



How to cite this article:

Cael A. Lincuna, Michelle Anne C. Tolentino & Ma. Carlota B. Decena. (2023). A time series analysis of tuberculosis incidences in Pasig City, Philippines. *Journal of Computational Innovation and Analytics*, 2(2), 219-251. <https://doi.org/10.32890/jcia2023.2.2.5>

A TIME SERIES ANALYSIS OF TUBERCULOSIS INCIDENCES IN PASIG CITY, PHILIPPINES

**¹Ma. Carlota B. Decena, ²Michelle Anne C. Tolentino,
& ³Cael A. Lincuna**

^{1,2&3}Department of Mathematics and Physics,
College of Science, University of Santo Tomas, Manila, Philippines

¹Research Center for the Natural and Applied Sciences,
University of Santo Tomas, Philippines

¹Corresponding author: mbdecena@ust.edu.ph

Received: 26/6/2022 Revised: 12/9/2022 Accepted: 30/5/2023 Published: 31/7/2023

ABSTRACT

Tuberculosis (TB) is a serious infectious disease caused by *Mycobacterium Tuberculosis* that mainly affects the lungs but can also attack various body organs. Globally, it has been reported that the annual number of people provided with TB treatment has grown from 6 million in 2015 to 7 million in 2018 and then 7.1 million in 2019 (WHO, 2020). In the Philippines alone, there had been an estimated 500,000 incident cases of TB in 2019. The study's objectives are to develop a model that indicates the occurrence of TB in Pasig City, determine the incidence rate in terms of gender and age of TB patients, and obtain the projected number of TB cases that will

occur in 2021. In this study, only age and gender were considered in the demographic profile of the TB patients. Note that the data were organized per year and grouped according to age and gender. Since data gathered from the Pasig City Health Office (PCHO) is annually tallied, researchers used cubic spline interpolation to get data points between the given data to create an Autoregressive Integrated Moving Average (ARIMA) model. Consequently, it was used to forecast the projected number of TB patients for 2021 compared to the actual data obtained in PCHO. The results present the best fit ARIMA model per age group followed by the predicted number of cases in 2021. It is worth noting that AG2M and AG3M have the least accurate model based on the Root Mean Square Error (RMSE), a measure of accuracy between the projected value and the actual, due to their proportion.

Keywords: Tuberculosis, incidence, cubic spline interpolation, autoregressive integrated moving average.

INTRODUCTION

Infectious diseases are disorders caused by organisms such as bacteria, viruses, and fungi (Kotra, 2007). Some are lower respiratory tract infections, including pneumonia, Tuberculosis (TB), malaria, and Human Immunodeficiency Virus (HIV) infection. Note that such diseases caused some of the pandemics that humanity experienced, such as the bubonic plague and the coronavirus disease 2019 (COVID-19) pandemic (Parry & Peterson, 2020).

Globally, it has been reported that the annual number of people provided with TB treatment has grown from 6 million in 2015 to 7 million in 2018 and then 7.1 million in 2019. In the Philippines alone, there had been an estimated 500,000 incident cases of TB in 2019. Likewise, rifampicin- and multidrug-resistant TB disease has grown annually from a global total of 122,726 in 2016 to 156,205 in 2018 and 177,099 in 2019 (WHO, 2020). One of the leading infectious diseases that cause death worldwide is TB. Despite being preventable and with an 85% success rate in treatment, data shows that 1.4 million people died from TB in 2019 (WHO, 2019). In the same year, it was recorded that 87% of the new cases came from 30 heavily tolled countries, with India on top of the list and including the Philippines.

STATEMENT OF THE PROBLEM

TB has been around for centuries. Innovations and medical breakthroughs have been made, but the death toll associated with TB remains significant despite different medicines tested to cure TB. Even though numerous studies about TB were done, locally-based research was limited in number. Therefore, this research aims to address the problems regarding TB by creating an age-gender-dependent time-series analysis of TB occurrences using data acquired from the Pasig City Health Office (PCHO) from 2015-2020. The analysis aims to present an initial profile of TB patients in Pasig City during this period. Furthermore, the dependency rate on gender and age for TB patients is also examined in this work.

The study focused on the years 2015 to 2020 when TB cases had significantly grown in South-East Asian countries. Specifically, only Pulmonary Tuberculosis (PTB) cases were considered in the scope of this study. The patients were categorized as new, relapsed, or previously treated together with their respective gender and age group. Pasig City has been chosen for the study cohorts since there have been reports that this city contains high levels of NO₂ and PM2.5. High levels of air pollution, with exposure to NO₂ and PM2.5, greatly affect the body against airborne diseases and would yield severe cardiovascular and respiratory health illnesses, that includes TB (Kloog et al., 2013). Air quality monitoring stations in Pasig City recorded an estimated Air Quality Index of 66 (Moderate) and a PM2.5 level of 397 µg/m³ in January 2020 (*Air Quality in Manila*, 2021), which severely exceeds the acceptable PM2.5 concentration of 6.1 µg/m³.

LITERATURE REVIEW

PTB is a type of TB with bacterial infection of the lungs, symptoms which include but are not limited to chest pain, breathlessness, and coughing. In 2019, 57% of pulmonary cases were bacteriologically confirmed globally. This has increased from 55% in 2018. In 2019, 7.1 million new and relapse cases were reported, whereas 5.4 million patients (84%) had PTB. Direct smear Microscopy alone directly correlates with many unconfirmed PTB cases. Nevertheless, in high TB-burden countries, the differences in reporting and diagnostics are the cause of variation in the proportion of pulmonary cases

that are bacteriologically confirmed. In the Philippines, 30% are bacteriologically confirmed. Furthermore, from 2018 to 2019, the increase in notifications in high TB-burden countries correlated with the decrease in bacteriologically confirmed cases.

Extrapulmonary TB has the same definition as pulmonary cases, but it is an infectious disease that occurs in organ systems other than the lungs. In 2019, extrapulmonary TB represented 16% of the total 7.1 million incident cases, whereas, for the Philippines, it ranges from 0 - 9.9%.

One of the leading causes of death in the Philippines is TB. With a life expectancy of 69.1 years old, as of 2016, it has come a long way since 1980, when the life expectancy at birth is just 62.2 years old, all the while combating diseases preventing its population from perishing (WHO, 2020).

The Philippines is one of the seven highest TB burden countries in the world, and these seven countries account for 57% of the global TB incidence. Note that 75% of the individuals seeking initial care are in the private sector of the high burden countries. Private healthcare providers were trained to manage TB but are not supported systematically regarding the recording, reporting, and monitoring treatment. Furthermore, private expenditure represents 61%-74% of the total health expenditure. The National Tuberculosis Control Program (NTP) in the Philippines has implemented the Public-Private Mix (PPM) strategy since 2000 but has risen to the top 4 high-TB incidence countries. In 2017, 317,266, 55%, out of 581,000 patients with TB, were notified by the NTP and reported to WHO, and 17% were from the private sector. Although an estimated 217,925 first-line treatment is sold annually for anti-TB drugs, which comprises 43% of the country's total market, it provides evidence of a lack of systematic support for privately managed TB cases (Wells & Stallworthy, 2019).

People aged 15-49 years are considered the most productive age group in TB in South-East Asia (SEA) (WHO, 2020). This study will follow the transmission model used by (Fu et al., 2020), where TB patients were divided into three age groups: children (<15 years old), adults (15-64), and elders (≥ 65). The Global Tuberculosis Report of WHO in 2020 presents that TB is the highest burden in adult men

aged 15 years old and above, accounting for 56% of all TB cases in 2019. It was the highest among all the age groups and sexes compared to adult women, with only 32% of the cases of that same year and 12% composed of children. Note that cases of adult males were significantly elevated.

In a study conducted by Zhang et al. (2011), where age, sex, and race are considered to have factors in the presentation of TB, it is concluded that from its four age groups (15-24, 25-44, 45-64, and ≥ 65 ; all in years), the second and third age group, 25-44 years old and 45-64 years old, respectively are more likely to develop Extrapulmonary Tuberculosis (EPTB) together with PTB than the other age groups. The same age group but those of Somali and Asian descent were at a higher risk of contracting EPTB. However, on the contrary, those from Greenland with the same age group are less likely to have EPTB than all other age groups.

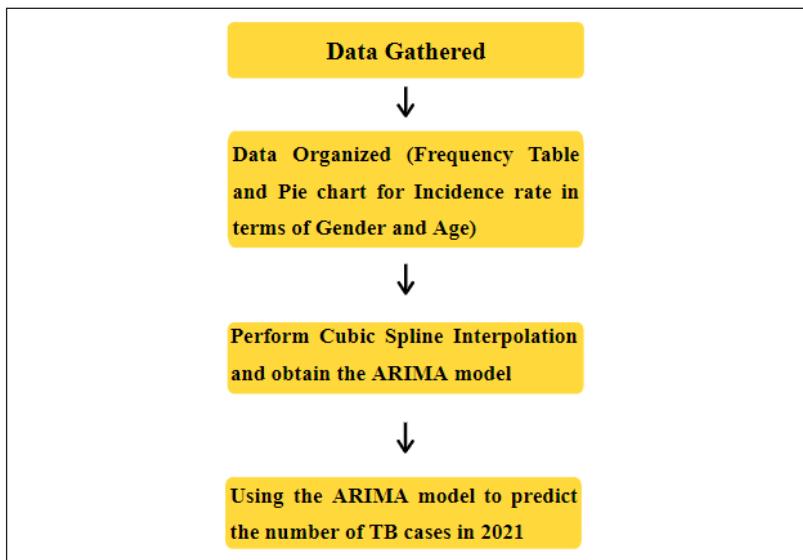
According to Cheng et al. (2020), elderly people risk developing TB due to immune responses. Male patients aged 70 years and above, living in rural areas with diabetes, congestive heart failure, chronic obstructive pulmonary disease, chronic kidney disease, and cancer were independent risk factors for TB in Taiwan Province (Cheng et al., 2020). Like global studies, male patients who smoke, with BMI < 18.5 , and with low annual household income are risk factors in developing TB. Moreover, elderly people have a higher chance of developing TB when they were/are a smoker with low BMI (Cheng et al., 2020). According to Cheng et al. (2020), a study conducted in 22 high TB-burden countries, mostly in Asia and Africa, reported the top contribution risk factors, namely, malnutrition (27.0%), smoking (21.0%) and HIV infection (16.0%).

METHODOLOGY

This section presents the research design and briefly describes the study population from Pasig City. The data gathering, statistical treatment, and applicable mathematical procedures are briefly presented below.

Figure 1

Methodology Flowchart



Data Gathering

Recorded files of TB patients in Pasig City from 2015 – 2020 were obtained with permission from data sources under the PCHO. Using retrospective random sampling, PTB patients were categorized into three age groups: less than 15, 15-64, and 65 years or above. Subsequently, collected data were filtered and organized in Microsoft Excel using the following format:

Table 1

Encoded Data

Total Pulmonary TB Cases	< 15 years old		15-64 years old		≥ 65 years old		Total	
	M	F	M	F	M	F	M	F
New								
Relapse								
Subtotal								
Total								

Cubic Spline Interpolation

Since the PTB cases were reported annually by the PCHO, a cubic spline interpolation was used to approximate the PTB cases in finer time intervals, resulting in monthly data points within the covered time range. Using the MATLAB cubic spline tool, a series of cubic polynomials were fitted in between each of the available data points (nodes), each resulting in continuous and smooth curves. The resulting cubic splines determine the cumulative change or rate of change on the given interval.

The equation is given as follows:

$$S(x) = \{S_1(x) \text{ if } x_1 \leq x \leq x_2, S_2(x) \text{ if } x_2 \leq x \leq x_3, \dots, S_{t-1}(x) \text{ if } x_{t-1} \leq x \leq x_t, \dots\} \quad (1)$$

where $i = 1, 2, 3, \dots, t - 1$

ARIMA Model t = time of diagnosis

The cubic spline-generated time series of PTB cases is then used and tested to fit an Autoregressive Integrated Moving Average (ARIMA) model. The ARIMA model was then used to forecast the projected number of TB patients for the year 2021 in comparison to the actual data reported by PCHO. It is a statistical model that uses time series data to predict the incidence rate and even forecast future trends. A shorthand notation for this is given as:

$$\text{ARIMA } (p, d, q), \quad (2)$$

where

p = non-seasonal autoregressive order

d = non-seasonal differencing

q = non-seasonal moving average order

The equation is given as follows:

$$\hat{y}_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} - \theta_1 e_{t-1} - \dots - \theta_q e_{t-q}, \quad (3)$$

where

μ = mean of the data

Φ = slope parameter

Θ = moving average parameter

The choice of the ARIMA model for this research was based on the study of Yan et al. (2019), where the ARIMA statistical model

was used to predict the incidence rate for 2018 - 2019 based on the monthly incidence of PTB in China from January 2005 to December 2017. Note that the predicted and actual data were compared to determine the effectiveness of the ARIMA model. Furthermore, it is recommended by Yan et al. (2019) that the ARIMA model is the best to use in the short-term prediction of TB cases.

Root Mean Square Error

To measure the accuracy of prediction for the 2021 PTB cases using the ARIMA model, the Root Mean Square Error (RMSE) is computed. RMSE is the standard deviation of the residuals or prediction errors. In contrast, it measures the distance of the data from the regression line to determine if it is within the best-fit line. It is also the square root of the average of the squared differences of the forecasted value and the observed value of the i th observation divided by the number of observations. The equation is given as follows:

$$\sqrt{\frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{N}}, \quad (4)$$

where

N = the number of observations available for analysis

x_i = forecasted value for the i th observation of data

\hat{x}_i = observed value for the i th observation of data

The RMSE is a good measure of the accuracy of the data and helps in comparing prediction errors of the computed model. A low RMSE value (≤ 0.2) of a model indicates that the model is highly accurate in predicting the data. Meanwhile, RMSE values between 0.2 and 0.5 show that the model can relatively predict the data accurately (Kenney & Keeping, 1962).

RESULTS AND DISCUSSION

Data Gathered

The data acquired from the PCHO were anonymized from 2015-2020. Since the study aims to present an age-gender dependent analysis of PTB occurrences, the demographic profile of patients obtained was

filtered to these two variables. The data were organized per year and grouped according to age and gender. From eight age groups, it was narrowed down to three based on the study of (Fu et al., 2020). The first group included children under 15 who were recorded to have TB and are classified as pediatric TB (CDC, 2021). The second age group, which ranges from 15 to 64 years of age, was considered the active age group among all three and where most of the working class was found. The last group, all 65 years old and above, were grouped because they were more susceptible to acquiring different diseases due to weakened immunity brought upon by older age, especially TB. Driving mechanisms for the prevention and treatment of TB are not clear on the age disparities, and age is a crucial factor in shaping TB epidemiology, progression risks, and TB control (Fu et al., 2020). The three groups were further categorized by gender, resulting in six study groups. The data for the year 2021 was also obtained to compare the forecasted occurrence corresponding to the empirical records.

Study Population

After the acquisition and sorting of the data, it is programmed in RStudio with respect to the research needs. TB patients diagnosed from the years 2015 - 2020 were the ones considered. The data has been divided according to TB into groups based on the given data of the PCHO. Subsequently, the age category of patients is presented in Table 2 and further categorized according to gender, shown in Table 3.

Table 2

Age Groups and Their Corresponding Codes

Age	Description	Code
0-4 years old		
5-14 years old	Pulmonary Tuberculosis	AG1
15-24 years old		
25-34 years old		
35-44 years old	Pulmonary Tuberculosis	AG2
45-54 years old		
55-64 years old		
>=65 years old	Pulmonary Tuberculosis	AG3

Table 3

Code for Cohorts (The Syntax For Coding is Age Group + Gender)

AG1M	Age Group 1 Male	AG1F	Age Group 1 Female
AG2M	Age Group 2 Male	AG2F	Age Group 2 Female
AG3M	Age Group 3 Male	AG3F	Age Group 3 Female

Frequency Distribution of the Sample

This part presents the frequency distribution of the TB data per year regarding their respective age group and gender. The summary distribution is presented in Table 4.

Table 4

Distribution Table of TB Cases by Age and Gender Group

Year	(AG1M)	(AG1F)	(AG2M)	(AG2F)	(AG3M)	(AG3F)
2015	223	200	888	517	89	52
2016	230	224	956	548	76	68
2017	210	178	916	608	105	70
2018	230	206	1414	775	124	96
2019	143	141	1859	1154	228	177
2020	85	50	1137	610	125	96
2021	106	65	1410	871	206	115

For example, in 2015, there were 223 male patients under age group 1 (AG1M), while 200 female patients were diagnosed with pulmonary TB under age group 1 (AG1F). A similar analysis is drawn for the years 2016 to 2021 on the age groups 2 and 3 for males and females. Here, a significant observation is seen in all years regarding gender, where more male TB patients are recorded than female patients.

For age groups throughout 2015 – 2020, 2,135, 2,290, 2,405, 3,172, 3,989, and 2,400 cases occurred, respectively. The distribution of TB cases for 2021 will be used as a reference for comparison on the ARIMA prediction of the pulmonary TB cases for 2021 per age

group. AG1 consists of the age intervals: 0 - 4 and 5 - 14 years old. As observed throughout the tables, AG2 has the largest population because it consists of 5 Age intervals: 15 - 24 years old, 25 - 34 years old, 35 - 44 years old, 45 - 54 years old, and 55 - 64 years old. It has been established that TB is the highest burden in adult men aged 15 years old and above, accounting for 56% of all TB cases in 2019 (WHO, 2020). Note that AG3 has fewer statistics than AG2 since it only consists of 1 age interval: ≥ 65 years old.

Based on the PCHO data, AG2 has the highest number of cases, mainly in the 15 - 34 age range. Adulthood and early adolescence are key risk periods for TB infection, diseases, and adverse effects. In 2020, the highest TB cases occurred in the WHO SEA Region, with 43% of new cases. The Philippines saw a large decrease in annual notifications between 2019 and 2020, with a 37% drop because of the COVID-19 pandemic and the heightening of restrictions. That is why there is a huge difference between the number of cases in the TB data from the PCHO in 2019 and 2020. In SEA, TB increases rapidly during adolescence to peaks in early adulthood (WHO, 2020). Based on the data, AG1 has higher cases than AG3, but both are consistently low throughout the years 2015 - 2019, but in the years 2020 and 2021, there is a rise in the number of cases for AG3. This is due to the restriction on movement for the ages 0 - 14 due to the COVID-19 pandemic, but as restrictions ease, TB detection improves (WHO, 2020). AG2, although it has the larger proportion of the two, remains consistently high. According to a study by Snow et al. (2018), over a quarter of TB patients in the Philippines are children, adolescents, and young adults, reflecting the country's young population. The Philippines' large number of patients under 25 has implications that indicate both in the short term when the disease burden among children and young people is significant. Meanwhile, in the long term, the current generation ages with a high incidence of latent TB infection.

Incidence Rate in Terms of Gender and Age

The rate of occurrence of PTB among patients of Pasig City is examined according to age and gender from 2015 to 2020. A visualization of the occurrence of PTB among patients in Pasig City according to gender and age is presented in Figures 2 and 3, respectively.

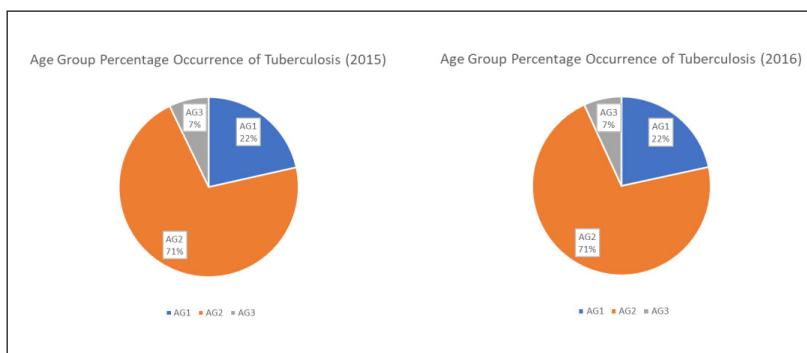
Figure 2

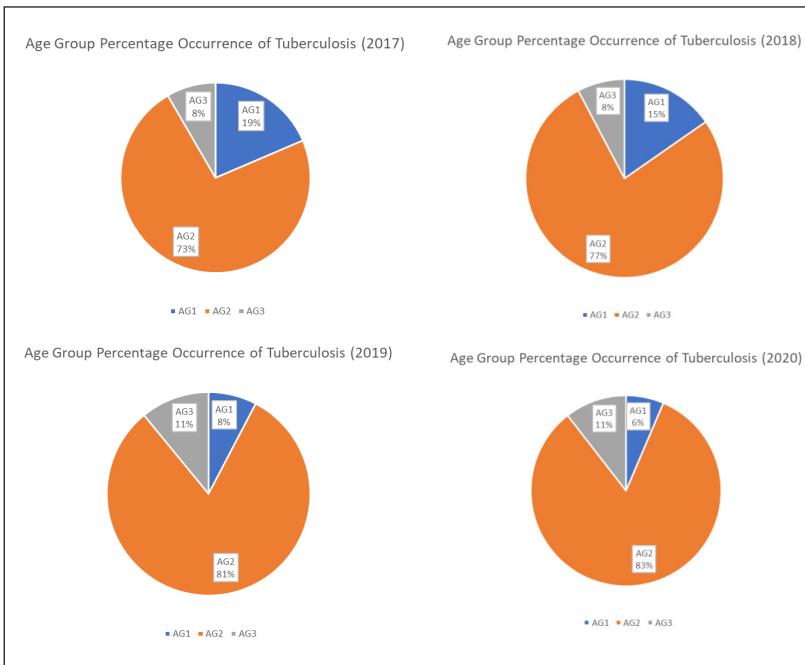
Gender Percentage Occurrence of Tuberculosis (2015-2020)



Figure 3

Age Group Percentage Occurrence of Tuberculosis (2015-2020)





Case notification rates are higher for males than for females. Based on the pie chart, the ratio of gender occurrence of TB cases is at least twice higher among men than women, approximately 60:40, which suggests that men are more likely to get TB than women. According to Horton et al. (2016), TB prevalence is significantly higher among men than women in countries with low- and middle-income, presenting strong evidence that men are disadvantaged in accessing TB care in many settings. Globally, males have 1.8 times the notification rate than females (WHO, 2020). As Pasig City recovers from the economic slump brought upon by the pandemic, the city entices more investors. An ordinance was passed to help the city recover from the economic collapse, creating an economic development and investment office. In that way, investors were more likely to set up their offices in the city through the help of investor-friendly ordinances that would bring upon jobs that included physical labor (Tucay Quezon & Ibanez, 2021), exposing these men to diseases, including TB. National TB programs and global strategies should recognize the male group as a high-risk group and improve access to diagnostic and screening services to effectively monitor and manage TB and ensure gender equity in TB care.

On the other hand, on Age Group Percentage of Tuberculosis throughout 2015 – 2020, as observed in Figure 3, consistently, the largest part of the pie chart is always AG2. It aligns with the statistics released by WHO in 2020, where it has been observed that TB is the highest burden in adult men aged 15 years old and above, accounting for 56% of all TB cases globally in 2019. The prevalence of TB is high in the ages 15-24 and 25-34 since they are more likely to develop TB. Although AG2 has the larger proportion of the two, it remains consistently high. Based on the national TB registry, the Department of Health (DOH) reported 311,000 TB cases in 2021 compared to 263,000 in 2020. The Philippines is one of the 16 countries where essential TB-related services were affected by the COVID-19 pandemic. In 2020, TB services decreased, comparable with the level recorded globally in previous years, with the Philippines as one of the countries being the most affected (WHO, 2020). This is why there is a huge difference in the 2019 and 2020 TB data of the PCHO.

Cubic Spline Interpolation

This part demonstrates cubic spline interpolation of the data points to indicate the data points in between the given data points. The given data is interpolated to have finer data points since the data from the PCHO is only per year. The researchers interpolated through the given yearly data and created 12 splines between data points to simulate monthly occurrences of TB. Note that the interpolation was done monthly to provide more data points to get the best ARIMA model in the next section. It can be observed that the graph of the actual data was smoothed through cubic spline interpolation, and more points are now used to create the time series analysis and ARIMA modeling. The following shows the interpolated distribution in comparison to the actual data:

Figure 4 illustrates the graph of the interpolated points versus the actual data points for Age Group 1 - Males. The highlighted points were the confirmed number of TB patients in the group with values 223, 230, 210, 230, 143, and 85, respectively.

Figure 5 depicts the graph of the interpolated points versus the actual data points for Age Group 2 - Males. The highlighted points were the confirmed number of TB patients in the group with values 888,

956, 916, 1414, 1859, and 1137, respectively. Figure 6 portrays the graph of the interpolated points versus the actual data points for Age Group 3 – Males. The highlighted points were the confirmed number of TB patients in the group, with values of 89, 76, 105, 124, 228, and 125, respectively. Figure 7 shows the graph of the interpolated points versus the actual data points for Age Group 1 – Female. The highlighted points were the confirmed number of TB patients in the group, with values of 200, 224, 178, 206, 141, and 50, respectively. Figure 8 shows the graph of the interpolated points versus the actual data points for Age Group 2 - Female. The highlighted points were the confirmed number of TB patients in the group with values 517, 548, 608, 775, 1154, and 1137, respectively. Figure 9 shows the graph of the interpolated points versus the actual data points for Age Group 3 - Female. The highlighted points were the confirmed number of TB patients in the group, with 52, 68, 70, 96, 177, and 96 values, respectively.

Figure 4

Interpolated vs. Actual Data of Age Group 1-Male

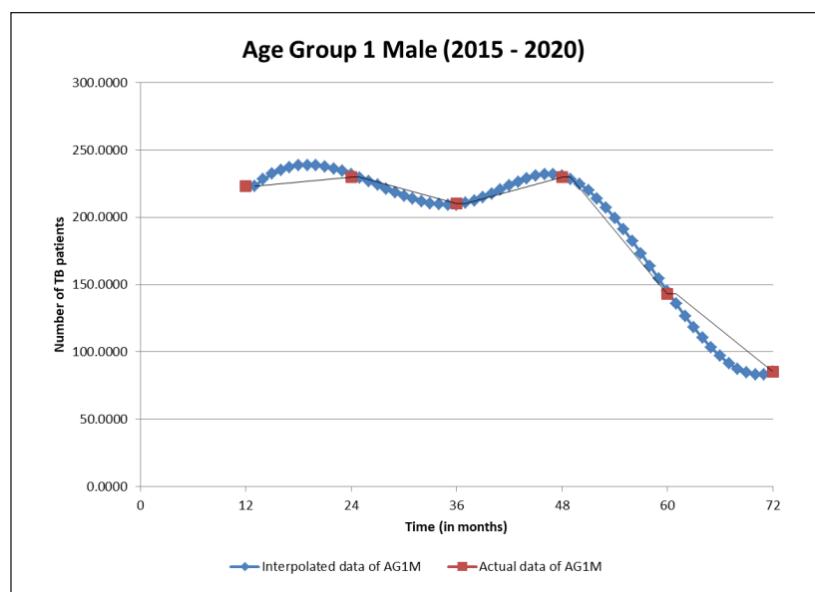


Figure 5

Interpolated vs. Actual Data of Age Group 2-Male

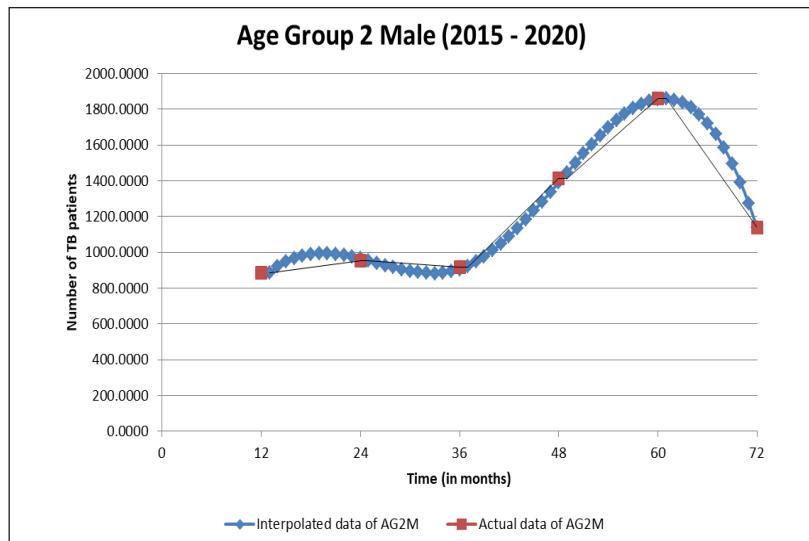


Figure 6

Interpolated vs. Actual Data of Age Group 3-Male

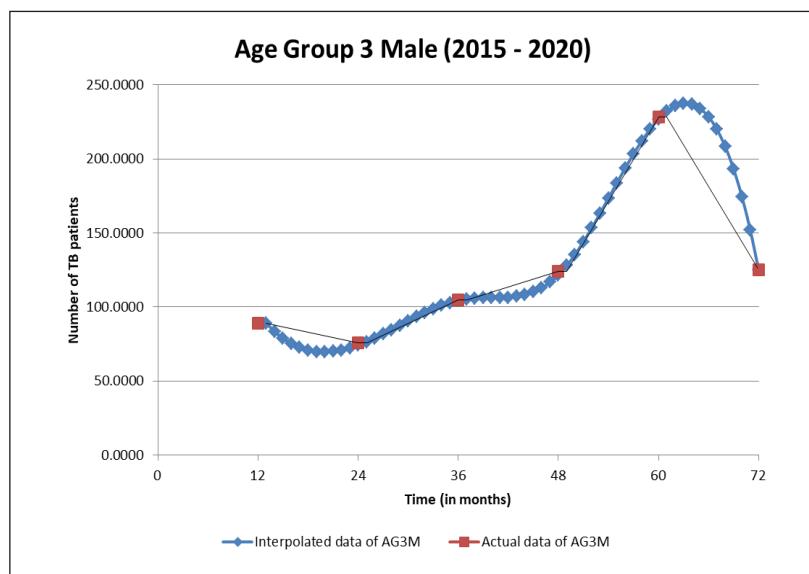


Figure 7

Interpolated vs. Actual Data of Age Group 1-Female

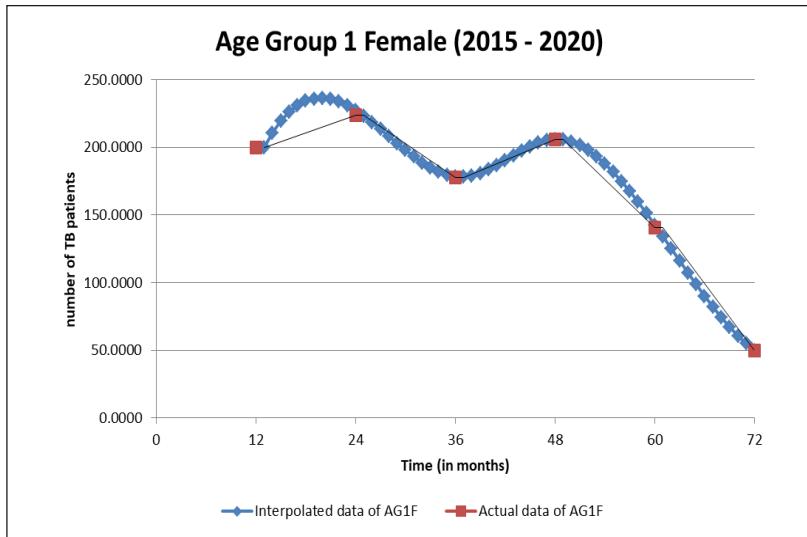


Figure 8

Interpolated vs. Actual Data of Age Group 2-Female

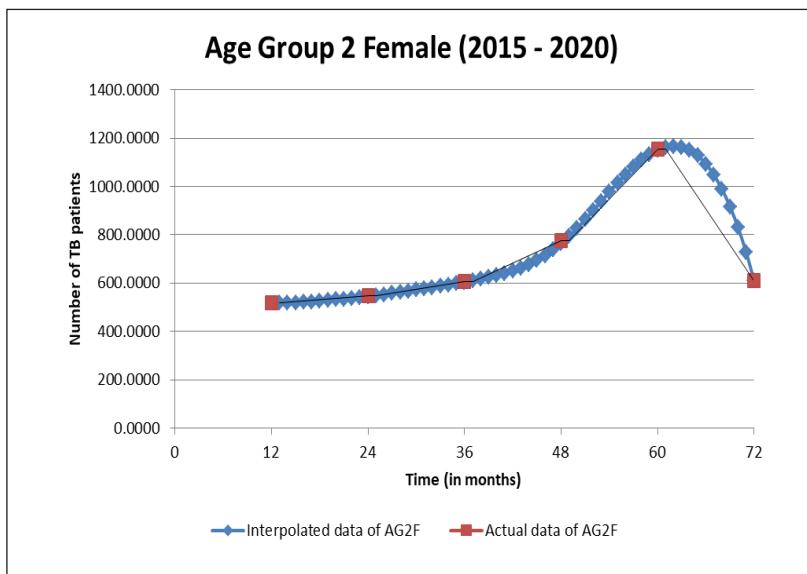
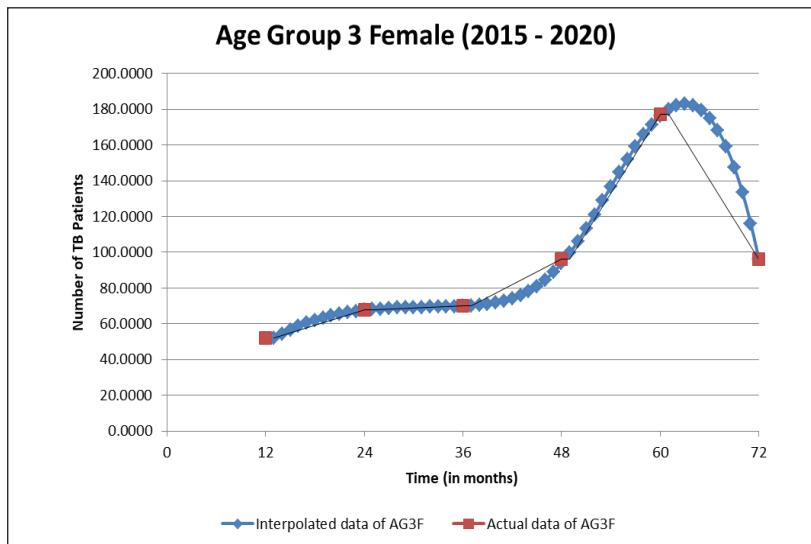


Figure 9

Interpolated vs. Actual Data of Age Group 3-Female



Statistical Analysis

ARIMA Model

After performing the cubic spline interpolation, each group was estimated to obtain the ARIMA parameters. This will aid in forecasting the number of TB patients for 2021. Note that only the values for December will be considered since the data only accounts for a yearly basis, and other values from January to November are considered as interpolated points to aid in the prediction of the graph.

Table 5

AGIM ARIMA Model Results

ARIMA(2,2,1)		
Coefficients:	ar1	ar2
	1.9058	-0.9215
Standard error	0.0284	0.0303
	ma1	0.7856
	0.0730	

(continued)

ARIMA(2,2,1)	
Coefficients:	
sigma^2	0.001609
log likelihood	190.59
AIC	-373.19

In Table 5, the ARIMA model that best fits AG1M is ARIMA (2,2,1) since it produced the lowest AIC with the value -373.19 using the `auto.arima()` function. It determines the best model using the R programming language. If $p = 2$, the predicted data will increase, and $d = 2$, which means it will also increase.

Table 6 presents the predicted values for AG1M for 2021 per month after using the ARIMA model to predict. It can be observed that the predicted data for 2021 has a value of 96.7013 with a standard error of 11.37. Compared to the actual data, 106, it is relatively close, with only a difference of approximately 10. Figure 10 illustrates a graphical representation of the predicted values.

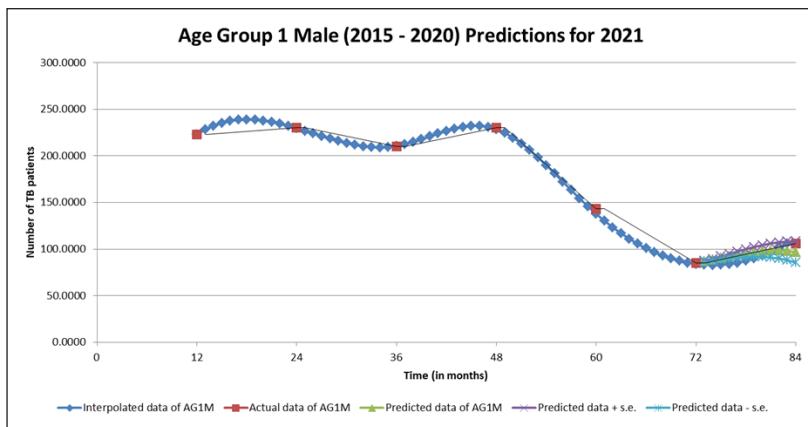
Table 6

Prediction of TB patients of AG1M 2021

	2021 Prediction of AG1M	2021 S.E of Prediction for AG1M	2021 Actual Data for AG1M
January	86.7563	0.447579	82.7721
February	88.5702	1.000817	82.7818
March	90.4563	1.674035	83.2085
April	92.3221	2.449261	84.0534
May	94.0751	3.314721	85.3177
June	95.6246	4.261944	87.0022
July	96.9083	5.284462	89.1083
August	97.8849	6.377128	91.6370
September	98.4775	7.535722	94.5894
October	98.5622	8.756705	97.9667
November	98.0125	10.03705	101.7698
December	96.7013	11.37413	106.0000

Figure 10

AG1M Prediction for 2021



For Table 7, the ARIMA model that best fits AG2M is the ARIMA (0,2,0) since it produced the lowest AIC value, 397.8. Since $d = 2$, it means that the AG2M data will either linearly increase or decrease. That is why there is a sharp decrease and then an increase in the curve for the predicted data in the graph. Here, a factor that affected the prediction that caused a downward trend is the huge difference in the actual data between 2019 and 2020, which are 1,859 and 1,137, respectively.

Table 7

AG2M ARIMA Model Results

ARIMA(0,2,0)				
Coefficients:				
	ma1			
	0.6726	sigma^2		14.89
Standard error	0.1136	Log-likelihood		-196.9
AIC	397.8			

Table 8

Prediction of TB patients of AG2M 2021

	2021 Prediction of AG2M	2021 S.E. of Prediction for AG2M	2021 Actual Data for AG2M
January	1018.80682	3.858209	1011.9841
February	895.84105	8.627218	982.7418
March	767.82756	14.4361	963.1226
April	634.49139	21.13228	954.2807
May	495.55742	28.61325	957.3698
June	350.75055	36.80497	973.5441
July	199.79586	45.65095	1003.9575
August	42.41819	55.10625	1049.7640
September	121.65745	65.13407	1112.1174
October	292.70622	75.70353	1192.1720
November	471.00304	86.78827	1291.0815
December	656.82307	98.36542	1410.0000

Figure 11

AG2M Predictions for 2021

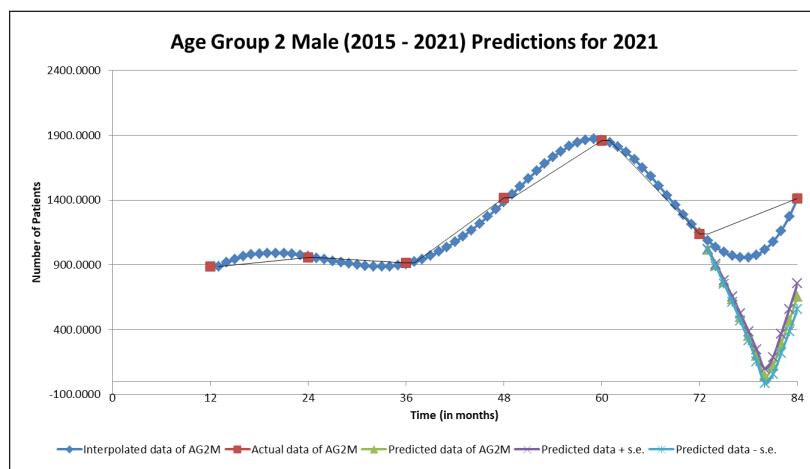


Table 8 shows the predicted values for AG2M for 2021 per month after using the ARIMA model to predict. The predicted number for 2021 is 656.82307, with a standard error of 98.3654, while the actual data

is 1,410 TB patients. As mentioned before, there is a huge difference in the total number of persons diagnosed with PTB between 2019 and 2020, which made the predicted sharp decrease then increase for 2021. Another factor that explains the sharpness of the curve is that the prediction was only obtained by interpolating the actual data from 2015 - 2020. The data given by the PCHO is only yearly. Hence, the accuracy of the prediction is affected. Figure 11 illustrates the graphical representation of the AG2M predictions for 2021.

Table 9 shows the best-fit model for AG3M is ARIMA (2,2,2) since it produced the lowest AIC value, -359.56. Since $p = 2$ and $d = 2$, the AG3M data predicted a linear increase or decrease. That is why it is seen in the graph there is a sharp decrease and then an increase in the curve for the predicted data, which is similar to the graph of AG2M.

Table 9

AG3M ARIMA Model Results

ARIMA (2,2,2)				
Coefficients:				
	ar1	ar2	ma1	ma2
	1.9555	-0.9577	0.8802	0.0769
Standard error	0.0091	0.0095	0.0084	0.0436
sigma ²		0.0005798		
Log-likelihood		184.78		
AIC		-359.56		

Table 10

Prediction of TB patients of AG3M 2021

	2021 Prediction of AG3M	2021 S.E of Prediction for AG3M	2021 Actual Data for AG3M
January	101.975454	0.982461	106.846
February	78.432064	2.196849	103.254
March	54.320198	3.676032	101.498
April	29.590258	5.381159	101.807
May	4.192648	7.286124	104.409
June	21.922266	9.372078	109.532

(continued)

	2021 Prediction of AG3M	2021 S.E of Prediction for AG3M	2021 Actual Data for AG3M
July	48.804081	11.62463	117.405
August	76.502464	14.03235	128.256
September	105.066943	16.58585	142.312
October	134.547186	19.27727	159.803
November	164.992824	22.09991	180.956
December	196.45342	25.04793	206.000

Figure 12

AG3M Predictions for 2021

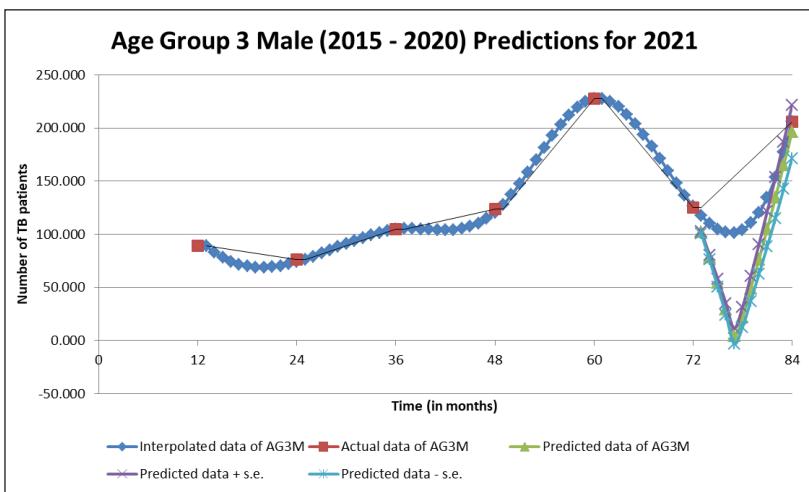


Table 10 presents the predicted values for AG3M for 2021 per month after using the ARIMA model to predict. The predicted number of PTB patients for 2021 in Pasig City is 196.45242, with a standard error of 25.04793, while the actual data for the count of patients in 2021 is 206. Although it can be observed that some of the monthly predictions were far from the actual points, they were only compared to the actual splined points. The only observation that should be the basis for the model's accuracy is the predicted value for December since the data is cumulatively collected at the end of the year. As observed in the graph, there is a huge dip in the 75th to 76th month. A factor that made the prediction rise and decrease sharply is the difference between the 2019 and 2020 actual data, which are 125 and 206, respectively. Similar to AG2M, the accuracy of the prediction is

affected by the interpolated data of AG3M. Figure 12 illustrates the graphical representation of the predicted values for AG3M for 2021. In Table 11, the best-fit model for AG1F is ARIMA (2,2,2) since it produced the lowest value of AIC, which is -362.57. Since $p = 2$ and $d = 2$, then it means that the AG1F data will either increase or decrease. Table 12 shows the predicted values for AG1F for 2021 per month after using the ARIMA model to predict. The predicted number of TB patients in Pasig City for AG1F is 50.19175, with a standard error of 25.8370. The actual data of TB patients in 2021 is 65. Figure 13 depicts the graphical representation of the predicted values for AG1F for 2021.

Table 11

AG1F ARIMA Model Results

ARIMA (2,2,2)				
Coefficients:				
	ar1	ar2	ma1	ma2
	1.9054	-0.9018	0.8320	0.0569
Standard error	0.0299	0.0319	0.1195	0.1179
σ^2		0.001298		
Log-likelihood		186.28		
AIC		-362.57		

Table 12

Prediction of TB patients of AG1F for 2021

	2021 Prediction of AG1F	2021 S.E of Prediction for AG1F	2021 Actual Data for AG1F
January	46.50937	0.036033	40.2821
February	43.65993	0.174467	38.1186
March	41.47283	0.494059	36.6930
April	39.95797	1.081354	36.0536
May	39.11479	2.026491	36.2482
June	38.93312	3.420171	37.3250
July	39.39405	5.351265	39.3319
August	40.47097	7.904887	42.3170
September	42.13051	11.16084	46.3284
October	44.33361	15.19238	51.4139
November	47.03651	20.06528	57.6218
December	50.19175	25.83708	65.0000

Figure 13

AG1F Predictions for 2021

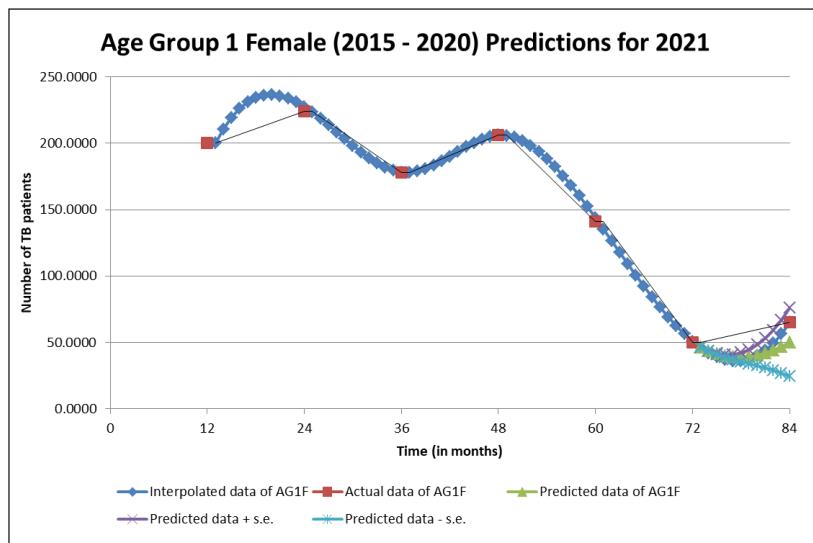


Table 13

AG2F ARIMA Model Results

ARIMA (0,2,1)			
Coefficients:			
ma1			
	0.3819	sigma^2	143.5
Standard error	0.1113	Log-likelihood	-272.73
AIC		549.46	

In Table 13, the best-fit model for AG2F is ARIMA (0,2,1) since it produced the lowest value of AIC, which is 549.46. Like in AG1F, $d = 2$ means that the data for AG2F is not stationary, and the prediction may still increase or decrease over time.

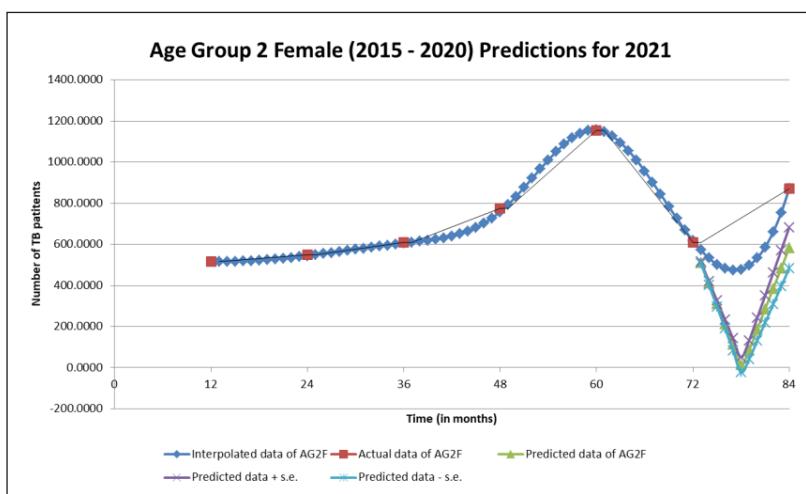
Table 14

Prediction of TB patients of AG2F for 2021

	2021 Prediction of AG2F	2021 S.E of Prediction for AG2F	2021 Actual Data for AG2F
January	510.5847	3.898112	513.3973
February	411.1694	8.716444	491.5226
March	311.7541	14.5854	477.5963
April	212.3388	21.35084	472.5947
May	112.9235	28.90918	477.4937
June	13.5082	37.18562	493.2696
July	85.9071	46.12309	520.8983
August	185.3224	55.67618	561.3561
September	284.7377	65.80771	615.6191
October	384.153	76.48649	684.6633
November	483.5683	87.68587	769.4649
December	582.9836	99.38276	871.0000

Figure 14

Interpolated vs Actual Data of Age Group 2 - Female



In Table 14, the projected number of TB patients in Pasig City is 582.9836, with a standard error of 99.38276, while the actual data for the count of patients in 2021 is 871. Note that the difference between

the actual data and the predicted data is large. As observed in Figure 14, there is a slight increase from 2020 up to 2021 but a sharp decrease from 2019 to 2020. Similar to AG2M, the accuracy of the prediction is affected by the large proportion of the data for AG2F.

Table 15

AG3F ARIMA Model Results

ARIMA(1,1,1)			
Coefficients:			
	ar1	ma1	
	0.8075	-0.3099	sigma^2
Standard error	0.0955	0.1529	Log likelihood
AIC			-76.64
			155.28

Table 16

Prediction of TB patients of AG3F for 2021

	2021 Prediction of AG3F	2021 S.E of Prediction for AG3F	2021 Actual Data for AG3F
January	79.0424	0.723228	78.4082
February	62.0848	1.617187	73.8573
March	45.1272	2.706072	70.4195
April	28.1696	3.961284	68.2491
May	11.2120	5.363604	67.5001
June	5.7456	6.899157	68.3266
July	22.7032	8.557351	70.8827
August	39.6608	10.32976	75.3225
September	56.6184	12.2095	81.8000
October	73.5760	14.19076	90.4694
November	90.5336	16.26862	101.4847
December	107.4912	18.43877	115.0000

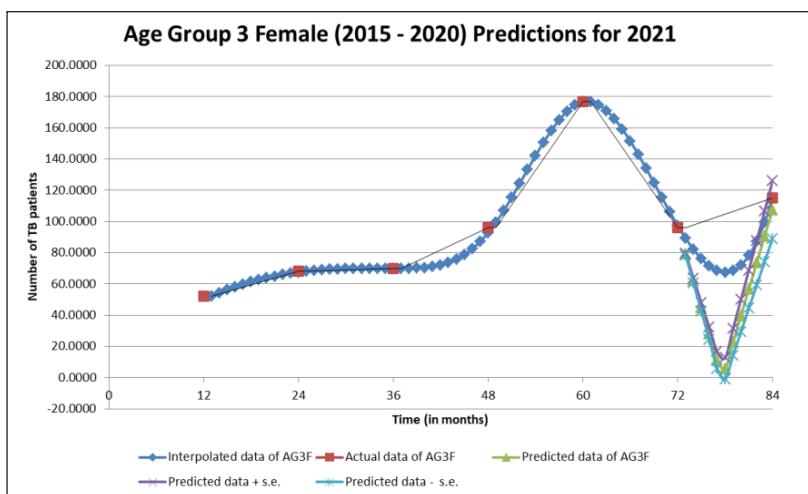
Table 15 presents the ARIMA model for AG3F. The best-fitted model for the age group is ARIMA (1,1,1), with AIC equal to 155.28. Since p

$= 1$, the data will increase or decrease linearly; since $d = 1$, the data is not stationary and will still tend to go up or down as time progresses.

In Table 16, the predicted number of TB patients in Pasig City for AG3F is 107.4912 with a standard error of 18.43877, while the actual data in 2021 is 115. The graph shows a sharp decrease from 2019 to 2020 because of the difference in the actual data from 2019 and 2020.

Figure 15

AG3F Predictions for 2021



Root Mean Square Error

After estimating the parameters of the ARIMA model per group, the predicted values will be compared to the actual values, and its accuracy will be measured through RMSE.

The researchers used ARIMA modeling to forecast the number of TB cases in 2021 using the data collected from 2015-2020. Due to the lack of data points, as the given data were cumulatively in years, the researchers performed cubic spline interpolation to facilitate the data smoothing.

As observed in Table 17, actual values of TB patients per age group of 2021 were given and compared to the generated predicted values

based on each year's fitted ARIMA models. AG2M has the highest RMSE recorded, which is equal to 217.4234516. This is because the data of this age group has decreased sharply during the last 2 years that were used for modeling. Note that AG2F was the second-highest RMSE among the groups, with a computed RMSE of 83.14317304, followed by AG1F with 4.274773562. The groups with similar RMSEs and the lowest ones were AG3M, AG1M, and AG3F, with values of 2.755860266, 2.684303474, and 2.167603851, respectively. The age groups with the lowest RMSEs were the same age groups whose predicted values are close to the actual values of the year 2021. It can be observed that ever since the onset of the pandemic in 2020, all age groups have had a sharp decrease in data that affected the accuracy of the prediction model used. This sharp decrease in data is because the transmission of infectious diseases was prevented because of the health and safety protocols implemented to prevent COVID-19. These measures helped curb the spread of such diseases, including TB.

For AG1M, in Table 6, it can be observed that the predicted data for 2021 has a value of 96.7013 with a standard error of 11.37 compared to the actual data, which is 106. Here, the RMSE value for AG1M is 2.684303474, which deems the predicted model accurate.

For AG2M, in Table 8, it can be observed that the predicted data for 2021 has a value of 656.82307 with a standard error of 98.3654 compared to the actual data, which is 1410. The RMSE value for AG2M is 217.4234516, which deems the model inaccurate. This is due to the large difference between the years 2019 and 2020.

For AG3M, in Table 10, it can be observed that the predicted data for 2021 has a value of 196.45342 with a standard error of 25.04793 compared to the actual data, which is 206. RMSE value for AG3M is 2.755860266, which deems the predicted model accurate.

For AG1F, in Table 12, it can be observed that the predicted data for 2021 has a value of 50.19175 with a standard error of 25.83708 compared to the actual data, which is 65. The RMSE value for AG1F is 4.274773562, which deems the predicted model accurate.

For AG2F, in Table 14, it can be observed that the predicted data for 2021 has a value of 582.9836 with a standard error of 99.38276 compared to the actual data, which is 871. The RMSE value for AG2F

is 83.14317304, which deems the model inaccurate. This is due to the large difference between the years 2019 and 2020. AG2F results are similar to that of AG2M.

For AG3F, in Table 15, it can be observed that the predicted data for 2021 has a value of 107.4912 with a standard error of 18.43877 compared to the actual data, which is 115. RMSE value for AG3F is 2.167603851, which deems the predicted model accurate. Table 17 summarizes the predicted values per group compared to the actual value and their computed root mean square.

Table 17

Prediction of TB patients of AG3F for 2021

Age Group	Actual Values	Predicted Values	RMSE
AG1 M	106	96.7013	2.684303474
AG2 M	1410	656.82307	217.4234516
AG3 M	206	196.45342	2.755860266
AG1 F	65	50.19175	4.274773562
AG2 F	871	582.9836	83.14317304
AG3 F	115	107.4912	2.167603851

Conclusions and Recommendations

Infectious diseases have been one of the causes of ailments all over the globe and can even cause death. One of these is TB, which is the focus of this study. This study aims to forecast TB cases in Pasig City for 2021 using the data collected from 2015-2020.

The researchers have acquired TB data from the PCHO to obtain the occurrence of TB per year. Based on the PCHO data, AG2 has the highest number of cases, mainly in the 15 - 34 age range. Adulthood and early adolescence are key risk periods for TB infection, diseases, and adverse effects. AG1 had higher cases than AG3 throughout the years 2015 - 2019, but in the years 2020 and 2021, it can be observed that there was a rise in the number of cases for AG3. This is due to the restriction in movement for AG1 due to the COVID-19 pandemic, making it hard to travel to healthcare facilities (WHO, 2020).

It was observed that through all the cases in the last six years, it is evident that there are more male patients recorded than female. Also,

men are more likely to be exposed to TB since they are more likely to obtain employment involving physical labor. In terms of age, AG2 comprises most of the recorded cases since it is also the largest age group, which indicates that the most productive age group was also highly exposed to TB among all the age groups. The best fit ARIMA models for each group are as follows: AG1M has the best fit ARIMA model of (2,2,1), AG2M with (0,2,0), AG3M with (2,2,2), AG1F with (2,2,2), AG2F with (0,2,1), and lastly, AG3F with an ARIMA model (1,1,1).

The computed RMSEs suggest how accurate the prediction was. AG1M, AG3M, and AG3F have the lowest values of RMSE with values of 2.755860266, 2.684303474, and 2.167603851, respectively. Therefore, AG1M, AG3M, and AG3F have the highest prediction accuracy among all groups. It is recommended to future researchers to 1) Expand the data set to at least 10 years and on a seasonal basis, either quarterly or monthly, for a more accurate prediction of TB cases. 2) Use other statistical treatments like Poisson regression to forecast data. 3) Use other analyses of TB cases, such as getting the survival rate of patients of different age groups.

Infectious diseases have been one of the causes of ailments all over the globe and can even cause death. One of these is TB, which is the focus of this study. This study aims to forecast TB cases in Pasig City for 2021 using the data collected from 2015-2020.

The best fit ARIMA models for each group are as follows: AG1M has the best fit ARIMA model of (2,2,1), AG2M with (0,2,0), AG3M with (2,2,2), AG1F with (2,2,2), AG2F with (0,2,1), and lastly, AG3F with an ARIMA model (1,1,1).

The computed RMSEs suggest how accurate the prediction was. AG1M, AG3M, and AG3F have the lowest values of RMSE with values of 2.755860266, 2.684303474, and 2.167603851, respectively. Therefore, have the highest accuracy of prediction among all groups. It is recommended to future researchers to 1) Expand the data set to at least 10 years and on a seasonal basis, either quarterly or monthly, for a more accurate prediction of TB cases. 2) Use other statistical treatments like Poisson regression to forecast data. 3) Use other analyses of TB cases, such as getting the survival rate of patients of different age groups.

ACKNOWLEDGMENT

The completion of this research has been possible with the help of the Pasig City Health Office, through Ms. Ma. Zenaida S. Ramos, in providing the data needed for the study. Without them, the research would not be possible.

REFERENCES

Air quality in Manila. (2021). IQA Ir. <https://www.iqair.com/philippines/ncr/manila>

CDC. (2021). *TB treatment for children.* Centers for Disease Control and Prevention. <https://www.cdc.gov/tb/topic/treatment/children.htm>

Cheng, J., Sun, Y. N., Zhang, C. Y., Yu, Y. L., Tang, L. H., Peng, H., Peng, Y., Yao, Y. X., Hou, S. Y., Li, J. W., Zhao, J. M., Xia, L., Xu, L., Xia, Y. Y., Zhao, F., Wang, L. X., & Zhang, H. (2020). Incidence and risk factors of tuberculosis among the elderly population in China: A prospective cohort study. *Infectious Diseases of Poverty*, 9(1), 64–76. <https://doi.org/10.1186/s40249-019-0614-9>

Fu, H., Lin, H. H., Hallett, T. B., & Arinaminpathy, N. (2020). Explaining age disparities in tuberculosis burden in Taiwan: A modelling study. *BMC Infectious Diseases*, 20(1), 1–12. <https://doi.org/10.1186/s12879-020-4914-2>

Horton, K. C., MacPherson, P., Houben, R. M. G. J., White, R. G., & Corbett, E. L. (2016). Sex Differences in Tuberculosis Burden and Notifications in Low- and Middle-Income Countries: A Systematic Review and Meta-analysis. *PLoS Medicine*, 13(9), e1002119. <https://doi.org/10.1371/journal.pmed.1002119>

Kenney, J. F., & Keeping, E. S. (1962). Root mean square. *Mathematics of Statistics*, 1, 59–60.

Kloog, I., Ridgway, B., Koutrakis, P., Coull, B. A., & Schwartz, J. D. (2013). Long- and short-term exposure to PM2.5 and mortality: Using novel exposure models. *Epidemiology*, 24(4), 555–561. <https://doi.org/10.1097/EDE.0b013e318294beaa>

Kotra, L. P. (2007). Infectious Diseases. In *xPharm: The Comprehensive Pharmacology Reference* (pp. 1–2). Elsevier. <https://doi.org/10.1016/B978-008055232-3.60849-9>

Parry, W., & Peterson, E. (2020). *28 devastating infectious diseases*. Live Science. <https://www.livescience.com/13694-devastating-infectious-diseases-smallpox-plague.html>

Snow, K., Yadav, R., Denholm, D., Sawyer, S., & Graham, S. (2018). Tuberculosis among children, adolescents and young adults in the Philippines: A surveillance report. *Western Pacific Surveillance and Response Journal*, 9(4), 16–20. <https://doi.org/10.5365/wpsar.2017.8.4.011>

Tucay Quezon, E., & Ibanez, A. G. (2021). Effect of covid-19 pandemic in construction labor productivity: a quantitative and qualitative data analysis. *American Journal of Civil Engineering and Architecture*, 9(1), 23–33. <http://pubs.sciepub.com/ajcea/9/1/4>

Wells, W. A., & Stallworthy, G. (2019). Meet the patients where they are: A greater ambition for private provider engagement for TB. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, 14, 14–15. <https://doi.org/10.1016/j.jctube.2018.12.002>

World Health Organization. (2019). *Global tuberculosis report 2019*. <https://rb.gy/ggfbh>

World Health Organization. (2020). *Global tuberculosis report 2020*. <https://rb.gy/xl45h>

Yan, M., Wilson, A., Bell, M. L., Peng, R. D., Sun, Q., Pu, W., Yin, X., Li, T., & Anderson, G. B. (2019). The shape of the concentration–response association between fine particulate matter pollution and human mortality in Beijing, China, and its implications for health impact assessment. *Environmental Health Perspectives*, 127(6), 1–13.

Zhang, X., Andersen, A. B., Lillebaek, T., Kamper-Jørgensen, Z., Thomsen, V. Ø., Ladefoged, K., Marrs, C. F., Zhang, L., & Yang, Z. (2011). Effect of sex, age, and race on the clinical presentation of tuberculosis: A 15-year population-based study. *American Journal of Tropical Medicine and Hygiene*, 85(2), 285–290. <https://doi.org/10.4269/ajtmh.2011.10-0630>