MATH*unesa*
Jurnal Ilmiah Matematika
e-ISSN : 2716-506X | p-ISSN : 2301-9115
Volume 11 No 01
Tahun 2023

MORTALITY BASELINE MODEL USING LINEAR TIME SERIES AND LINEAR MIXED MODEL IN EXCESS MORTALITY CALCULATION DURING COVID-19 IN DKI JAKARTA

Maulia Dita Agistia
Magister Manajemen, Fakultas Ekonomi dan Bisnis, Universitas Indonesia
email: maulia.dita@ui.ac.id

Lenny Suardi
Magister Manajemen, Fakultas Ekonomi dan Bisnis, Universitas Indonesia
email: lennys@ui.ac.id

Abstract

COVID-19 pandemic has affected the mortality across a globe. To investigate the impact of COVID-19, many countries have recorded number of deaths due to COVID-19. In Indonesia, DKI Jakarta reports the highest number of mortality due to COVID-19. However, the reported data may have discrepancy, for example the scope of testing for COVID-19 that has not been widely implemented, false-negative on testing results and deaths that occur before COVID-19 test. The measurement of excess mortality has been suggested to cover the lack of data. Baseline mortality will be the main component in calculating excess mortality. Monthly deaths data of DKI Jakarta from January 2018 up to February 2021 will be used to generate baseline mortality model. The analysis will compare two models, linear time series and linear mixed model. Model accuracy will be calculated to choose the better baseline mortality model. The better model of baseline mortality will give better estimation of excess mortality during COVID-19. Linear time series provides a better accuracy on baseline mortality model in DKI Jakarta. The result shows that there are 25,553 excess mortality during COVID-19 pandemic in DKI Jakarta from June 2020 until June 2021. The SMR during pandemic COVID-19 is around 133%.

Keywords: *Linear time series, linear mixed model, excess mortality, standardized mortality rate.*

INTRODUCTION

The pandemic has been associated with negative impacts on various aspect, especially health. There have been many significant disease outbreaks and pandemics recorded in history, including Spanish Flu, Hong Kong Flu, SARS, H7N9, Ebola, Zika (Qiu et al., 2017). In December 2019, Coronavirus Disease 2019 (COVID-19) pandemic was first identified in Wuhan, China. It has affected more than 200 countries around the world, including Indonesia (Setiati & Azwar, 2020). World Health Organization (WHO) was declared COVID-19 as the pandemic since March 11, 2020. As of June 20, 2022, there were 6,316,655 deaths due to COVID-19 in the world.

As of June 20, 2022, there were 156,695 deaths due to COVID-19 in Indonesia, of which 15,309 were from DKI Jakarta. Based on data from the Ministry of Health of Indonesia, most of the positive confirmed cases and deaths due to COVID-19 in Indonesia have occurred in the DKI Jakarta. This can happen due to several factors such as the process of tracing and recording is much better compared to other provinces, DKI Jakarta is the national capital and is the center of business in Indonesia and DKI Jakarta as the largest transmission center.

The increasing mortality due to COVID-19 is in line with the increasing of death claim in insurance company. Deaths due to COVID-19 can be insured by Insurance company in Indonesia. Indonesian Life Insurance Association (AAJI) recorded an increase in death claims at life insurance companies in Indonesia by 17.1% for 2020 and 65.7% for 2021. The increasing of death claims causes the higher exposure of risk.

To manage the risk, most countries record daily active cases and daily deaths due to COVID-19 as their monitoring strategy. The insurance companies record deaths due to COVID-19 and non-COVID-19 regularly for their analysis to manage the mortality risk. However, reported deaths due to COVID-19 depend on the completeness and method of counting deaths (Verbeeck et al., 2021). The reports are prone to misreporting (Wijaya, 2022). There are five reasons that caused gaps between reported and excess deaths related to the COVID-19 (Wang et al., 2022):

a. Healthcare reporting systems do not list COVID-19 as the cause of death without a positive SARS-CoV-2 test in locations with low testing capabilities
b. Country systems for registering deaths differ in quality and comprehensiveness, including the definitions used to count the deaths due to COVID-19
c. There is no universal agreement regarding when the death of someone infected with SARS-CoV-2 must be reported as a death from COVID-19
d. Political considerations that may hinder the reporting of deaths in several locations
e. Other causes of death that may have changed due to social, economic and behavioral responses to the pandemic, including strict lockdown policies

Therefore, excess mortality calculation has been suggested to assess the potential of inaccuracy mortality data during COVID-19 pandemic. Excess mortality due to the COVID-19 pandemic, defined as the net difference between the number of deaths during the pandemic (measured by observed or estimated all-cause mortality) and the number of deaths that would be expected on the basis of past trends in all-cause mortality (Wang et al., 2022). Excess mortality will give more accuracy measurement to assess the risk and provide policies that are in accordance with actual conditions. This article will assess the best estimate of excess mortality during COVID-19 pandemic between two models of baseline mortality.

LITERATURE REVIEW

The statistical modeling of lifetimes is a main topic in actuarial science, as it has consequences for the pricing and risk management of insurance and pension products (Albrecher et al., 2022). The statistical modeling is related to the modeling of mortality or death rates. According to the Centers for Disease Control and Prevention (CDC), the mortality rate is a measure of the frequency of the number of deaths in a population in a certain period of time.

There are several various types of mortality measurement, including the Crude Death Rate (CDR) and Case Fatality Rate (CFR) (CDC, 2022). Both of these measures, CDR and CFR, are calculated based on observed data. Hence the value of CDR and CFR can change quickly based on the daily observed data. It is important to use other mortality measurement to give more comprehensive analysis about COVID-19 pandemic. Several studies have been suggested the excess mortality to assess the overall impact on mortality due to pandemic.

The pattern of past mortality before the pandemic is one of the parameters in excess mortality calculation. Hence, the mortality model prior to the pandemic will be a critical part in calculating the excess mortality. The expectations from the mortality model before the pandemic is called the baseline mortality model. A simple model to calculate baseline mortality is using the weekly average of the past 5-years mortality based in historical data (Verbeeck et al., 2021). However, the methodology ignores year-specific trends in mortality.

The other baseline modeling in the calculation of excess mortality during COVID-19 pandemic has been made in some recent studies. Some studies using linear time series for the baseline mortality model because it utilizes serial correlation in historical mortality data to project base mortality rates into the future (International Actuarial Association., 2010). International Actuarial Association (IAA) applies the linear time series for baseline mortality model by using mortality data from USA. The result showed that the baseline model does not project large increases in mortality rates. Another method used to estimate baseline mortality model is the linear mixed model. Linear mixed models offer the flexibility to model longitudinal data with the inclusion of mean and variability structures (Wijaya, 2022). Verbeeck conducted baseline mortality modeling using the linear mixed model method and 5-year weekly average in calculating excess mortality due to COVID-19 for Belgian and Dutch data. The result showed that the linear mixed model improves the estimation of excess mortality in Belgium and Netherlands. Wijaya also used a linear mixed model for modeling baseline mortality in calculating excess mortality for COVID-19 for DKI Jakarta data. The result showed 13,507 excess mortality in DKI Jakarta between June 2020 until December 2020.

In this study, we propose a comparative analysis between Linear Time Series and Linear Mixed Model for baseline mortality model in calculating excess

mortality during COVID-19 in DKI Jakarta. It is important to make comparisons in baseline mortality modeling in order to obtain a better estimate of excess mortality due to the COVID-19 pandemic. An accurate excess mortality can show the real impact of the COVID-19 pandemic related to mortality in DKI Jakarta. By utilizing the accuracy test, Root Mean Square Error (RMSE), %RMSE and Mean Absolute Percentage Error (MAPE), a better method can be identified in modeling baseline mortality. Further, we will use Standardized Mortality Rate (SMR) to assess whether there is an excess in mortality during the pandemic in DKI Jakarta or not.

METHODS

Open source all-cause DKI Jakarta mortality data from January 2018 up to June 2021 are available from Open Data Jakarta. Total number of data is 42 monthly basis data. These data provide the number of deaths that classified by gender and six administrative cities in DKI Jakarta (West Jakarta, Central Jakarta, South Jakarta, East Jakarta, North Jakarta and Administrative of Kepulauan Seribu). The data from January 2018 up to February 2020 will be used to estimate the baseline mortality model as non-pandemic period.

First method is linear time series. By linear time series, historical mortality rates are used to calibrate the time series model (International Actuarial Association., 2010). The mortality rate can be defined as below:

$$q_t = \frac{d_t}{l_t} \tag{1}$$

where d_t is the number of deaths recorded in the population for time t and l_t is the number of lives in the population for time t. The time series of the mortality rate will generate a pattern that provide information regarding the projected future data. The linear time series method uses the assumption that the data being analyzed is stationary, that is, the mean and variance are constant (Cryer & Chan, 2008). Linear autoregressive is formulated as follows:

$$y_t = q_0 + q_1\, y_{t-1} + \cdots + q_n\, y_{t-n} + \varepsilon_t \tag{2}$$

where y_t is the change in mortality rate from time $t-1$ to time t and ε_t is error. The parameters will be estimated by using Maximum Likelihood Estimation (MLE), which are selected to maximize the likelihood function.

There are three main steps in the time series process (Cryer & Chan, 2008). First step is model specification to identified the observed series data. Second step is model fitting to find the best possible parameters estimator. Third step is model diagnostic to assess the quality of the model that has been generated.

The second method is linear mixed model. The linear mixed model is an extension of the simple linear model to analyze data that are not mutually independent, have levels, or are correlated (Wijaya, 2022). The formula of linear mixed model:

$$y_{tj} = \begin{cases} g_0 + b_j + f(t) + \varepsilon_{tj}, & \text{for male} \\ g_1 + b_j + f(t) + \varepsilon_{tj}, & \text{for female} \end{cases} \tag{3}$$

where y_{tj} is mortality related to each group j in time t; g_0, g_1 are regression constants for male and female, respectively; b_j is random constants for each group j; $f(t)$ is time function to capture the time effect; and ε_{tj} is error. The parameters will be estimated by using Restricted Maximum Likelihood (REML). REML perform better by removing the finite-sample bias than Maximum Likelihood Estimators (MLE) (Wijaya, 2022).

RMSE and MAE are commonly applied in model evaluation studies (Chai & Draxler, 2014). The result of the two model will be assessed by using RMSE, %RMSE (RMSE percentage variant) and MAPE (MAE percentage variant). The formula as below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_t - y_t)^2} \tag{4}$$

$$\%RMSE = \frac{\frac{1}{n}\sqrt{\sum_{i=1}^{n}(\hat{y}_t - y_t)^2}}{\frac{1}{n}\sum_{i=1}^{n} y_t} \times 100 \tag{5}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{y}_t - y_t|}{y_t} \times 100 \tag{6}$$

where $\hat{y}_t$ is predictive mortality in time t; y_t is actual mortality in time t; n is number observation data. RMSE, %RMSE and MAPE values are in the range 0 to ∞. The smaller the RMSE, %RMSE and MAPE, the better model will be. In other words, smaller RMSE,

%RMSE and MAPE values will give the best model among the models that being analyzed.

After knowing the best model for baseline mortality model, the next step is calculating the excess mortality. Excess mortality is defined as the difference between reported death and expected death. Expected death is generated by the baseline mortality model that has been performed earlier. To know the significance of the excess in mortality during COVID-19, this study will calculate the Standardized Mortality Rate (SMR). SMR is one of the principal measure of excess mortality (England & Haberman, 1993). SMR is the ratio of the number of deaths observed (y) to the number of deaths estimated (y') (Klein & Moeshberger, 1984). If the SMR value is more than one, then it can be said that there is excess mortality due to COVID-19 in DKI Jakarta.

RESULTS AND DISCUSSIONS

DESCRIPTIVE STATISTICS

Based on data from the Badan Pusat Statistik (BPS), the total population of DKI Jakarta as of June 2021 reached 10.6 million people. It is also known that East Jakarta is the region with the most population with a total population of around 3 million lives and Kepulauan Seribu is the region with the least population with a total population of only 28 thousand lives.

Central Jakarta, South Jakarta, East Jakarta and North Jakarta. This is to obtain more appropriate prediction modeling results related to the calculation of excess mortality in DKI Jakarta.

Table 1. Descriptive Statistics of Number of Death in DKI Jakarta

	West Jakarta	Central Jakarta	South Jakarta	East Jakarta	North Jakarta	Kep. Seribu	DKI Jakarta
Minimum	424	239	464	695	320	4	2,150
Median	1,150	621	1,120	1,508	881	15	5,228
Mean	1,222	691	1,200	1,676	928	17	5,734
Maximum	1,942	1,130	2,070	2,847	1,452	46	9,224

Mortality rate among the 5 administrative cities also have similar patterns. There was a decrease in the mortality rate in April 2020. This is suspected due to a transitional period of Large-Scale Social Restrictions (PSBB) policy implemented in DKI Jakarta, resulting in delays in reporting.

BASELINE MORTALITY MODEL

Baseline mortality model will be compared between two methods, linear time series and linear mixed model. Mortality data from January 2018 to February 2020 will be observed and simulated using RStudio.

There are three steps in linear time series modeling, as follows:

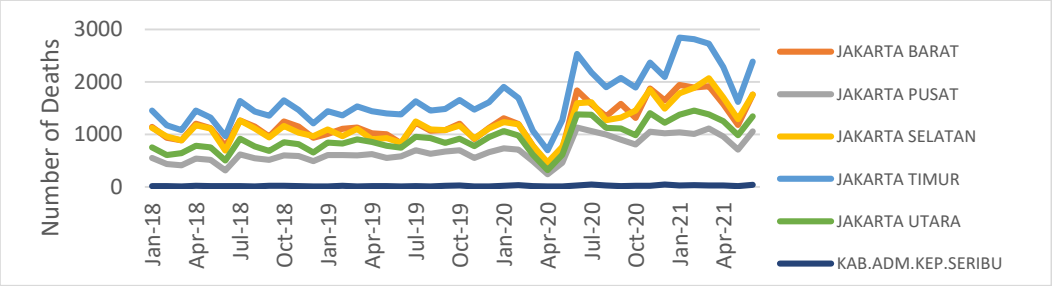

Figure 1. Number of Deaths among 6 Regions in DKI Jakarta

There is a similar pattern of the number of deaths between the 5 cities in DKI Jakarta (Figure 1). However, the number of deaths' pattern of Kepulauan Seribu cannot clearly visible due to the small number of deaths compare to others regions (Table 1). Hence, this research will observe only for 5 administrative cities in DKI Jakarta: West Jakarta,

a. Model identification

Linear time series requires data to be stationary. However, mortality rate in DKI Jakarta is not stationary (Figure 2). It needs to be differentiated once to make it stationary (Figure 3).

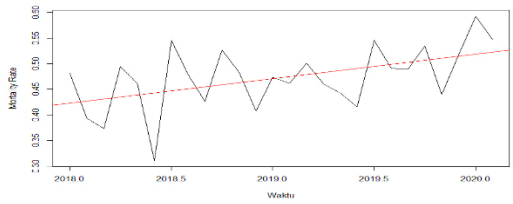

Figure 2. Mortality rate in DKI Jakarta

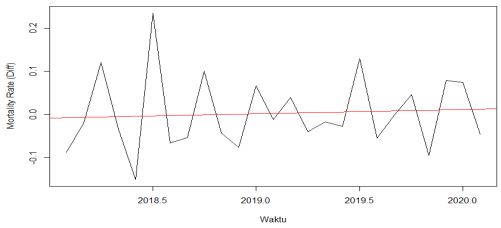

Figure 3. First Differentiation of Mortality Rate in DKI Jakarta

b. Model fitting

DKI Jakarta mortality rate, that has been differentiated once, forms Autoregressive model with a lag of two months or denoted as AR(2). It can be interpreted that the current mortality rate will be affected by two previous periods (monthly) mortality rates. Using MLE parameter estimator, the AR(2) equation as follows:

$$\hat{y}_t = -0.7848y_{t-1} - 0.6920y_{t-2} \qquad (7)$$

The comparison of actual and model is showed by Figure 4.

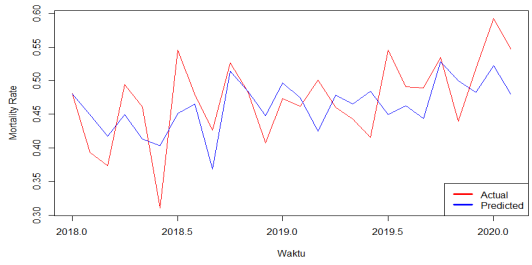

Figure 4. Actual vs Model Estimation of Mortality Rate in DKI Jakarta

c. Model diagnose

To evaluate whether the model is adequate or not, normal plot will be analyzed (Figure 5). The error is around the standard normal line, hence it can be concluded that the model fits the standard normal distribution error assumption and the adequacy model is fulfilled.

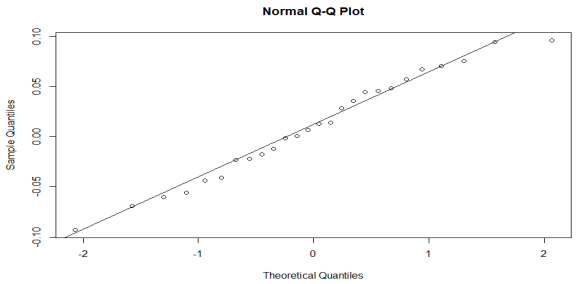

Figure 5. Normal Plot

Linear mixed model does not require the stationary in data. In this model, the number of monthly deaths per regions and per gender from January 2018 to February 2020 will be observed. Using REML parameter estimator, the linear mixed model equation as follows:

$$\hat{y}_{tj} = \begin{cases} 505{,}6181 + b_j + 4{,}4330t, & \text{for male} \\ 367{,}0305 + b_j + 4{,}4330t, & \text{for female} \end{cases} \qquad (8)$$

Table 2. $\boldsymbol{b_j}$ coefficients for each regions $\boldsymbol{j}$

j	b_j
West Jakarta	477.8696
Central Jakarta	229.5598
South Jakarta	465.1466
East Jakarta	663.4839
North Jakarta	345.5616

Linear mixed model gives prediction in linear form (Figure 6).

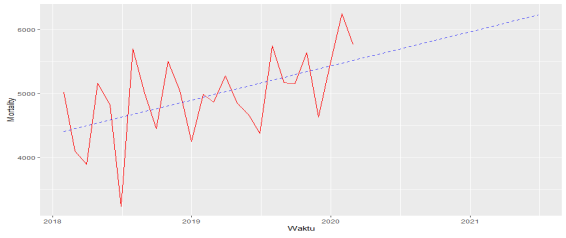

Figure 6. Actual vs Model Estimation of Mortality in DKI Jakarta using Linear Mixed Model

To assess the best model, accuracy test calculations were performed using RMSE, %RMSE and MAPE. The accuracy test was carried out on the actual value and the estimated value from the mortality baseline model for the period January 2018 to February 2020. RMSE, %RMSE and MAPE of linear time series are less than linear mixed model (Table 3). It shows that linear time series gives better estimation in mortality baseline model than linear mixed model. Hence, we will apply linear time series to predict the mortality baseline model and evaluate the excess mortality during COVID-19.

Table 3. Accuracy Test of Linear Time Series and Linear Mixed Model

Accuracy Test	Linear Time Series	Linear Mixed Model
RMSE	546.92	552.97
%RMSE	2.16	2.19
MAPE	9.56	9.64

EXCESS MORTALITY AND SMR

Number of death from March 2020 to May 2020 shows outlier values. This could occur due to the Large-Scale Social Restrictions (PSBB) policy. Therefore, the excess mortality that will be calculated is for the period June 2020 to June 2021. Based on data from Open Data Jakarta, total deaths recorded in DKI Jakarta was 101,446 people from June 2020 to June 2021. However, the prediction results of AR(2), the total number of deaths is estimated at 75,893 people. It indicates excess mortality during the COVID-19 pandemic is 25,553 people.

It is also known that based on data recorded on the Jakarta Responsive COVID-19 website, there were only 7,966 deaths in DKI Jakarta as of June 2020 to June 2021. This could be an indication that the data recorded is underestimated. Further, we calculate the SMR is 133.67%. It can be concluded that the COVID-19 Pandemic in DKI Jakarta in the period June 2020 to June 2021 has caused an increase in mortality of 33.67% from normal conditions.

CLOSING

CONCLUSIONS

This article presents that the linear time series is more suitable than linear time series for baseline mortality model in DKI Jakarta. The estimate of baseline mortality using linear time series proves the existence of excess mortality during COVID-19 pandemic in June 2020 to June 2021 in DKI Jakarta. The excess mortality is 25,553 people and the SMR is 133.67%. This research is expected to provide an additional insight in pandemic research. For example, additional premium for life insurance product during pandemic or future outbreak.

RECOMMENDATION

For further research, baseline mortality model and excess mortality calculation should be improved with other methods that consider more factors such as age band or health services. Other research related to pandemics can also be analyzed by considering other pandemics that have occurred around the world. The pandemic research will help people to manage the pandemic risk better.

BIBLIOGRAPHY

Albrecher, H., Bladt, M., Bladt, M., & Yslas, J. (2022). Mortality modeling and regression with matrix distributions. *Insurance: Mathematics and Economics*, *107*, 68–87. https://doi.org/10.1016/j.insmatheco.2022.08.001

CDC. (2022). *Measures of risk search*.

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? - Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Cryer, J. D., & Chan, K.-S. (2008). *Time series analysis*. Springer.

England, P. D., & Haberman, S. (1993). *A new approach to modeling excess mortality*. http://digitalcommons.unl.edu/joaphttp://digitalcommons.unl.edu/joap/170

International Actuarial Association. (2010). *Stochastic modeling : theory and reality from an actuarial perspective*. AAI, IAA.

Klein, J. P., & Moeshberger, M. L. (1984). Survival Analysis. In *Springer* (Vol. 19, Issue 5).

https://doi.org/10.1145/390011.808243

Qiu, W., Rutherford, S., Mao, A., & Chu, C. (2017). The pandemic and its impacts. *Health, Culture and Society*, *9*, 1–11. https://doi.org/10.5195/hcs.2017.221

Setiati, S., & Azwar, M. K. (2020). *COVID-19 and Indonesia*. *April*.

Verbeeck, J., Faes, C., Neyens, T., Hens, N., Verbeke, G., Deboosere, P., & Molenberghs, G. (2021). A linear mixed model to estimate COVID-19-induced excess mortality. *Biometrics*. https://doi.org/10.1111/biom.13578

Wang, H., Paulson, K. R., Pease, S. A., Watson, S., & Comfort, H. (2022). Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020-21. *Lancet (London, England)*, *399*(10334), 1513–1536. https://doi.org/10.1016/S0140-6736(21)02796-3

Wijaya, M. Y. (2022). The estimation of excess mortality during the COVID-19 pandemic in Jakarta, Indonesia. *Kesmas*, *17*(1), 25–31. https://doi.org/10.21109/kesmas.v17i1.5413