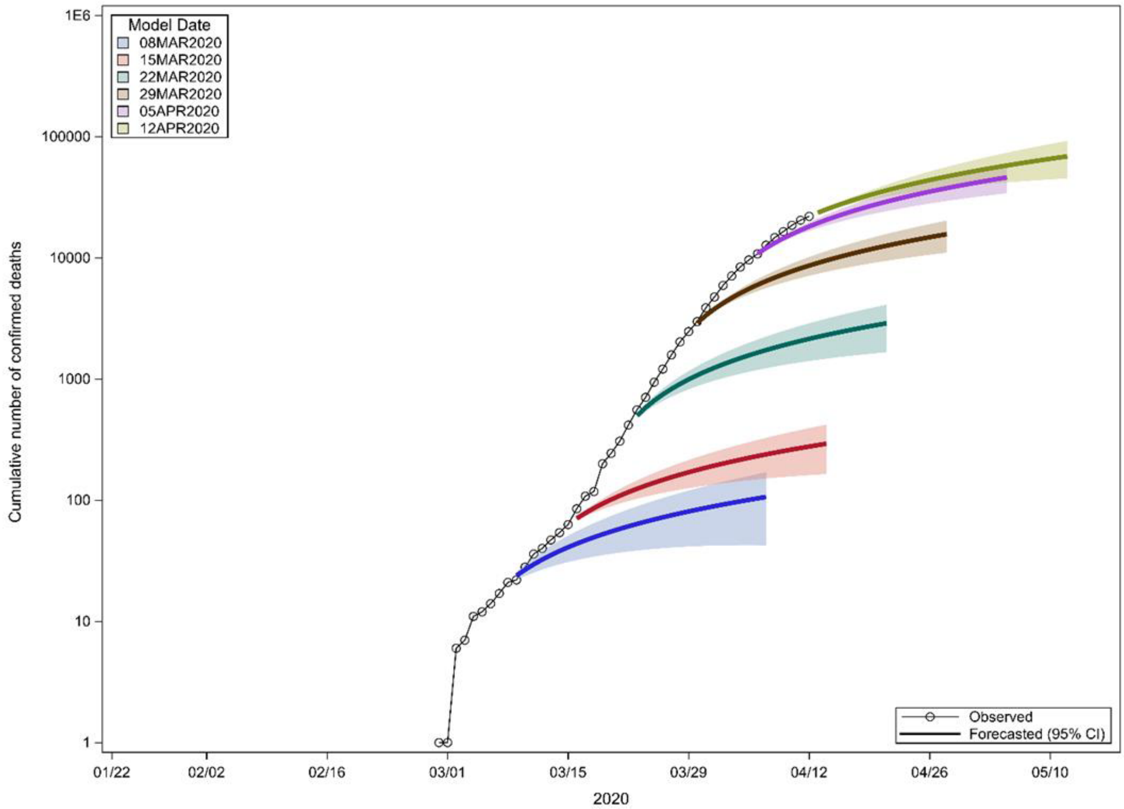

Figure 4a: Cumulative Number of Confirmed Deaths (US)

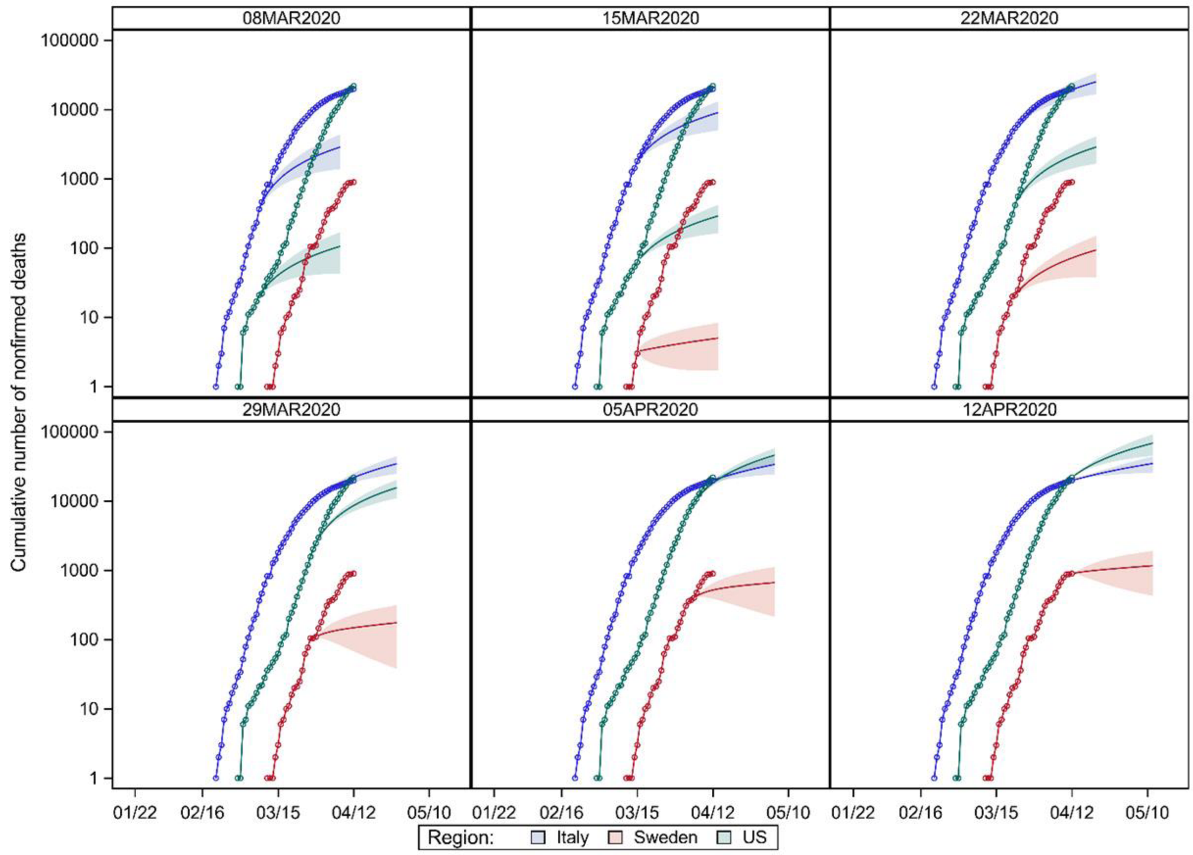

Figure 4b: Cumulative Number of Confirmed Deaths

Final ARIMA (1,1,1) parameter estimates for the US are presented in Table 1. Current forecasts, based on these models, predict about 33,847 new cases per day in the US (Figure 1) and over 1.42 (95% CI: 1.11, 1.73) million total confirmed cases by mid-May. Death rates are forecasted to be about 1915 deaths per day (Figure 3), and 68,82 (95% CI: 45,239, 92,408) cumulative deaths by mid-May.

	μ		MA(1,1)		AR(1,1)	
Scenario	**Estimate**	**P-value**	**Estimate**	**P-value**	**Estimate**	**P-value**
Confirmed cases	352.8	0.04	-0.94	0.001	-1	<0.001
Cumulative confirmed cases	2934.2	0.09	0.04	0.72	1	<0.001
Confirmed deaths	20.1	0.1	0.2	0.85	0.08	0.94
Cumulative confirmed deaths	100.6	0.42	0.01	0.94	1	<0.001

Table 1: Parameter estimates for ARIMA (1,1,1) models for the US

Discussion

The primary purpose of this analysis was to compare observed coronavirus cases and deaths with forecasts by model date and region. Early forecasting models tended to underestimate the number of new cases and deaths attributable to the coronavirus. Later models appear to be converging on estimates. However, the impact of early data continued to influence the model forecasts resulting in underestimates of observed frequencies.Timely and accurate data collection is difficult in any situation, and the global coronavirus pandemic is no exception [16]. Data quality is important for monitoring the pandemic and can impact patient health [17,18]. Unreliable data can bias forecasted results and influence models for extended periods [19]. The ARIMA models presented here demonstrate the impact on forecasting future events given the policies enacted for the coronavirus pandemic. Model forecasts were inconsistent with poor fits, and tended to underestimate observed frequencies. The effect of early observations on forecasted results continued for nearly 2-months. This study is limited by regional efforts to collect and report accurate data. Frequencies are likely to be underreported for a variety of reasons including a lack of screen test availability and mild symptoms for which patients recovered and did not seek testing [20].

Conclusion

Early data on coronavirus cases and deaths tend to provide poor forecasting results for the observed future events. Models may be biased from the inclusion of early data, and have difficulty correcting for this bias during later stages. The timely and accurate collection of data can improve model forecasts which may positively impact societal response and outcomes.

References

1. Chang R, Wang H, Zhang S, Wang Z, Dong Y, et al. (2020) Phase- and epidemic region-adjusted estimation of the number of coronavirus disease 2019 cases in China. Front Med pp. 1-11.

2. Coccolini F, Sartelli M, Kluger Y, Pikoulis E, Karamagioli E, et al. (2020) COVID-19 the showdown for mass casualty preparedness and management: the Cassandra Syndrome. World J Emerg Surg 15: 26.

3. Viner RM, Russell SJ, Croker H, Packer J, Ward J, et al. (2020) School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. Lancet Child Adolesc Health 4: 397-404.

4. Chatterjee P, Nagi N, Agarwal A, Das B, Banerjee S, et al. (2020) The 2019 novel coronavirus disease (COVID-19) pandemic: A review of the current evidence. Indian J Med Res.

5. Prachand VN, Milner R, Angelos P, Mitchell C Posner, John J Fung, et al. (2020) Medically-Necessary, Time-Sensitive Procedures: A Scoring System to Ethically and Efficiently Manage Resource Scarcity and Provider Risk During the COVID-19 Pandemic. J American College Surg.

6. Kim S, Kim YJ, Peck KR, Jung E (2020) School Opening Delay Effect on Transmission Dynamics of Coronavirus Disease 2019 in Korea: Based on Mathematical Modeling and Simulation Study. J Korean Med Sci 35: e143.

7. Chung CJ, Nazif TM, Wolbinski M, Hakemi E, Lebehn M, et al. (2020) The Restructuring of Structural Heart Disease Practice During The Covid-19 Pandemic. J Am Coll Cardiol.

8. Hilsenrath PE (2020) Ethics and Economic Growth in the Age of COVID 19: What Is a Just Society to Do? J Rural Health.

9. Qian X, Ren R, Wang Y, Guo Y, Fang J, et al. (2020) Fighting against the common enemy of COVID-19: a practice of building a community with a shared future for mankind. Infect Dis Poverty 9: 34.

10. Petropoulos F, Makridakis S (2020) Forecasting the novel coronavirus COVID-19. PLoS One 15: e0231236.

11. Anastassopoulou C, Russo L, Tsakris A, Siettos C (2020) Data-based analysis, modelling and forecasting of the COVID-19 outbreak. PLoS One 15: e0230405.

12. Wynants L, Van Calster B, Bonten MMJ, , Collins GS, Debray TPA, et al. (2020) Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 369: m1328.

13. Novel Coronavirus (COVID-19) Cases Data (2020) United Nations Office for the Coordination of Humanitarian Affairs Human Data Exchange.

14. Helfenstein U (1996) Box-Jenkins modelling in medical research. Stat Methods Med Res 5: 3-22.

15. SAS for Windows [computer program]. Version 9.4 TS Level 1M3.

16. Rojas Vallejos J (2020) Strengths and limitations of mathematical models in pandemicsthe case of COVID-19 in Chile. Medwave 20: e7876.

17. de Lusignan S, Williams J (2020) To monitor the COVID-19 pandemic we need better quality primary care data. BJGP Open pp. 1-3.

18. Matrajt L, Leung T (2020) Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease. Emerg Infect Dis 26(8).

19. Panovska Griffiths J (2020) Can mathematical modelling solve the current Covid-19 crisis? BMC Public Health 20: 551.

20. Li R, Pei S, Chen B, Song Y, Zhang T, et al. (2020) Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). Sci 368: 489-493.