Proceedings of the 10th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/2026.32955

Time Series Forecasting of Diabetes Mortality in the United States Based on ARIMA

Yueyue Tian

Department of Business Analytics, University of California, San Diego, USA
yut023@ucsd.edu

Abstract. It looks at diabetes death data in the U.S. for certain age groups from 1999 to 2020. The goal of this article is to identify the patterns of mortality rate over time and whether past patterns can be applied to predict what will happen in the future. Looking first at the data showed that the time series was not stationary, which meant that it was unable to be modeled directly. Data was differenced before the model was built to fix this problem. This paper chose the ARIMA(0,1,0) model after looking at a number of other options. Although this model was pretty simple, it fit the dataset well. It accurately showed the main trends in the data without making things more complicated than they needed to be. One interesting finding about the data is that the number of deaths went up a lot in 2020. It doesn't fit with how matters have changed in the past, which makes it stand out as a major anomaly.

Keywords: Diabetes, Time Series Analysis, ARIMA, Mortality Forecasting, Public Health Modeling

1. Introduction

Diabetes has been accepted as a huge health problem, both in the USA and the whole world. Many patients have serious base diseases, such as heart or kidney diseases, that increase their risk of death [1]. Researchers have analyzed diabetes from multiple parts. Earlier research showed that individuals with diabetes is show more higher death rates in contrast to the general population [2]. Some researchers focus on finding key risk factors, while some apply statistical or machine learning methods to detect high-risk populations [3]. Time series analysis enables the identification of long-term trends and allows for forecasting based on historical data [4]. Scholars often use the ARIMA model, because it is easy to use and has a structure that is not too hard to understand. This model is often used in public health and in other fields as well, like predicting the economy [5].

2. Methodology

This paper selects the age adjusted mortality data of diabetes in the United States from 1999 to 2020 as the research sample. The analysis focuses on long-term changes rather than short-term ones because the data is collected every year. ARIMA was selected; it has been used a lot before and

DOI: 10.54254/2754-1169/2026.32955

works well with this kind of data [6]. The ARIMA model is usually written as ARIMA(p, d, q), where each letter stands for a different part of the model's structure [7].

This article obtained the information used in this analysis from the National Center for Health Statistics (NCHS), which keeps official death records [8]. Time series analysis was appropriate because the dataset spanned multiple years.

The author must check that the data are stationary before using the model. For a general idea of what the data looked like, they were first plotted. The plot shows that the series level changes over time, which means that the data might have not be stable. To further confirm this, the author has the Augmented Dickey-Fuller (ADF) test was done [9].

The test results indicated that the original series was non-stationary, necessitating a first-order differencing step. Next, a bunch of Autoregressive Integrated Moving Average (ARIMA) models were compared, and the final choice was mostly based on things with Akaike's Information Criterion (AIC).

3. Results and discussion

3.1. Time series characteristics

Figure 1 shows the age-adjusted diabetes death rate in the United States from 1999 to 2020. Not all groups of people saw this trend happen in the same way. There was a small rise in the early 21st century, but from 2003 to 2010, the death rate slowly went down. After that, the time series seemed to level off, though there were still small changes. This time of relative stability could be connected to better treatment options or standards for managing diseases, but the author cannot be sure of that just from this one observation. Also, there may be other factors that affect the results that are not included in this dataset. The increase in the death rate seen in 2020 is an especially surprising piece of information. This phenomenon may have been affected by external factors. Previous research has shown that people with long-term illnesses, like diabetes, are more likely to be affected by public health emergencies [5]. The changes in the mean level of this time series indicate that the raw data may not be stationary; this was later verified by the Augmented Dickey-Fuller (ADF) test. Therefore, before constructing the ARIMA model in this article, the data was first subjected to first-order differencing.

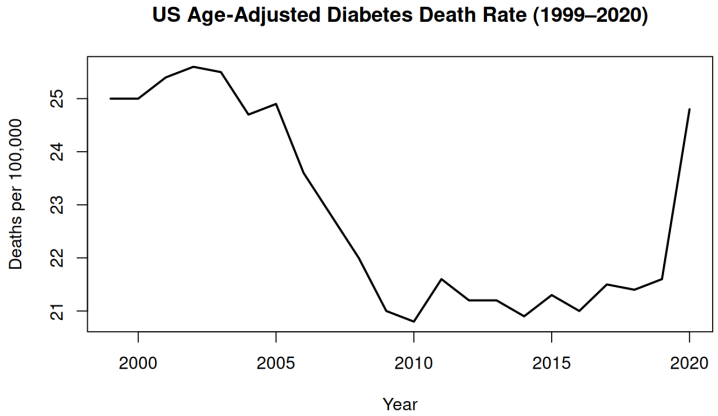

Figure 1. US age-adjusted diabetes death rate (1999–2020)

DOI: 10.54254/2754-1169/2026.32955

3.2. ARIMA model estimation

The data looked more stable and better for modeling after differencing. In conducting the analysis, this article studied several different ARIMA models. Some of them were more complicated, but they didn't seem to make the model work better in a clear way. Thus, a simpler method was chosen. This article chose the ARIMA(0,1,0) model because it was simple enough to capture the main features of the data without adding any extra complexity. The model is not too complicated, but it does fit the general pattern pretty well (see Table 1).

Table 1. ARIMA results

Model	p	d	q	AIC	BIC
ARIMA(0,1,0)	0	1	0	56.63	57.68

3.3. Residual diagnostics

Residual diagnostics were performed to ascertain if the fitted ARIMA model sufficiently represented the temporal structure of the diabetes mortality series. A properly defined time series model ought to yield residuals that exhibit characteristics of white noise, signifying they are independently distributed with a stable mean and variance over time [8]. Figure 2 shows the residual time plot, the autocorrelation function (ACF) of the residuals, and the histogram of how the residuals are spread out. The residual time plot shows that the residuals move around zero in a random way, with no clear pattern. This means that the model has successfully found the underlying trend in the data. The residual ACF also doesn't show any big spikes outside of the confidence bounds, which means that there is no more autocorrelation in the residual series. To further check if the model was good enough, the Ljung–Box test was used to look for residual autocorrelation [10]. The findings suggest that the null hypothesis of no autocorrelation cannot be dismissed, indicating that the residuals approximate white noise and that the ARIMA model sufficiently represents the observed mortality series.

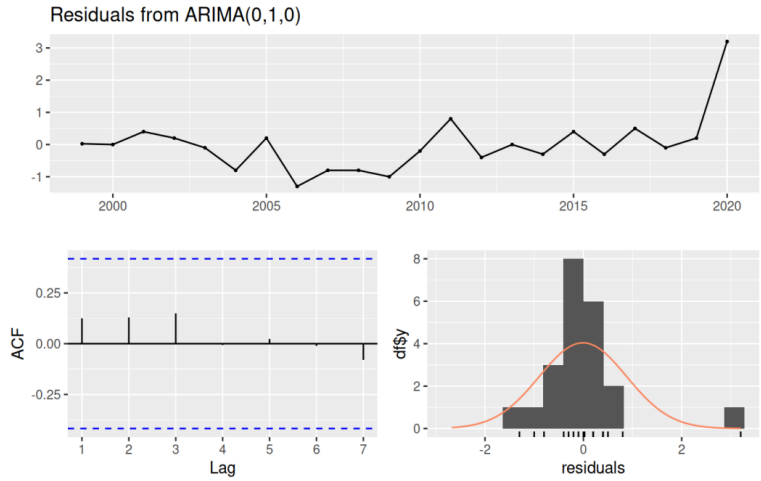

Figure 2. Residual diagnostics for the fitted ARIMA (0,1,0) model

DOI: 10.54254/2754-1169/2026.32955

3.4. Forecasting results

It used the fitted ARIMA(0,1,0) model to make predictions for the years 2021 to 2025 (see Figure 3 and Table 2). The expected death rate stays the same at about 24.8 cases per 100,000 people, with only small changes over time.

This result isn't surprising. After differencing, the model behaves like a random walk, indicating that subsequent values are expected to be near the latest observations. So, the forecast mostly shows where the time series is right now and not any big changes in direction.

But the big jump in 2020 makes it harder to make sense of these results. It's still hard to tell if this change will last or if it will only last for a short time. This suggests that the model might not fully show how uncertain things will be in the future.

Previous research indicates that individuals with chronic illnesses may face higher risks during public health emergencies [5], and additional research has yielded analogous results [11]. These things make it even harder to make sense of the most recent data.

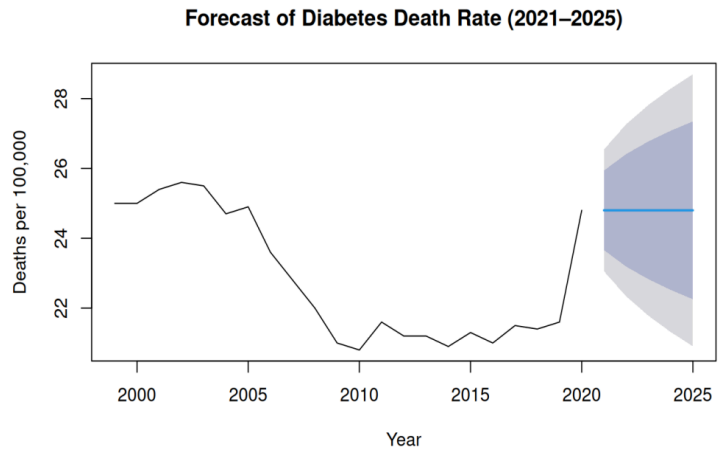

Figure 3. The ARIMA (0,1,0) model predicts diabetes death rates in the US from 2021 to 2025

Table 2. Estimated death rates from diabetes (2021–2025)

Year	Forecast
2021	24.8
2022	24.8
2023	24.8
2024	24.8
2025	24.8

3.5. Effects on policy

The findings of this study have numerous implications for policies designed to mitigate diabetes-related mortality in the United States. Early screening and timely treatment should be prioritized, especially for older adults and those who are overweight, as early intervention can management the risk of severe complications.

DOI: 10.54254/2754-1169/2026.32955

It is also important to have access to healthcare in order to manage diabetes well. Patients can keep their blood sugar levels stable and stop the disease from getting worse by taking some more cheaper and normal medicine, going to the doctor same time, and getting professional advice. Promoting healthier ways of living, such as eating a balanced diet and getting regular exercise, is also important.

Some people who have long-term health problems may be very affected by sudden changes in the public health environment. In these situations, better emergency response plans and stronger healthcare systems can help keep special groups safe and lower the number of deaths that happen during unexpected events.

4. Conclusion

A time series methodology was utilized to analyze diabetes mortality trends in the United States from 1999 to 2020. Even death rates seem to have been stable in recent years, but the rise in 2020 makes the pattern unclear. The ARIMA(0,1,0) model, which is a simple way to explain the data and make short-term predictions, but it doesn't do a good job of picking up on changes that weren't expected. When looking at the results, people should keep this in their mind.

There are also a number of problems. This data only has 22 annual observations in the dataset, which limits the amount of detail that can be captured. Yearly data may also miss short-term changes. Future research might investigate higher-frequency data or other models, and external factors may require more thorough consideration. In general, the results give a broad picture of death trends, but they should be taken with a grain of salt.

References

[1] Zheng, Y., Ley, S. H., & Hu, F. B. (2018). Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. Nature Reviews Endocrinology, 14(2), 88–98.

[2] Gregg, E. W., Li, Y., Wang, J., Burrows, N. R., Ali, M. K., Rolka, D., Williams, D. E., & Geiss, L. (2014). Changes in diabetes-related complications in the United States, 1990–2010. New England Journal of Medicine, 370(16), 1514–1523.

[3] Sisodia, D. S., & Sisodia, D. S. (2018). Prediction of diabetes using classification algorithms. Procedia Computer Science, 132, 1578–1585.

[4] Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. Computational and Structural Biotechnology Journal, 15, 104–116.

[5] Barron, E., Bakhai, C., Kar, P., Weaver, A., Bradley, D., Ismail, H., Knighton, P., Holman, N., Khunti, K., Sattar, N., Wareham, N. J., Young, B., & Valabhji, J. (2020). Associations of type 1 and type 2 diabetes with COVID-19-related mortality in England. The Lancet Diabetes & Endocrinology, 8(10), 813–822.

[6] Shumway, R. H., & Stoffer, D. S. (2017). Time series analysis and its applications: With R examples (4th ed.). Springer.

[7] Brockwell, P. J., & Davis, R. A. (2016). Introduction to time series and forecasting (3rd ed.). Springer.

[8] Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: principles and practice. International Journal of Forecasting, 34(3), 555–560.

[9] Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association, 74(366), 427–431.

[10] Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. Biometrika, 65(2), 297–303.

[11] Zhang, X., Zhang, T., Young, A. A., & Li, X. (2014). Applications and comparisons of four time series models in epidemiological surveillance data. PLoS ONE, 9(2), e88075.