Research Article

Prediction Model for COVID-19 Patient Visits in the Ambulatory Setting

Li RC[1]*, Harrison CK[4], Jurkovitz CT[1], Ndura K[1], Papas MA[1], Kerzner RK[2], Teal C[3] and Chiam T[5]

[1]Institute for Research on Equity and Community Health (iREACH), ChristianaCare Health Services, USA

[2]Medical Group, ChristianaCare Health Services, USA

[3]Family and Community Medicine, ChristianaCare Health Services, USA

[4]Connecticut Children's and UConn School of Medicine, USA

[5]Infection Prevention & Control, Children's Hospital of Philadelphia, USA

*Corresponding author: Li RC, Institute for Research on Equity and Community Health (iREACH), ChristianaCare Health Services, Avenue North, 4000 Nexus Drive, Wilmington, DE 19803, USA

Abstract

Objective: Healthcare systems globally were shocked by coronavirus disease 2019 (COVID-19). Policies put in place to curb the tide of the pandemic resulted in a decrease of patient volumes throughout the ambulatory system. The future implications of COVID-19 in healthcare are still unknown, specifically the continued impact on the ambulatory landscape. The primary objective of this study is to accurately forecast the number of COVID-19 and non-COVID-19 weekly visits in primary care practices.

Materials and Methods: This retrospective study was conducted in a single health system in Delaware. All patients' records were abstracted from our electronic health records system (EHR) from January 1, 2019 to July 25, 2020. Patient demographics and comorbidities were compared using t-tests, Chi square, and Mann Whitney U analyses as appropriate. ARIMA time series models were developed to provide an 8-week future forecast for two ambulatory practices (AmbP) and compare it to a naïve moving average approach.

Results: Among the 271,530 patients considered during this study period, 4,195 patients (1.5%) were identified as COVID-19 patients. The best fitting ARIMA models for the two AmbP are as follows: AmbP1 COVID-19+ ARIMAX(4,0,1), AmbP1 nonCOVID-19 ARIMA(2,0,1), AmbP2 COVID-19+ ARIMAX(1,1,1), and AmbP2 nonCOVID-19 ARIMA(1,0,0).

Discussion and Conclusion: Accurately predicting future patient volumes in the ambulatory setting is essential for resource planning and developing safety guidelines. Our findings show that a time series model that accounts for the number of positive COVID-19 patients delivers better performance than a moving average approach for predicting weekly ambulatory patient volumes in a short-term period.

Keywords: Ambulatory; COVID-19; ARIMA; Time series analysis; Family medicine

Introduction

Healthcare systems were globally shocked by a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and the resulting disease, coronavirus disease 2019 (COVID-19) [1-3]. On March 11, 2020, COVID-19 was declared a global pandemic by the World Health Organization (WHO) [4]. By mid-March, transmission of COVID-19 had rapidly accelerated, increasing case counts throughout the United States, and it was found that many patients with severe disease also had common comorbidities such as hypertension, obesity and diabetes [5,6]. In the state of Delaware, the first presumptive positive case of COVID-19 was reported by the Delaware Division of Public Health on March 11, 2020 [7]. In order to mitigate the spread of the virus, the Governor of Delaware declared a state of emergency on March 13, 2020. The weeks that followed included several modifications to the original state of emergency to minimize the spread of the virus.

In response to the growing pandemic, ChristianaCare Health Services, Inc. (ChristianaCare), which serves the majority catchment area of Northern Delaware and the most populous county in the state, followed suit with its own measures to mitigate spread, postponing all elective procedures in hospitals and all ambulatory practices effective March 17, 2020 to adhere to state and CDC guidelines. The ambulatory services at ChristianaCare adjusted the delivery of healthcare services by reducing the number of in-person visits to minimize the risk to patients and healthcare providers, redirecting patients to telehealth when appropriate. By April 2020, more than 80% of ambulatory visits were telehealth visits (Phone or video). This proportion rapidly decreased the following months to approximately 35% in September 2020. Primary care practices were screening patients according to CDC guidelines to determine eligibility for in person versus virtual visits [8]. The majority of patients who were suspected of Covid had telehealth visits. Although our proportion of Telehealth visits are somewhat larger, the trend is similar to the trend described by the Vizient organization from a cohort of 39 large organizations including ChristianaCare [9]. This resulted in a decrease of patient volumes throughout the ambulatory system. With the uncertainty that COVID-19 presented then, the Phase 1 reopening that occurred on June 1, 2020, and the rise in cases occurring, it is essential to understand how the ambulatory setting will continue to be affected in order to develop proper guidelines.

To understand the impact of the novel virus, scientists rely on community spread models to predict possible transmission. The popular susceptible, infected, and recovered (SIR) epidemiologic model and variations of this model have been used to gauge community spread of a variety of infectious diseases such as influenza and dengue fever [10-14]. SIR models have also been applied to inpatient settings to predict hospital capacity regarding admissions, ICU beds, and ventilators [11,13,15,16]. In addition to SIR models, the current literature on predicting patient volume varies from descriptive statistics to Discrete-event Simulation (DES), Markov modeling, and advanced time series models, with most of the studies that have used time series forecasting models focusing on emergency department and hospital admissions [17].

Time series forecasting in ambulatory visits prior to the COVID19 pandemic have been described in a few reports but other types of modeling for both in-person and telehealth visits are lacking [18-25]. The most used method for time series forecasting is the Box-Jenkins method otherwise known as the AutoRegressive Integrated Moving Average (ARIMA) model [26]. The ARIMA model has been used for its simplicity and flexibility in capturing linear patterns in a time series [17,19-22].

Significance

The future implications of COVID-19 in healthcare are still unknown, specifically how it will continue to affect the ambulatory landscape. This work aims to inform COVID-19 and nonCOVID-19 ambulatory resources allocation as well as guide ambulatory practices for in-person visits as in-person care might have been delayed. Integrated health systems, such as ours, could benefit from having insights into both ambulatory and inpatient predictions to optimize resources throughout the health system. We propose an ARIMA time series model to capture the changes in ambulatory patient volumes as a result of COVID-19.

Objective

The primary objective of this study is to accurately forecast the number of COVID-19 and nonCOVID-19 weekly visits in primary care practices for both Telehealth and in-person visits. The ability to forecast patient volumes in primary care locations by accurately evaluating the dynamic changes in patient visits and fitting these data to a statistical model is useful for the appropriate allocation of human and material resources for future planning. With the uncertainty that COVID-19 presents, healthcare systems have been adapting their ambulatory practices to adhere to state guidelines. Therefore, we developed a time series model that provides an 8-week future forecast for ambulatory practices and compared it to a moving average approach.

Materials and Methods

Study design

This retrospective study was conducted in primary care practices that are part of a single integrated health care system in Delaware (ChristianaCare), serving the primary catchment area of New Castle County. New Castle County is in the northernmost region of Delaware and as of 2019 has an estimated population of 558,753, accounting for nearly two-third of the entire state population [27]. We selected the patients' records from the two practices that had the highest historical patient volumes among all clinics affiliated with ChristianaCare. Although the two practices were conducting both Telehealth and in-person, due to the small number of visits we combined them to obtain a total overall weekly visit count. Our study population included (1) COVID-19 patients who had prior family medicine ambulatory services within ChristianaCare in 2019 and had been previously hospitalized and discharged for COVID-19 or COVID-19 positive who were self-monitoring at home and had not been hospitalized. (2) Any patient who utilized ambulatory services from the same practices during the same time period and were not diagnosed with COVID-19. COVID-19 patients currently hospitalized were excluded from the population.

We extracted all patients' records from our electronic health records system (EHR) from January 1, 2019 to July 25, 2020 and built two datasets. One included patient-level data (e.g. age, gender, race, ethnicity, insurance, marital status, and Elixhauser comorbidities) and the other ambulatory practice-related data (e.g. encounter location, encounter providers, and weekly patient volumes) [28]. Patient-level data were used for characterizing the study population and were not included in the forecasting model. Ambulatory practice-related data were primarily used for our time series models. For model development, we used one year of data between January 2019 and December 2019, and for model validation data from January 2020 until July 2020.

Statistical and forecasting methods

Descriptive statistics: Patient demographics and comorbidities were compared using t-tests, Chi square, Mann Whitney U analyses as appropriate according to the distribution.

Time-series: A time series is a sequential set of data points, measured typically over successive times. The ARIMA model was created for auto-correlated and non-stationary time series data [11]. The framework for ARIMA is displayed in Table 1.

The forecasting method used is a non-seasonal ARIMA and ARIMA with exogenous variables (ARIMAX) to predict COVID-19 and nonCOVID-19 weekly patient volumes for Ambulatory Practice 1 (AmbP1) and Ambulatory Practice 2 (AmbP2). The COVID-19 models included an exogenous variable, the weekly number of positive COVID-19 patients, and were significant at p-value <0.05. When incorporating the exogenous variable to the nonCOVID-19 models, we found them to be insignificant. The complete dataset was split, 75:25 for training and validation sets. The validity of our models was evaluated using the difference between the forecasted patient volume and the actual patient volume beyond the period on which the model was trained. For each model out-of-sample forecast errors were assessed by calculating the Root Mean Square Error (RMSE) and mean absolute percentage errors (MAPE) as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \qquad (1)$$

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y_i}}{y_i}\right|\right) * 100 \qquad (2)$$

Where n represents the total data points, y_i represents the observed values at time i and $\hat{Y_i}$ represents the forecasted value at time i. Once we evaluated the models performance by computing the

RMSE and MAPE, the *(p,d,q)* parameters were used to forecast the next 8-week patient volumes for both populations in each location.

The model selection was performed using the Python 3.8.3 software package *statsmodels*.

Results

Among the 271,530 patients considered during this study period, 4,195 patients (1.5%) were identified as COVID-19 patients. The COVID-19 patients were younger, and majority was non-white compared to the COVID-19 negative patients. Both COVID-19 positive and negative patients had multiple comorbidities such as obesity, diabetes and hypertension despite the COVID-19 patients being younger. Details of the patient demographics and comorbidities for the study population are provided in Tables 2 and 3.

Figure 1 and 2 present the results of our ARIMA models for COVID-19 and nonCOVID-19 weekly patient volumes for both AmbP1 and AmbP2 and their 95% confidence interval. The best ARIMA models shown in Figures 1 and 2 are an aggregate of the weekly models that were generated during May 11-July 27, 2020. Table 4 represent the average RMSE and MAPE for all the models generated. The ARIMA models had lower RMSE and MAPE apart from the AmbP1 COVID-19 moving average model, which had a slightly lower MAPE of 23.21 compared to 24.86.

Discussion

The novelty of the COVID-19 virus created a conundrum, not only for the inpatient world, but for the ambulatory outpatient clinical environment as well. While many people were being hospitalized due to complications from the virus, an overwhelming majority were being evaluated and treated in the outpatient setting, either by urgent care or by their primary care provider. At the height of the pandemic, prediction models only existed for hospitals to anticipate the need for staffing, personal protective equipment (PPE), equipment and other resources as the cases surged, which made it very challenging to anticipate staffing, equipment and logistical needs for primary care practices. There were many questions/scenarios to consider such as designating specific practices to care for COVID-19 infected patients, estimating the number of staffing and PPE necessary at each practice site for in person care *vs.* delivering telehealth care; redeployment of staff to our Ambulatory COVID-19 treatment center (in person care for non-emergent patients with COVID-19) and to our Virtual COVID-19 primary care practice, which monitored moderately ill patients infected by the disease via video visits and secure texting, based on ambulatory patient volumes. Development of the

Table 1: ARIMA framework.

1	Visualize the data as a time series.
2	Test time series data for stationarity (e.g. Augmented Dickey Fuller Test); if non-stationary, transform the data.
3	Estimate model parameters *(p,d,q)* (e.g. Autocorrelation Function (ACF), and Partial Autocorrelation Function (PACF).
4	Identify best model parameters using fit criteria (e.g. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)).
5	Apply diagnostic tools to determine model fitness (e.g. Plot of standardized residuals, Histogram plus estimated density, Normal Q-Q plot, and Correlogram plot).
6	Forecast *n* weeks using an independent dataset.
7	Evaluate the accuracy of the forecast using error statistics (e.g. RMSE and MAPE).

Table 2: Study Population Characteristics.

	nonCOVID-19 (n=267,335)	COVID-19 (n=4,195)	p value
Age (mean, SD)	48.8 (22.7)	44.9 (19.0)	<0.01
Age category, n (%)			<0.01
<18	31515 (11.8)	179 (4.3)	
18-<45	75394 (28.2)	1970 (47.0)	
45-<65	84435 (31.6)	1394 (33.2)	
≥65	75991 (28.4)	652 (15.5)	
Male, n (%)	104677 (39.2)	1856 (44.2)	<0.01
Race, n (%)			<0.01
White	185300 (69.3)	1634 (39.0)	
African American	58965 (22.0)	1477 (35.2)	
Asian	9648 (3.6)	81 (1.9)	
Other	8186 (3.1)	564 (13.4)	
Missing	5236 (2.0)	439 (10.5)	
Hispanic, n (%)	13527 (5.1)	896 (21.4)	<0.01
Insurance, n (%)			<0.01
Commercial	150097 (56.2)	2189 (52.2)	
Medicaid	32835 (12.3)	640 (15.3)	
Medicare	71910 (26.9)	645 (15.4)	
Self-Pay	12493 (4.6)	9 (0.2)	
Missing	0 (0.0)	712 (16.9)	
Married, n (%)	124563 (46.6)	1593 (38.0)	<0.01
Have any Comorbidities, n (%)	260920 (97.6)	3666 (87.4)	<0.01

Table 3: Study Population Comorbidities.

	nonCOVID-19 (n=267,335)	COVID-19 (n=4,195)	p value
Comorbidity Count (median, IQR)	2.0 (0.0 - 4.0)	2.0 (0.0 - 4.0)	<0.01
Hypertension, n (%)	104331 (39.0)	1363 (32.5)	<0.01
Congestive heart failure, n (%)	15790 (5.9)	236 (5.6)	<0.01
Diabetes, n (%)	42194 (15.8)	741 (17.7)	<0.01
Liver Disease, n (%)	19082 (7.1)	253 (6.0)	<0.01
Renal Failure, n (%)	16264 (6.1)	287 (6.8)	<0.01
Chronic Lung, n (%)	60898 (22.8)	804 (19.2)	<0.01
Depression, n (%)	62104 (23.2)	748 (17.8)	<0.01
Obesity, n (%)	70752 (26.4)	1022 (24.4)	<0.01
Coronary Heart Disease, n (%)	28210 (10.6)	355 (8.5)	<0.01
Cardiac Arrhythmia, n (%)	50207 (18.8)	763 (18.2)	<0.01

ambulatory COVID-19 model provided the opportunity to identify volume trends and anticipate the need to modify our care delivery models based on the estimates.

We found that the ARIMA/ARIMAX forecasting models considered in this study were able to capture the temporal changes in weekly visits when compared to the moving approach. Although a MAPE of <10% is considered an accurate forecast, the COVID-19 ARIMA models we generated provided a more dynamic prediction than the moving average forecast. Our model included the number
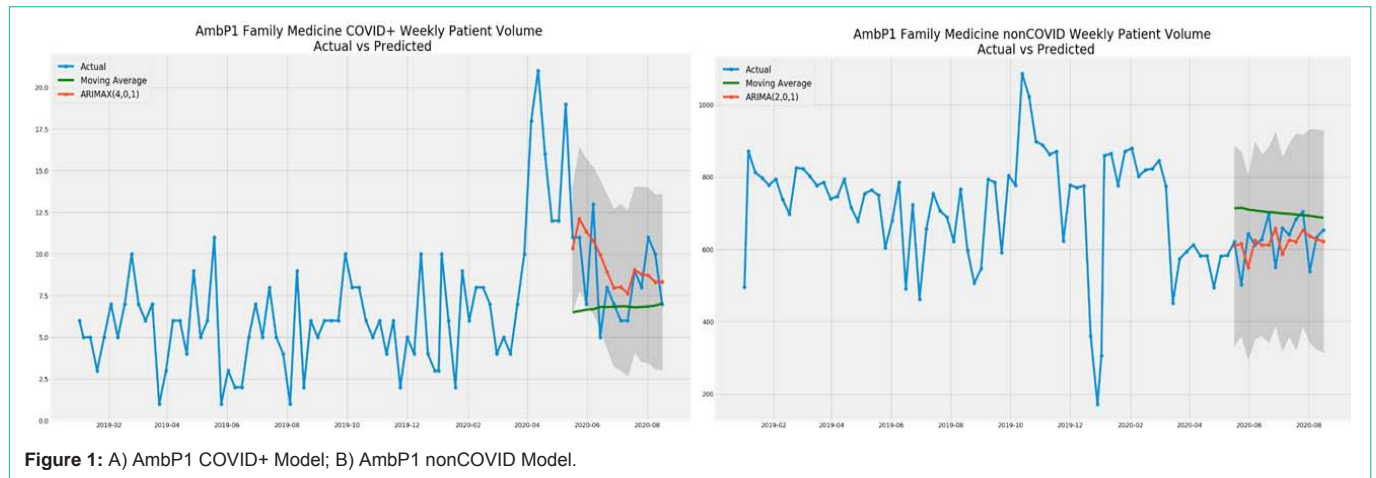
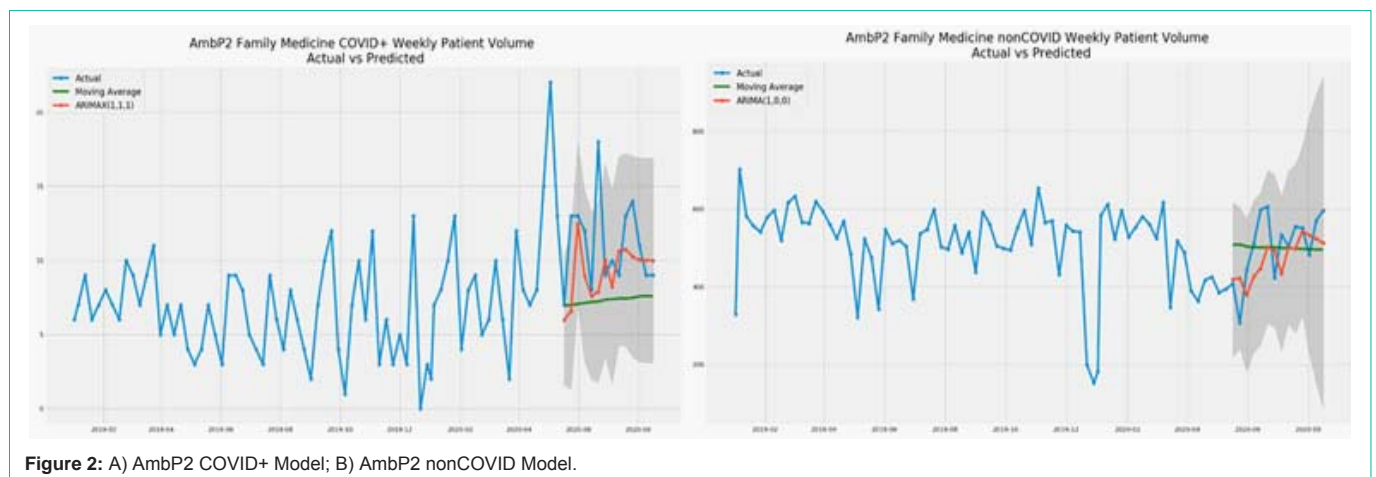**Figure 1:** A) AmbP1 COVID+ Model; B) AmbP1 nonCOVID Model.

Figure 2: A) AmbP2 COVID+ Model; B) AmbP2 nonCOVID Model.

Table 4: Validation Dataset Performance *RMSE; MAPE for each model.

		ARIMA Models			Moving Average	
		ARIMA (*p,d,q*)	RMSE	MAPE	RMSE	MAPE
AmbP1	COVID-19	ARIMAX (4,0,1)	2.22	24.86	2.92	23.21
	nonCOVID-19	ARIMA (2,0,1)	67.13	7.83	96.05	13.18
AmbP2	COVID-19	ARIMAX (1,1,1)	3.62	14.49	4.73	29.75
	nonCOVID-19	ARIMA (1,0,0)	79.46	8.55	86.15	15.53

of weekly positive COVID-19 cases as an exogenous variable. The availability of COVID-19 vaccinations has decreased the number of positive cases; however, new variants of COVID-19 are still causing an increase of cases. Since the length of efficacy of the vaccine is uncertain, a method to predict the variant trends could be helpful in improving our model.

Limitations

The current study is limited to retrospectively using electronic health records and positive COVID-19 results from only one hospital system. Our results may not be generalizable to other hospital systems, particularly those who serve patients with different characteristics. Other forecasting methods may be appropriate for different hospitals due to the differences in organizational structure and resources. Our study period is limited to one year of historical

data, ignoring potential factors such as weather and seasonal affects that could possibly improve forecasting accuracy. However, due to the unpredictable nature of COVID-19 our regular volumes and trends were disrupted. Therefore, using volumes from more years might not actually give us any more accurate prediction since our system is in a transient state, especially during the time of this study. Also, our weekly predictions did not differentiate between in-person and Telehealth volumes. In future studies, dividing the volumes between in-person and Telehealth could improve accuracy and provide additional information to healthcare providers for resource planning. Lastly, our models are short-term forecasts. Long-term forecasts can be generated, although the error rate will increase as the prediction period increases.

Conclusion

Accurately predicting future patient volumes in the ambulatory setting is essential for resource planning and developing guidelines for safely providing appropriate in-person visits. This study contributes to the exploration of time series modeling to forecast ambulatory patient volumes in ChristianaCare during the COVID-19 pandemic. We compared the forecasting accuracy of a moving average approach and ARIMA. Our findings show that a time series model that accounts for the number of positive COVID-19 patients delivers better performance for predicting weekly ambulatory patient

volumes in a short-term period. This improved forecasting ability can be used to provide health systems administrators decision support for clinic operations.

Declarations

Author's contributions: TC designed and conceptualized the overall study. RCL performed time series model development and evaluation. CKH performed the statistical analysis. KN performed data extraction and cleaning. RCL and CKH led the writing of this manuscript. CTJ, MAP, RK, CT, and TC provided input in the interpretation of the results, reviewed the manuscript, and contributed to revisions. All authors gave their approval for the final version to be submitted and published.

References

1. Centers for Disease Control and Prevention. Symptoms of Coronavirus. CDC. 2021.

2. Helmy YA, Fawzy M, Elaswad A, Sobieh A, Kenney SP, Shehata AA. The COVID-19 Pandemic: A Comprehensive Review of Taxonomy, Genetics, Epidemiology, Diagnosis, Treatment, and Control. J Clin Med. 2020; 9: 1225.

3. Pascarella G, Strumia A, Piliego C, et al. COVID-19 diagnosis and management: a comprehensive review. J Intern Med. 2020; 288: 192-206.

4. World Health Organization. Coronavirus disease (COVID-19) pandemic.

5. Centers for Disease Control and Prevention. Coronavirus disease 2019 (COVID-19): cases in US Title.

6. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes among 5700 Patients Hospitalized with COVID-19 in the New York City Area. JAMA - J Am Med Assoc. 2020; 323: 2052-2059.

7. Delaware, Health D of P, Response C. Coronavirus Disease (COVID-19).

8. Centers for Disease Control and Prevention. COVID-19. 2021.

9. Vizient. Effects of the COVID-19 Pandemic on Telehealth. 2021.

10. Li J, Wang J, Jin Z. SIR dynamics in random networks with communities. J Math Biol. 2018; 77: 1117-1151.

11. Woodul RL, Delamater PL, Emch M. Hospital surge capacity for an influenza pandemic in the triangle region of North Carolina. Spat Spatiotemporal Epidemiol. 2019; 30.

12. Malavika B, Marimuthu S, Joy M, Nadaraj A, Asirvatham ES, Jeyaseelan L. Forecasting COVID-19 epidemic in India and high incidence states using SIR and logistic growth models Logistic growth model SIR model Time interrupted regression model Projection. 2020.

13. Aguiar M, Stollenwerk N. SHAR and effective SIR models: from dengue fever toy models to a COVID-19 fully parametrized SHARUCD framework. Commun Biomath Sci. 2020; 3: 60-89.

14. Moss R, Zarebski AE, Carlson SJ, McCaw JM. Accounting for healthcare-seeking behaviours and testing practices in real-time influenza forecasts. Trop Med Infect Dis. 2019; 4.

15. Billingham S, Widrick R, Edwards NJ, Klaus SA. COVID-19 (SARS-CoV-2) Ventilator Resource Management Using a Network Optimization Model and Predictive System Demand. 2020.

16. Weissman GE, Crane-Droesch A, Chivers C, et al. Locally Informed Simulation to Predict Hospital Capacity Needs During the COVID-19 Pandemic. Ann Intern Med. 2020; 173: 21-28.

17. Chiam T, Papas M. Overcoming a pandemic: Delaware. J Public Heal. 2020; 6: 44-48.

18. van Bussel EM, van der Voort MBVR, Wessel RN, van Merode GG. Demand, capacity, and access of the outpatient clinic: A framework for analysis and improvement. J Eval Clin Pract. 2018; 24: 561-569.

19. Batal H, Tench J, McMillan S, Adams J, Mehler PS. Predicting patient visits to an urgent care clinic using calendar variables. Acad Emerg Med. 2001; 8: 48-53.

20. Capan M, Hoover S, Jackson EV, Paul D, Locke R. Time series analysis for forecasting hospital census: Application to the neonatal intensive care unit. Appl Clin Inform. 2016; 7: 275-289.

21. Reich NG, Brooks LC, Fox SJ, et al. A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. Proc Natl Acad Sci USA. 2019; 116: 3146-3154.

22. Luo L, Luo L, Zhang X, He X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. BMC Health Serv Res. 2017; 17: 1-13.

23. Delurgio S, Denton B, Cabanela RL, et al. Forecasting Weekly Outpatient Demands at Clinics within a Large Medical Center. Prod Invent Manag J. 2009; 45: 35-46.

24. Elgohari H, Bakr AM, Majeed MA. Forecasting the number of outpatient visits in tertiary hospital using time series based on ARIMA and ES models. Aust J Basic Appl Sci. 2019; 13: 70-77.

25. Ibrahim M, Jamil M, Akhtar AM, Mir Z-H, Akbar S, Ahmad M. Forecasting of Patient ' S Influx At Outpatient Medical Laboratory (Opml); Prof Med J. 2016: 1-9.

26. Brockwell PJ, Davis RA. Introduction to Time Series and Forecasting. 3rd ed. Springer International Publishing. 2016.

27. United States Census Bureau. Quick Facts New Castle County, Delaware.

28. Healthcare A for R and Q, Project HC and U Elixhauser Comorbidity Software, Version 3.7.