The Plight of COVID-19 in Ethiopia: Describing Pattern, Predicting Infections, Recoveries and Deaths Using Initial Values from Different Sources

Girma Taye [1*], Mengistu Yilma[1], Awgichew Kifle[1], Wondimu Ayele[1] and Ephrem Biruk[1]

Abstract

Background: On 31rd December 2019, China reported a cluster of cases of pneumonia of unknown etiology in Wuhan city, Hubei province. Eventually, a coronavirus was identified which was called "COVID-19" by World Health Organization (WHO) and was declared as a Public Health Emergency Concern globally.

Experts suggested a country context evidence to reduce the impact of COVID-19 in Africa region. To this end, this study aimed to model the course of the outbreak towards understanding the spread of the disease and the effect of integrated intervention.

Methods: The SEIR and other relevant models were fitted to determine the effect of integrated intervention towards prevention and control of the virus. Comparative visualization of data was conducted to show the pattern and progress of the disease in Ethiopia in relation with other countries.

Results: The overall trend of the virus in Ethiopia showed linear increase since the first case on March 13, 2020, and exponential increase after May 24, 2020. The confirmed cases in Ethiopia reached 5034 within 67 days, while South Africa and Italy reached 22,556 and 205,425 respectively within 67 days after passing 100 cases.

The SEIR model considered integrated intervention measures (social distancing, facemask, and hand hygiene) with rho values of 0.7 and 0.5. Without intervention, about 9% of the population can be infected, while the proportion reduced to 5.5% and 2.5% with implementation of 30% and 50% integrated intervention measures, respectively.

The Prophet model showed prediction accuracy of 78.3% (95%CI = 74.2% – 82.3%) for confirmed cases.

Conclusion: Ethiopia showed the slow progress of COVID-19 compared with South Africa and Italy. The implementation of integrated measures could reduce the proportion of infection significantly. The integrated intervention measures could also extend the peak time to a longer period. The Prophet model showed promising prediction accuracy as it increases when the data increase. [*Ethiop. J. Health Dev.* 2021; 35(SI-1):82-89]

Key Words: COVID-19, patterns, predicting, infections, recovery and death

Introduction

On 31 December 2019, China reported a cluster of cases of pneumonia of unknown etiology in Wuhan city, Hubei province. Eventually, a coronavirus was identified which was called "COVID-19" by WHO and was declared as a Public Health Emergency Concern globally (1). The first confirmed COVID-19 cases in Africa was announced in Egypt on 14 February 2020 (1). The Federal Ministry of Health (FMoH) has confirmed a coronavirus disease (COVID-19) case in Addis Ababa on the 13th of March 2020(2).

During the end of March, when this study was initially started, the number of cases in Ethiopia was only 23. Although the increase in the number of cases was slow for most of May, the cases suddenly rose around the end of May and continued into June. By 7th of June, daily cases reached as high as 150 with a total of cases reaching 2020 with 27 deaths and 344 recoveries. By the end of July, however, the daily cases exceeded 700 with total cases climbing beyond 5000 (3). Since there were evidence of community-level transmissions this time around, the importance of the modelling study was at its peak.

Cleo Anastassopoulou et. al. (4) used early one-month data of newly confirmed cases from Hubei province and estimated parameters and predicted future new cases. Later, it was found that predictions made earlier were very close with actual data obtained at later times. Dalin Li, et.al.,(5) also used 20 new cases in the US on 29/02/20. The authors used this small data set to estimate parameters of interest and determine the fate of the infections. Zhihua Liu, et al. (6) predicted the number of cases based on early observed cases in China. These are just examples among the many researchers in different countries rushed to model the cases, however small the data they obtained may be.

Because of lack of appropriate data for the COVID-19 in the early phase of the outbreak, some (7, 8) used the serial intervals (SI) of the two known coronavirus diseases, MERS and SARS, as approximations for the unknown parameter.

Several authors (9-11) used mathematical modelling for their countries to shed light on important epidemiologic parameters that help researchers and officials understand the nature of the epidemic a little better.

It is little known about the rate of infection, hospitalization, and recovery; severity and incubation period of the COVID-19 at the time of this study. Therefore, it is important to provide further information on the plight of this virus to help planners and policymakers to observe the situation from different angles.

Methodology

We have conducted data visualization, modelling and prediction of COVID-19 using data from different publicly available sources. We used Python and Facebook Prophet Software for visualization, fitting the models and conduct prediction.

Data sources

Data were extracted from different sources such as GitHub web page, Information Network Security Agency (INSA) COVID-19 monitoring platform, Worldometer web page, and the FMOH and Ethiopian Public Health Institute (EPHI) report. We used data

[1]School of Public Health, College of health Sciences, Addis Ababa University, Addis Ababa, Ethiopia

from these public sources of daily time series confirmed, recovered, critical and death cases for the analysis.

Data visualization
The geographical distribution of cases by region was visualized using the latitude and longitude coordinates of the area. The comparative visualization of confirmed versus recovered cases were displayed by the region. The distribution of coronavirus by sex and age were displayed. The trend of cumulative cases was visualized to show how the cases raised since the first report of a single case in Ethiopia.

Modeling
Since the Susceptible-Infected-Removed (SIR) model (12), several other compartmental models and prediction tools have been developed or the original work extended. A thorough mathematical derivation using differential equations was provided in (13) to show how the model formulation took place. The SIR model is modified to the SIRD model to include death, with four groups of patients: *susceptible*, *infected*, *recovered*, and *died*. This model is also known as lethal (as opposed to the SIR model) since patients are assumed to die or recover once infected and those recovered re-infected. The SIR model was also extended to SIRS, S indicating Susceptible group.

The SEIR (Susceptible-Exposed-Infectious-Recovered) model is a four-compartment model containing S, E, I and R. It is assumed that those in the exposed group are infected but not infectious yet. Compared to the SIR model, this model extends the time in which patients are infectious, because one must be transferred from the 'Susceptible' compartment to 'Exposed" group before they become infectious.

Based on systemic review, basic reductive number and average incubation period were determined and used for model development, considering a different scenario in addition to the use of local data. The population of Ethiopia is assumed to be 114,963,588 (24,463,423 Urban and 90,500,165 rural).

Prophet is an open-source fully automatic software and available for installation in python. It was developed and released by the Facebook's Core Data Scientist team. Prophet is a robust software for predicting time series data based on the additive model by considering the pattern of the historical data and its seasonal and holiday effect(11). We used Prophet to forecast COVID-19 cases, recovered and deaths in Ethiopia of a five-days ahead with a 95% prediction confidence interval.

The input to Prophet model contains a data frame with two columns, namely, ***ds*** is a date stamp or date format, and ***y*** is the numeric measurement we want to forecast. The predicted value contains for columns that are ***ds, yhat, yhat-lower and yhat-upper***. The ***forecast object*** is a new data frame that includes a column ***yhat*** with the forecast, as well as columns for the lower and upper value of 95% prediction interval.

Ethical consideration
Ethical clearance was secured from Addis Ababa University, College of Health Sciences IRB. Secondary data was gathered from EPHI and Addis Ababa health bureau through permission letter from the School of Public Health. There was no involvement of individuals in this study; hence, ethical issue was very minimal.

Results
Visualization of corona virus outbreak in Ethiopia: The Ethiopia daily time series confirmed cases, recovered cases and deaths were collected from March 13 to June 24, 2020 from different sources (2, 3, 14). The largest majority of confirmed cases (73.4%) recovered cases (81.9%) and deaths (83.1%) were reported from Addis Ababa City administration. The total number of cases in all regions was 4971. The total number of active cases was 3468 from March 13 to June 24, 2020.

The geographical distribution of the disease was visualized using latitude and longitude coordinate of the regions where the size of the marks represented the number of cases of the region. The comparative visualization of total active cases, recovery and death were displayed using seaborn in python. The overall proportion of recovery and death among the total confirmed cases were 28.81% and 1.43% respectively up to June 24, 2020 (Figure1).

Figure1: **The Geographical distribution of coronavirus cases in Ethiopia by Region**

The overall trend of cases in Ethiopia showed linear increase since the first case on March 13 and exponential increase after May 24, 2020. The disease took 70 days to reach 582 cases from March 13 to May 24, 2020 but took only 8 days to be doubled. Ethiopia already crossed 5000 cases by June 24 and the increment will be beyond the capacity of the Country's health system if it is not contained on time (Figure 2).

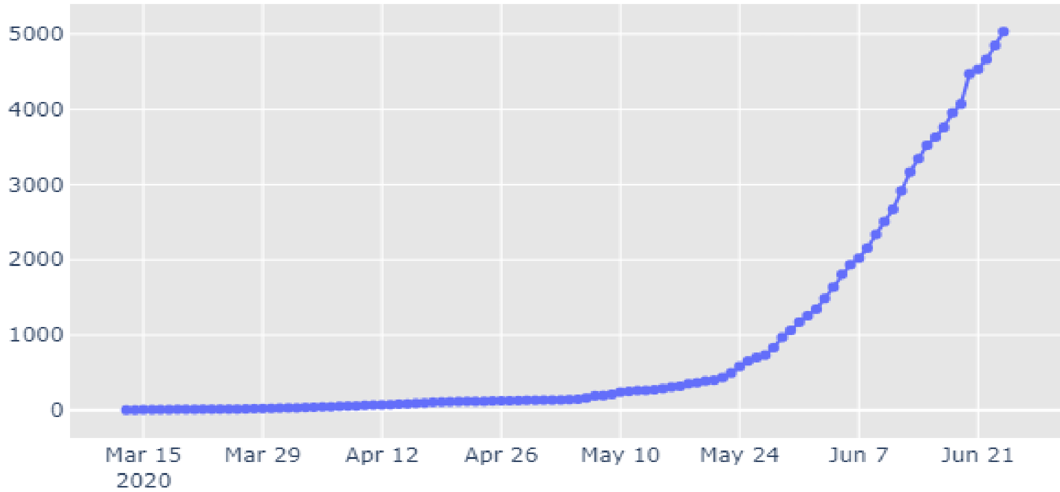

Figure 2: **The cumulative trend of coronavirus cases in Ethiopia from March 13 to June 24, 2020**

We compared the cumulative cases in Ethiopia with South Africa and Italy to show variation of cases from country to country. The rising cases in Ethiopia were slower in the first two months when compared with South Africa and Italy. However, the number of cases rose exponentially after the second month even though the rate of testing for coronavirus still varied from country to country (Figure 3).

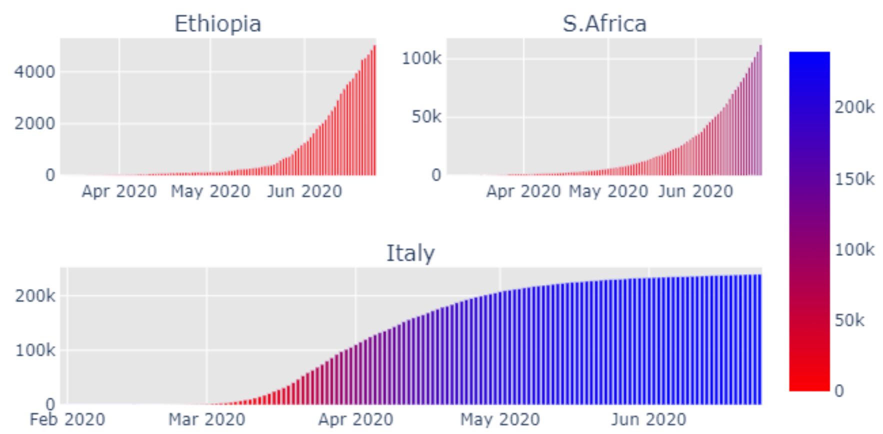

Figure 3: **Comparison of coronavirus cases rising in Ethiopia, South Africa, and Italy up to June 24, 2020.**

The trend showed that the number of confirmed cases in Ethiopia reached 5034 within 67 days, while South Africa and Italy reached 22,556 and 205,425 respectively within 67 days after passing 100 cases. This variation could be due to the difference in the number of testing of people for coronavirus or the difference in transmission rate from country to county. In Italy, cases took around two months to reach the peak and start to slow down, but for Ethiopia and South Africa cases continued to rise even after two months (Figure 4).

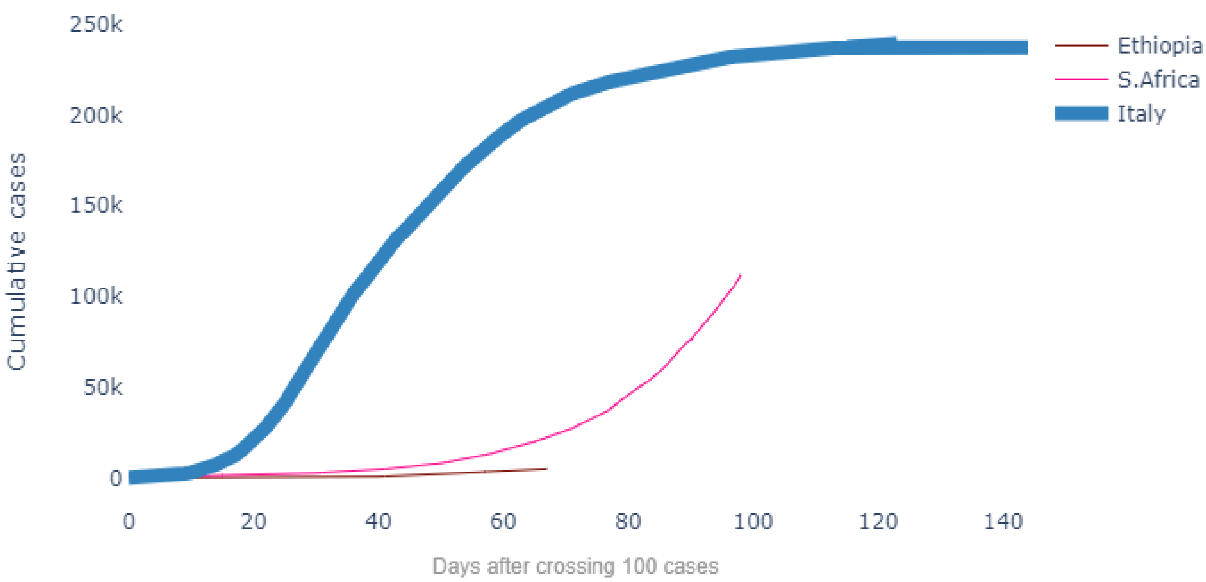

Figure 4: **The trend of coronavirus disease after crossing 100 cases in Ethiopia, S. Africa, and Italy**

SEIR Modeling

SEIR model was fitted using simulated parameters. Pooled value of basic reproduction number (R0) of 3.32 from a meta-analysis (13), average estimated incubation period of 5.52 days calculated from five different studies (15-19), and gamma value of 0.5 taken assuming that the mean infection period of 2 days(20) were used. Based on this information the parameter values are:

- α = 1/incubation period -> incubation period = 5.5 -> α = 0.18
- β = R0*γ -> 3.32*0.5 = 1.66
- γ = 0.5 (20)

Population size of 114,963,588 provided initial values as S_0, E_0, I_0, and R_0. We used these values into the python with codes using a 1-day time step as an initial model without applying the integrated intervention measures.

We introduced integrated intervention measures (social distancing, facemask, and hand hygiene) into the model as rho. This secondary model was developed assuming the rho values of 0.7 and 0.5, which means 30% and 50% of the population implemented the integrated intervention measures, respectively. The base case model without intervention (rho =1) showed more than 9% of the population can be infected while the infection reduced to about 5.5% and 2.5% with implementation of 30% and 50% integrated intervention measures, respectively. The time to reach the peak of the outbreak also pushed from 75 days (without intervention) to 125 and 225 days with application of the 30% and 50% integrated interventions, respectively (figure 5).

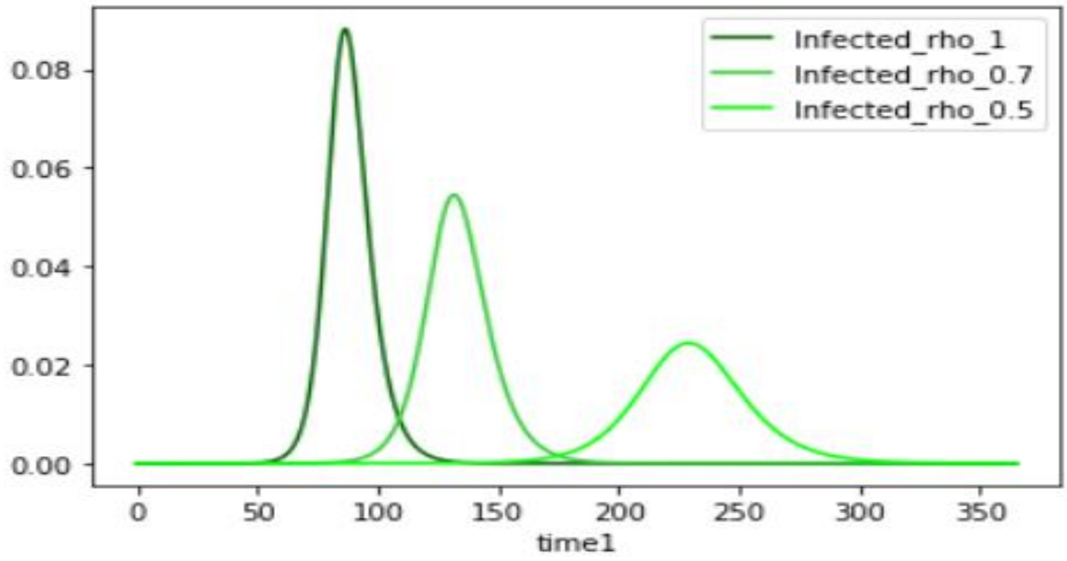

Figure 5: **COVID-19 SEIR Infection model with integrated intervention (α = 0.18, β =1.66 and γ = 0.5)**

The same principle was applied for modelling the exposure that the base case model without intervention (rho =1) showed about 25% of the population can be exposed for the disease while the exposure reduced to about 15% and 6.5% with implementation of 30% and 50% integrated intervention measures, respectively (Figure 6).

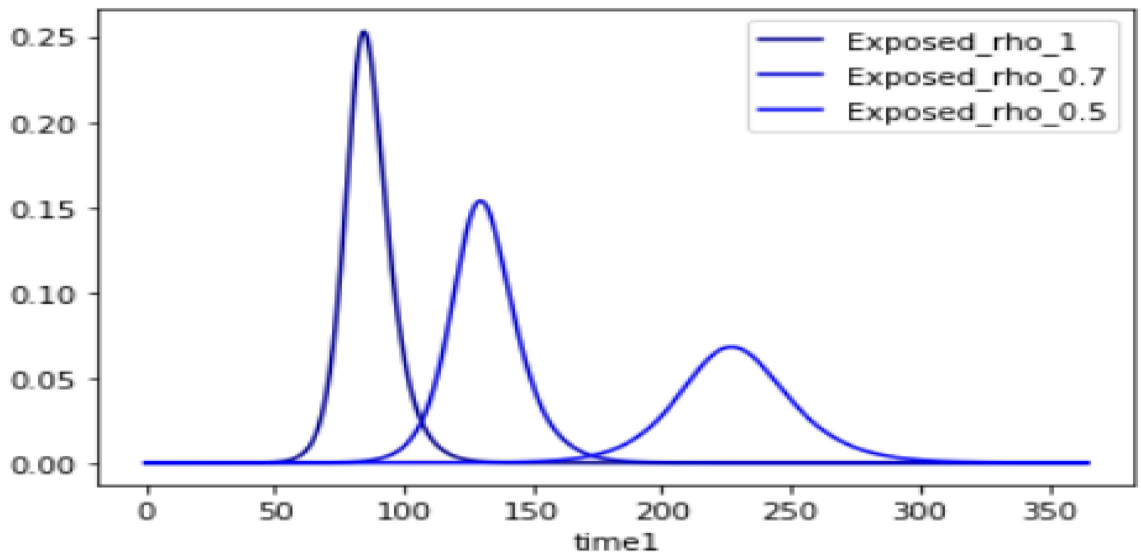

Figure 6: **COVID-19 SEIR Exposure model with integrated intervention (α = 0.18, β =1.66 and γ = 0.5)**

Prophet model output

We have projected 5 days ahead forecast of confirmed cases, recovered and deaths of corona virus in Ethiopia using the data from March 13 to June 5, 2020, with Facebook's Prophet software. This machine-learning software projected the future by understanding the pattern of the provided historical data with the principle of artificial intelligence. The prediction accuracy of the Prophet model of the 5 days with the 95% prediction interval from June 5 – 10, 2020 were 78.3% (95%CI = 74.2% – 82.3%) of confirmed cases, 68.8% (95%CI = 63.3% – 76.1%) of recovered cases and 48.5 % (95%CI = 40% -57.1%) of deaths. The prediction accuracy of the model reduced as the data size reduced in recovered and deaths since machine learning required large data by its nature. The model can be powerful when fitted with large data (Table 1).

The predicted values shown in the table below are results for the last (5th) day.

Table 1: **The 5 days forecasting outcomes of confirmed cases, recovered cases and deaths on June 10, 2020.**

SN	Prediction subjects	Actual	Predicted	Percent	95% CI
1	Confirmed cases	2506	1963	78.3	1859–2063
2	Recovered cases	401	276	68.8	250–305
3	Deaths	35	17	48.5	14–20

The graphical presentation showed how the predicted values fitted with the actual values with the 95% CI. There is an outlying confirmed case between 4th and 18th of May, which might have affected the model.

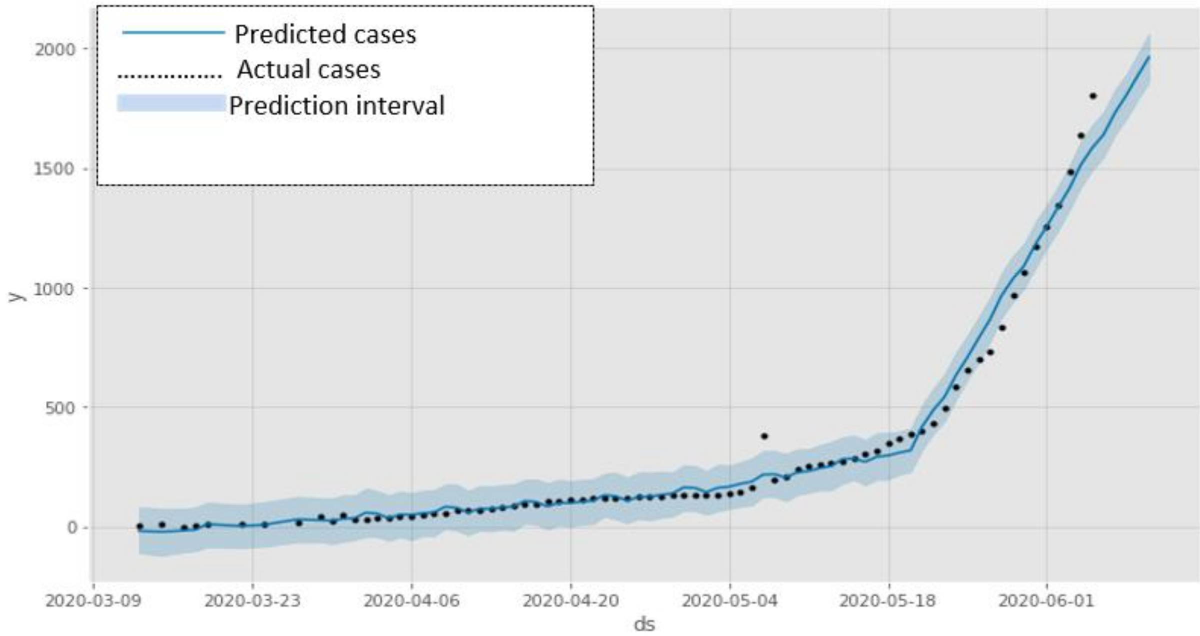

Figure 7: **COVID-19 confirmed cases forecasting with Prophet Software ahead of 5 days beyond June 5, 2020.**

Discussion

Although COVID-19 cases were detected around the same time in Ethiopia and South Africa, the amount and rate of increase differed among these countries very much from the beginning. This might be highly attributed to a combination of the capacity of testing and level of transmission of the disease in these countries. This can even be validated by the comparison between the countries in terms of patterns of infections after crossing 100 cases mark. Comparison between three countries showed that within 67 days after detection of the virus, cases in South Africa increased fourfold and cases in Italy increased by 14-fold compared to Ethiopia. Therefore, this evidence are simply indicators of the worst scenario that is just at our doorstep.

The models indicated that reducing the rate of transmission and the unreported rate of case would have a major impact on reducing the infection(21). Extension of SEIR showed that isolation, quarantine and treatment/monitoring influence the change in case rate and infection rate as well as the time to the peak (12).

The simple reproduction number (R0), mortality rate and recovery rate were estimated by a SIDR model in China (41) and SEIR modeling to predict COVID-19 outbreaks within and outside China (42). Models such as Gompertz, Logistic and Artificial Neural Networks were used to project the infection of COVID-19 until the end of the epidemic, (22).

The SEIR model can be fitted in different ways, using actual data on confirmed cases, recovered and deaths, or using overall parameters without using observed cases. We have followed the second approach for two reasons; the first is capacity of testing of the country was extremely low in the past and very low even now, and the second is there is no systematic way of isolating and testing suspected cases in a mass situation and that community level transmission is expected to have gone very far and deep in the society compared to available evidence about the spread of the virus. Therefore, using the current observed cases may not be representative of disease status, cannot reflect the country's actual situation, and may not produce trustworthy predictions. Assuming 50% and 30% successfully combined (facemask, social distancing, and hand hygiene) interventions, the models provided a hint that the number of days to a peak range from 125 to 225 days with 2.5% and 5.5% of the suspected cases (suspected cases assumed to be about 75% of the population of the country), respectively being a victim. This may not be an exact value on which planning may have to be depending but can serve as an indicator of where we are going.

The 95% confidence band for confirmed cases, compared to the other graphs included here, is narrower, showing better accuracy for prediction. This is due to the fact that the number of confirmed cases has increased recently, while the number of deaths was very low before June 24, forcing the confidence band to be wide and unstable indicating compromised prediction accuracy. Before June 2020, deaths were not encountered daily, and constant value was used in the model for days. For example, the same value of reported death was used between mid of April to mid of May. This forced the prediction line to highly oscillate and the 95% confidence band to be very wide, rendering prediction less reliable and less accurate.

Short time interval predictions, for five days, was made in this report due to the low number of confirmed cases, recoveries, and deaths by the time the report was written; but prediction interval can improve as the number of cases increase. This model can be used by authorities weekly since short term prediction is highly accurate.

Conclusion

The burden of COVID-19 prevailed in Addis Ababa City Administration than in other regions of the country. The disease was progressed slowly in the first two months while it increased exponentially after the second month. The trend of death also increased after the second month. People with the age group of 20 to 40 years and males were the most affected with coronavirus infection in Ethiopia. The rising pattern of the disease in Ethiopia was very slow when compared with that of South Africa and Italy, especially in the first two months. The trend of disease increment after crossing of 100 cases showed that Ethiopia will take far longer time to reach the peak compared with South Africa and Italy.

The SEIR model showed that implementing the integrated intervention measures like physical distancing, use of facemask and hand hygiene can reduce the contact rate and flatten the peak to lower level. The interventions can also push the time to reach the peak to longer time.

The Prophet model showed a promising forecasting outcome to project the future trend of the confirmed cases, recovered cases and deaths ahead of a certain time especially when we have larger data.

Recommendations

The FMoH and Addis Ababa City Administration should strengthen the integrated intervention measures to contain the spread of the disease in Addis Ababa and to limit its spread to other regions of the country. The government should use the instance of slow progress of the disease as an opportunity to be more prepared for the mitigation and equip the health sector's capacity.

Based on the evidence from the SEIR model, the government needs to strengthen the enforcement of implementation of integrated intervention measures. The health service providers, volunteers and media should increase the communities' awareness towards applying the integrated measures to prevent the disease. The health service providers and decision makers can use the Prophet model forecasting outcome for planning ahead of a certain time, be it short or long, with updating the data to the latest date.

Limitation of the study

Some of the models applied performed poorly due to the small size of data such as death by June 24, 2020.

Acknowledgment

This work was supported by Addis Ababa University, Research and Technology transfer office and the Ministry of Health, Ethiopia through CBMP project.

References

1. World Health Organization, Novel coronavirus(2019-nCoV). Situation report 21. Geneva, Switzerland: World Health Organization; 2020; 2020. Available from:https://www.who.int/docs/default-source/coronaviruse/ situation-reports/20200210-sitrep-21-ncov.pdf?sfvrsn=947679ef_2.
2. AAU. Ethiopian Covid_19 Data ecosystem 2020 [Available from: https://etcovdataecosystem.aau.edu.et/.
3. Worldometer. Worldometer Ethiopia coronavirus 2020 [Available from: https://www.worldometers.info/coronavirus/country/ethiopia/.
4. Anastassopoulou C, Russo L, Tsakris A, Siettos C. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. Plos One. 2020;13(5).
5. Dalin L, Jun L, Gregory B, Jonathan B, Weihua C, Liming L. Estimating the scale of COVID-19 Epidemic in the United States: Simulations Based on Air Traffic directly from Wuhan, China. 2020; Available from: doi: https://doi.org/10.1101/2020.03.06.20031880.
6. Liu Z, Magal P, Seydi O. Predicting the cumulative number of cases for the COVID-19 epidemic in China from early data. Quantitative Biology. 2020; Available from: https://arxiv.org/abs/2002.12298v1.
7. Zhang Z, Yuan M, Wenbo D, Zhaofen L, Dechang C. Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis. Int J Infect Dis. 2020.
8. Kucharski A, WRussell T, Diamond C, Liu Y, Edmunds J, Funk S. Early dynamics of transmission and control of COVID-19: a mathematical modelling study,. 2020; Available from: https://doi.org/10.1016/S1473-3099(20)30144-4.
9. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet. 2020;395:507-13.
10. Worldometer. Coronavirus (COVID-19) Mortality Rate 2020 [Available from:

https://www.worldometers.info/coronavirus/coronavirus-death-rate/.
11. Covid_19 Pandemic in India 2020 [Available from: https://www.edureka.co/.
12. Alexandre P, Cristina L. Mathematical Modeling and Epidemic Prediction of COVID-19 of the State of São Paulo, Brazil. IJAERS. 2020;5(7).
13. Alimohamadi Y, Taghdir M, Sepandi M. The Estimate of the Basic Reproduction Number for Novel Coronavirus disease (COVID-19): A Systematic Review and Meta-Analysis. 2020.
14. INSA. Ethiopia COVID-19 Monitoring Platform 2020 [Available from: https://www.covid19.et/covid-19
15. Guan W, Ni Z, Hu Y, Liang W, Ou C, He J, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. engl j med. 2020;382(18).
16. Kong T-k. longer incubation period of coronavirus disease 2019 (COVID-19) in older adults. Aging Medicine 2020;2020(00):1-8.
17. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann Intern Med. 2020.
18. Nie X, Fan L, Mu G, Tan Q, Wang M, Xie Y, et al. Epidemiological Characteristics and Incubation Period of 7015 Confirmed Cases With Coronavirus Disease 2019 Outside Hubei Province in China. The Journal of Infectious Diseases. 2020;2020(222):26-33.
19. Ye Q, Wang B, Mao J, Fu J, Shang S, Shu Q, et al. Epidemiological analysis of COVID-19 and practical experience from China. J Med Virol. 2020:1-15.
20. Hubbs C. Social Distancing to Slow the Coronavirus 2020 [Available from: https://towardsdatascience.com/social-distancing-to-slow-the-coronavirus-768292f04296.
21. Chen S, Li Q, Gao S, Kang Y, Shi X. Mitigating COVID-19 outbreak via high testing capacity and strong transmission-intervention in the United States. 2020.
22. Torrealba-Rodriguez O, Conde-Gutiérrez RA, Hernández-Javier AL. Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models. Chaos, Solitons and Fractals 138 (2020) 109946. 2020;138(2020):109946.