NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

BIOSTATISTICS DEPARTMENT

TIME SERIES ANALYSIS OF 2015-2019 ALZHEIMER'S DEATHS RECORDS IN THE UNITED STATES, AND FORECASTING (2020-2024)

.

VIDAL FRANCK YMELONG GHOKENG

MASTER THESIS

NICOSIA

(2021)

TIME SERIES ANALYSIS OF 2015-2019 ALZHEIMER'S DEATHS RECORDS IN THE UNITED STATES, AND FORECASTING (2020-2024)

VIDAL FRANCK YMELONG GHOKENG

NEAR EAST UNIVERSITY INSTITUTE OF GRADUATE STUDIES

FACULTY OF MEDICINE

DEPARMENT OF BIOSTATISTICS

MASTER THESIS

THESIS SUPERVISOR

PROF. DR. ILKER ETIKAN

NICOSIA

(2021)

ACCEPTANCE/APPROVAL

We as the jury members certify the “Time Series Analysis of 2015-2019 Alzheimer’s Deaths Records in the United States, and Forecasting (2020-2024)”.Prepared by Vidal Franck YMELONG GHOKENG on **.....**has been found satisfactory for the award of Degree

of Masters.

Thesis Committee;

Chair of the committee: **Prof. Dr. İlker ETİKAN (Advisor)**
Near East University
Sig: …….

Member: **Prof. Dr. Zeki AKKUŞ**
DicleÜniversitesi
Sig: …….

Member: **Asst. Prof. Dr. Özgür TOSUN**
Near East University
Sig: …….

Approved by: Prof. Dr. K. Hüsnü Can BAŞER
Director of Institute of Graduate Studies
Near East University
Sig: …….

Statement (Declaration)

Hereby I declare that this thesis study is my own study, I had no unethical behavior in all stages from planning of the thesis until writing thereof, I obtained all the information in this thesis in academic and ethical rules, I provided reference to all of the information and comments which could not be obtained by this thesis study and took these references into the reference list and had no behavior of breeching patent rights and copyright infringement during the study and writing of this thesis.

O The full extent of my Thesis can be accessible from anywhere.

O My Thesis can only be accessible from Near East University.

O My Thesis cannot be accessible for two (2) years. If I do not apply for extension at the end of this period, the full extent of my Thesis will be accessible from anywhere.

Date:

Signature:

Name, Surname: Vidal Franck YMELONG GHOKENG

Acknowledgement

We obviously can't present this research project without expressing our gratitude to those who near or far away have participated to it. They are:

- God, the Almighty, for all the blessings he granted me during this research work.
- My academic supervisor Pr. Dr. ILKER ETIKAN, head of the department of Biostatistics in Near East University; for his guidance, his availability at any time to answer my numerous questions, and his strong support. They only teacher to whom you can sent an email at 00:30 and he replies to you 5 minutes later. You are just the best Sir, you have not only participated to this research project, you have also greatly impacted my future carrier.
- Asst. Pr. Dr. ÖzgürTosun of the department of Biostatistics in Near East University; for all his assistance and intervention during the analysis of this research data. The one that makes me love Biostatistics, always available to direct students, no matter if he is their academic supervisor or not. Stay blessed sir.
- Asst. Pr. Dr. Etienne GNIMPIEBA ZOHIM of the department of Biomedical Engineering in the University of South Dakota; for his financial support, and his assistance during the implementation of this research. No matter the jet lag, as soon as you notify that it is an emergency, he would reply to you. Thank you, my dear cousin, to have pushing me to this level, you have all my gratitude.
- My lovely parents without who, I wouldn't have reached this level of study. You have always told me that I will be your doctor. For all the difficulties I faced through the implementation of this project, that statement makes me to never give up and push to the best I could. This is your project; I will always love you.
- My two brothers Cedric JIONGO and Serge TONLIO, also MSc student in the department of Biostatistics of Near East University for the good atmosphere that was always present at home during this project writing, it has really helped my brain to think efficiently.
- My family for all the motivation and the psychologic assistance that the provided to me. Just the statement "it is a small thing" that all of you use to say whenever I was calling you, given though you did not know anything about Biostatistics has helped me a lot.

Contents

Abbreviation And Symbol List

AD: Alzheimer's disease

TS: Time Series

TSA: Time Series Analysis

TSF: Time Series Forecasting

FDA: Food and Drugs Administration

CDC: Center for Disease Control

ACF: Autocorrelation Function

PACF: Partial Autocorrelation Function

CI: Confidence Interval

UCL: Upper Confidence Limit

LCL: Lower Confidence Limit

OR: Odd Ratio

ES: Exponential Smoothing

AR: Autoregressive

MA: Moving Average

BIC: Bayesian inclusion criteria

AIC: Akaike information criterion

List of illustration

- **List of Figures**

- **List of Tables**

Öğrenci: VIDAL FRANCK YMELONG GHOKENG.
Mentor: İLKER ETİKAN
Bölüm: Biyoistatistik

Türkçe Özet

Amaç: Çalışmanın amacı, popülasyondaki varyasyonu karakterize eden sürdürülebilir bilgi elde etmek için ABD'deki 2015-2019 Alzheimer ölüm kayıtları zaman serisi üzerinde istatistiksel testler yapmak; ve gelecek yıllar için ölümlerin sayısını tahmin etmektir.

Gereç ve Yöntem: Veriler, Hastalık Kontrol ve Önleme Merkezleri (CDC) WONDER veri tabanından çevrimiçi olarak elde edildi. Üç araştırma sorusu araştırıldı: eşit yaştaki cinsiyetler arasındaki Alzheimer ölümlerinin sayısındaki fark; Irklar ve ölüm yerleri arasındaki Alzheimer ölüm sayısı farkı; ve Alzheimer ölümlerinin sayısı ile yaş arasındaki korelasyon. İki fark $\chi 2$ testi ile test edildi ve korelasyon ρ testi ile değerlendirildi. Basit mevsimsel model, Ocak 2020'den Aralık 2025'e kadar gelecekteki Alzheimer ölümlerini tahmin etmek için kullanıldı.

Bulgular: Eşit yaş gruplarında, erkek ve kadın Alzheimer ölümleri arasındaki fark 70 yaşın altında ($\chi 2$ =0.580) ve 70 yaşın üzerinde ($\chi 2$ = 0.0001) anlamlı değildi.). Veri kümemizdeki farklı ölüm konumlarının her birinde Alzheimer ölüm sıklığında ırklar arası büyük bir fark vardı ($\chi 2$ = 0.0001). Ölüm yaşı ile yaşa göre kaba ölüm hızı arasında 0.01 düzeyinde mükemmel bir pozitif korelasyon vardı (ρ = 1). 2025'te ABD'de 139.732 Alzheimer ölümü olacağını tahmin ettik; bu, en düşük 129.505 ölümden en yükseği 149.960'a (129.505 < 139.732 ölüm 2025 > 149.960) doğru bir değişiklik.

Sonuç: Alzheimer, erkeklerden daha fazla kadını etkiler; Hispanik olmayan veya Latino'dan daha çok Hispanik veya Latino; ve Beyaz olmayanlardan daha fazla Beyaz. Hastalar başka yerlerde olduğundan daha fazla huzurevinde/uzun süreli bakımda ve evde ölmektedir; ve ölme veya hastalığa yakalanma riski yaşla birlikte artar. Alzheimer erken erişkinlik döneminde (35 yaş civarında) ortaya çıkabilir. Alzheimer hastalarının yaz, sonbahar ve kış aylarında ölme olasılığı daha yüksektir; ve ilkbaharda ölme olasılığı daha düşüktür.

Anahtar Kelimeler: Zaman Serileri, Alzheimer ölümleri, Tahmin, Amerika Birleşik Devletleri.

Student: VIDAL FRANCK YMELONG GHOKENG.

Mentor: ILKER ETIKAN

Department: Biostatistics

English Summary

Aim: The study's objective was to perform statistical testing on the 2015-2019 Alzheimer's death registry time series in the USA to extract sustainable information characterizing variation in the population; and predict the number of deaths for the years to come.

Materials and Methods: Data were obtained online from the Centers for Disease Control and Prevention (CDC) WONDER database. Three research questions were explored: the difference in the number of Alzheimer's deaths between genders of equal age; The difference in the number of Alzheimer's deaths between races and places of death; and the correlation between the number of Alzheimer's deaths and age. The two differences were tested with $\chi 2$ test and the correlation was evaluated with ρtest. The simple seasonal model was used to predict future Alzheimer's deaths from January 2020 to December 2025.

Findings: In equal age groups, the difference between male and female Alzheimer's deaths was not significant under 70 years ($\chi 2 = 0.580$) and significant over 70 years old ($\chi 2 = 0.000$). There was a large interracial difference in Alzheimer's mortality frequency at each of the different death locations in our dataset ($\chi 2 = 0.0001$). There was an excellent positive correlation at 0.01 level between age at death and crude death rate by age ($\rho = 1$). We estimated that there would be 139,732 Alzheimer's deaths in the US in 2025, a change from the lowest of 129,505 deaths to the highest of 149,960 (129.505 < 139.732 deaths in 2025 > 149.960).

Conclusion: Alzheimer affects more women than men; More Hispanic or Latino than not-Hispanic or Latino; and more Whites than non-Whites. Patients die more in nursing homes/long-term care and at home than elsewhere; and the risk of dying or contracting the disease increases with age. Alzheimer can occur in early adulthood (around age 35). Alzheimer's patients are more likely to die in summer, autumn and winter; and less likely to die in the spring.

Keywords: Time Series, Alzheimer's deaths, Forecast, United States.

1. Introduction and Aim

Each 3 seconds around the world someone develops inability to think critically, remember, or makes decisions that correlate with activities done every day (Link, Accession

date: 03 February 2021); a set of signs/symptoms that generalize Dementia. Several forms exist in dementia; Vascular dementia, Lewy bodies dementia, dementia-Parkinson complex, Frontotemporal dementia, Alzheimer's disease, etc.**(Garre-Olmo, 2018)**.Vascular dementia is caused by brain damaged by vascular injury **(Prathipati et al., 2021)**; Lewy bodies dementia is a form of dementia caused by an anormal deposition of alpha-synuclein protein (Lewy bodies) in the brain **(Phillips et al., 2021)**]; dementia-Parkinson complex is a form of dementia that develops in people living with Parkinson disease long time later after diagnosis (W. Yang et al., 2007);Frontotemporal dementia is caused by the loss of brain cells in the frontal and temporal cortices **(Häkkinen et al., 2020)**]; Alzheimer's disease is a neurodegenerative dementia characterized by progressive loss of cognitive function **(Liu et al., 2021).**Among all the forms, Alzheimer's disease (AD) is the most common**(Xie et al., 2021)** with approximately 50 million peoples affected worldwide and a wide manifestation in Western Europe and North America (link, Accession date: 05 April 2021) since 2010 **(Wimo et al., 2013)**. Between the 23 countries that constitute the North America, the United Stated are the most affected by AD. In 2018 the disease was classified as the sixth leading cause of deaths, with 122.019 official death certificates records **("2020 Alzheimer's Disease Facts and Figures," 2020)**,and the literature predicts a continuous increase of this value in upcoming years. The wide impact of the disease in the country has leaded to the implementation of many researches trying to explain its death rate with regards to the population's demographics characteristic, the country geographic characteristics, or the country disease management. With some resolutions taken by the country, and considering the possible variation that might have affected the disease manifestation, an update of previous research results sees al its importance. Therefore, our study aims to apply statistical test on time series 2015-2019 Alzheimer's death record in the USA to extract sustainable information characterizing the disease death variation in the population and forecast values for upcoming years; a useful way to update the literature for an evaluation and/or a reorganization of the disease management in the country.

2. General Information

2.1. Alzheimer's disease history.

Alzheimer's disease is a neurodegenerative disorder first identified by the German psychiatrist Dr. Alois Alzheimer in a fifty-year-old patient who he followed from 1901 till her dead in 1906**(Shampo et al., 2013)**.There was not a real interest in the disease until 11 similar

cases were identified in the medical literature the next five years**(Berchtold & Cotman, 1998)**. The brain degeneration leading to the disease evolves progressively, and is caused by abnormal formation of proteins in (neurofibrillary tangles of hyperphosphorylated tau protein) and around (aggregation of β-amyloid peptide into amyloid plaques) brain cells **(Tackenberg et al., 2020)**.

The disease progression is generally described in three stages on the basis of a diagnostic: the mild stage, the moderate stage, and the severe stage**(Lancet, 2011)**. At the early or mild stage, we observe small difficulties in basic human's functions (execution of movement, executive function, perception) than memory issues. The middle or moderate stage is characterized by the inability of the affected to perform common daily living activities (progressive loss of living and writing skills, vocabulary mistake etc.). Reaching the severe stage, the patient manifest neurological disturbance (loss of verbal language ability, inability to perform daily simple task …), and is completely caregivers' dependent**(Förstl & Kurz, 1999)**.

The diagnostic of AD is usually based on the patient's history from relative, behavioral observations, and medical history, and the idea of others dementia or cerebral pathology is generally excluded with advanced medical imaging**(Schroeter et al., 2009)**. Some biochemical tests involve the analysis of cerebrospinal fluid for total tau protein and phosphorylated $tau_{181P,}$ and beta-amyloid protein concentrations**(Sharma & Singh, 2016).**

Several trials have been implemented for Alzheimer's drugs development, from 2002 to 2012 about 244 compounds were assessed up to phase III trial, but only one (memantine) received approval from the United State Food and Drugs Administration (FDA)**(Cummings et al., 2014)**and is used in patient who are at the moderate to the severe stage of the disease and are showing difficulties with alertness and attention**(Cummings et al., 2014)**, **(Weller & Budson, 2018)**.Up to date, in addition to memantine, researches in pharmacologyalso propose cholinesteraseinhibitorsgalantamine, rivastigmine, and donepezil as therapy for thosewithsevere, moderate, or mildform of the disease**(Howard et al., 2012)**.Nevertheless, Alzheimer's deaths are day to day increasing worldwide, theanalysis of series of deaths records then sees its usefulness for a good disease management.

2.2. Time series

Time series (TS)or time series data, aresequence of data collected in equal time interval for a certain period. These data can be of various proprieties directing the classification of the TS. According to the number and to the characteristics of variable

involved, we distinguish univariate TS when a single variable is collected over time, from multivariate TS when two or more variables are involved in the collection; and discrete time series (when the set of observations is discrete), from continues time series (when the set is continuous)**(Daniels & Pourahmadi, 2009)**. Furthermore, When the collected TS data have a mean and variance that sparsely vary over time, we define the time series process as stationary; otherwise, when the TS data presents a trend (either increase or decrease over time), or a seasonality pattern (presence of regular repeating cycles over time) we qualify the time series process as non-stationary **(ROSCA, 2010)**.

In general, three main objectives direct the use of TS: better clean the data, understand the data, and forecast upcoming values**(Montgomery et al., 2008)**.The first two objectives generalize the time series analysis (TSA); the performance of statistical analysis on TS data to extract useful information **(Søren et alt., 2016)**. This technic started since the 18 century (Yule 1927) with main applicationin business and economy **(Tsay, 2001)**; and further followed the application in life science. The third objective characterize time series forecasting; the greatest profit of TSD.

2.3. Time series forecasting

Time series forecasting (TSF)is the prediction of future values based on historical data (TSD) using a model**(Jalil & Rao, 2019)**.In accordance to the mathematical process involved, TSF models can be grouped as deterministic or stochastic. Deterministic models assume a time series process not affected by any variations (randomness); as soon as a mathematical expression is built on the series, upcoming values are predicted based on this expression without considering anything that might affect the data. Conversely, stochastic models accept some randomness that might occurs, and therefore the series can't be explicitly expressed by an analytic expression, but rather by a probability distribution**(John & Mark, 2013)**. Probably in nature no phenomenon is entirely deterministic, most time series forecasting is based on stochastic modelsfrom which the most used are: exponential smoothing (ES) models, autoregressive models (AR), moving average (MA) models, and the popular Box-Jenkins model (association of autoregressive and moving average models).A time series model building is generally done in three phases: the model identification, the model estimation, and the model validation.

2.3.1. Model identification

The identification phase consists of determining the stationary status of the series and select the model suitable for the forecasting.

Stationarityis usually detected with a run sequence plot of the series, and when seasonality is doubted to explain the variation in mean and variance, a seasonal subseries plot, or multiple box-plotcan be used for confirmation**(John & Mark, 2013)**. In addition to these plots, various statistical tests are also usedfor stationarity checking like the most known Dickey-Fuller and Augmented Dickey-Fuller Test, which assume as null hypothesis that the series is not stationary, and the KPSS (Kwiatkowski-Phillips-Schmidt-Shin) Test which conversely to the other assume as null hypothesis that the series is stationary **(Shin c Schmidt, 1992)**; **(Aditya Satrio et al., 2021)**.

Stationarity is a main assumption in TSF; and most be obtained prior to the forecastfor series that are not stationary. A non-stationary series can be made stationary by:

- Simple differencing:

This method is used for non-seasonal series; the stationarity is obtained by simply subtracting consecutive observation in the series. The mathematical formula can be written as:

$$Xt' = Xt - X(t-1)$$

Xt'= differenced data

Xt = data at rang t

- Seasonal differencing:

This is used when there is seasonality effect in the series; here, unlike simple differentiation, the difference is done between a term and its precedents of the same season. For example, a term recorded in January will be subtracted from a term recorded in the previous January. The mathematical expression can be written as:

$$Xt' = Xt - X(t-n)$$

Xt'= differenced data

Xt = data at rang t

n = seasonal term

- Transformations (square root, power transform, and log transform etc.):

This is used when after differencing, the series' variance is still non constant. For example, the log transformation replaces every value of a series by its logarithm; mathematically, we express it as:

$$\mathbf{Xt' = log\ (Xt)}$$

(Anderson, 1976; Luetkepohl & Xu, 2009)

Once stationarity is obtained on the series, the suitable model for the forecast has to be identify. This is done by plotting and analyzing the autocorrelation function (ACF) and the partial autocorrelation function (PACF) of the series. The ACF is defined as the correlation between a variable and a copy of itself shifted by a certain number of units (time lags); mathematically, it is the quotient of the autocovariance at some lag on the variance of the series**(Gubner, 2006; Mrna & Hornik, 2016)**, and expressed as:

$$A_k = \frac{\sum_{t=k+1}^{n}(x_t - \overline{x})(x_{t-k} - \overline{x})}{\sum_{t=1}^{n}(x_t - \overline{x})^2}$$

x_t= dataset ranged by ascending date

x_{t-k}= dataset shifted by k units

$\overline{x}$ = the mean of the original dataset.

The PACF is an ACF that takes into account the values at previous lag. Mathematically, it is the correlation at lag k with a regression of the values of the series at all lags k-1**(Daniels & Pourahmadi, 2009)**, and expressed as:

$$\mathbf{P_k} = \frac{c\left(\left[x_t \middle| x_{(t-1)}, x_{(t-2)} \ldots x_{(t-k+1)}\right], \left[x_{(t-k)} \middle| x_{(t-1)}, x_{(t-2)} \ldots x_{(t-k+1)}\right]\right)}{\sqrt{\left[x_t \middle| x_{(t-1)}, x_{(t-2)} \ldots x_{(t-k+1)}\right]\left[x_{(t-k)} \middle| x_{(t-1)}, x_{(t-2)} \ldots x_{(t-k+1)}\right]}}$$

c **=** covariance

x = dataset

x_i = data at the rang i in the dataset

k = time lag

It exists in the literature several ways to identify suitable TS models through the ACF and PACF.**(Jain & Mallick, 2017)**proposed the table 1 below for model's identification, and highlighted the use of ES when dealing with a univariate series with a trend or seasonal component. Box Jenkins model can also be used for data with trend and seasonal component as the ES model, and here the order **P** and **Q** for the seasonal AR and MA parameters has to be multiplied to **p** and **q(Valipour, 2015).**

Table 1: Time series model identification

The figure presents the method to identify AR, MA or ARMA models through analysis of the shape of the ACF and PACF. The columns represent the different models, and the lines represents the characteristics of the plots directing the choose of the models.

	MA (q)	**AR (p)**	**ARMA (p, q)**
ACF	Tail off	Cut off q lag	Tail off
PACF	Cut off, p lag	Tail off	Tail off

2.3.2. Model estimation

The estimation phase is only applied in Box and Jenkins model, and consist of determining the order **p** and **q** for the respective AR and MA parameters. The order **p** is determined on the PACF and **q** on the ACF; and these orders commonly represent the lag number beyond which there is a cutoff in the respective plot (i.e. The lag out of the confidence interval)**(ArunKumar et al., 2021; Khan & Gupta, 2020)**. In general, there is not a unique accepted method to determine the optimal p and q, practitioners sometimes try different values of p and q, and pick the suitable combination for their dataset.

For example, let us use the ACF and PACF in the figure 1 below to determine the orders for the AR and MA process, and identify the suitable TS model. The blue color represents the CI.

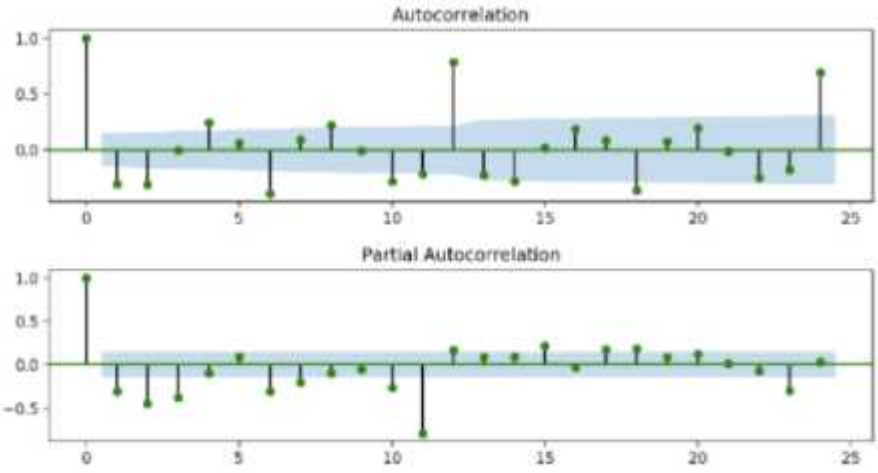


Figure 1: Example of ACF and PACF plot.

According to our table 1 above, the suitable model for the series is an ARMA. In the ACF the cut of is after lag 2, indicating the use of a MA term 2 (q = 2); and in the PACF the cut off is after lag 3, directing the use of an AR term 3 (p = 3).

Another common way to identify the AR and MA terms is by counting the number of bars out of the CI (either above or below, and without including the bar at lag 0) before the next bar that is included in the CI.

2.3.3. Model validation

The last step in TSF is the model validation (or diagnostic in Box-Jenkins), it consists of verifying if the chosen model well fit the data.In Box Jenkins method it helps to choose the suitable model when several are tested. The different appropriate way to diagnose a model are by analyzing the correlogram of the residuals, by analyzingsome parameters (the significance of the model's component, the volatility, the log likelihood ratio, the Bayesian inclusion criteria (BIC), the adjusted R^2, and the Akaike information criterion (AIC)), or by running one among the various diagnosis test**(Chawsheen & Broom, 2017)**.

The residuals correlogram checking allows to detect whether some information are not captured by the model i.e., if there are some lags out of the CI. For instance, the most ideal is a flat correlogram because it signify that all the lag are included in the 95% CI**(Yu et al., 2015)**. Nevertheless, it can arrive that some lags are still significant in one or both of the correlogram (ACF and PACF). When this is the case,while some authors suggest to re-estimate the model and include the lag(s) that is(are) still significant to capture all the information; others suggest to not re-estimate the model, take parsimony as the watchword

(the principle of parsimony in TS sustains a model selection with the possibility to minimize the number of parameters that fully express the linear dependence structure, offering greater prediction and new observations' generalization), ignore this(ese) lag(s), and use the chosen modelfor the forecast to avoid overfitting**(Ledolter & Abraham, 1981; Moffat & Akpan, 2019)**.

In addition to the analysis of the residuals correlogram, several diagnostic test can be used for model diagnostic. In practice these tests are regrouped in three classes: the tests of serial uncorrelatedness, the tests of martingale difference, and the tests of serial independence**(Kuan, 2008)**.

- The tests of serial uncorrelatedness test if the residuals from different period are correlated.Wedistinguish the popular Q tests of Ljung and Box (1978) andBOX and Pierce (1970), the robust spectral test of Deo (2000), etc. **(Nankervis & Savin, 2010)**.
- The tests of martingale difference verify if the predicted value is a martingale (a stochastic process/sequence of random variables for which, at a given time, the expected next value in the sequence is equal to the currentobserved value, without consideringall previous values**(Victor, n.d.)**). Statistically, it test the hypothesis that the mean of a time series is independent of its pass **(Domínguez & Lobato, 2003; Salisu et al., 2016)**. We can cite as example the Kolmogorov-Smirnov test, and the Cramer- von Mises.
- The tests of serial independence access if the sequence of the residuals data is identically distributed and independent **(Ghoudi & Rémillard, 2018)**. The null hypothesis tested here is that the residuals series is serially independent.This can be performed with the Cramer-Von Mises test, the Kolmogorov-Smirnov test, the Durbin and Watson test …,according to the user's expectations and conveniences**(Diks & Panchenko, 2007; Ghoudi et al., 2001).**

Another important tool for model diagnostic in TSF is the analysis of parameters: the significance of the model's component, the volatility, the log likelihood ratio, the BIC, the adj R^2, and the AIC. These parameters are most important when several models have been estimated (almost the case in Box-Jenkin methodology); they help to choose the one that most suit for the forecast.

- Among different model, the suitable is the one with the highest significant components (p, q).
- The volatility is the standard deviation of the series (residuals), and the suitable model among several is the one with the lowest volatility.

- The log likelihood ratio assesses the goodness of fit of the estimated models, based on the number of times the observed log values are more likely to occur under one than the other**(Li & Babu, 2019)**. Between several models tested, the best is the one with the highest log likelihood ratio **(Dickey & Fuller, 1981)**.
- During model fitting, when adding parameter to capture all information, the likelihood has tendence to increase, and consequently we might result in overfitting. The BIC and AIC resolve the problem of choosing the good model in this case by penalizing the number of parameters in the model. The difference between the two is that the penalty term is slighter in AIC than in BIC. The suitable model between several is the one with the lowest AIC and BIC. **(Lian, 2014; Portet, 2020)**.
- The R^2 indicates the proportion of the variance in the dependent variable that is explained by the independent variable. Highest is the R^2, best is the prediction. Commonlyin TSF users increase the number of predictors to capture all the information. Since the R^2 tends to increases with the number of predictors, this can lead to an unwarranted high R-squared value, and consequently to an overfitted model. Therefore, the adj R^2 resolves this problem by penalizing variables that do not fit the model. We end up with a higher or lower adj R^2, depending on a small or high number of unsignificant predictors. The adj R^2 is positive and always lower than the R^2, and the best model among different implemented is the one with the highest adj R^2**.(Karch, 2020)**.

Once the model is validated, it can now be used to forecast future values for a determined period.

TS is applied in many fields of science and engineering. In medical research, the method helps to understand disease manifestation in a population and helps for disease management, and numerous of its implementation have been done on AD in the United State.

2.4. Literature review

As said earlier, the United State is among the country most affected by AD in the world, and a broad of TSA studies have been conducted on AD or on deaths by AD to verify the different hypothesis behind the disease management in the country. The idea of an existing relationship between AD death rate and Per Capita Personal Income was investigated by **(Stępkowski et al., 2015),** who analyzed the correlation between historical per capita personal income (PCPI) from 1929-2011 for all states of the USA and corresponding age-adjusted AD decease rates (AADR) for years 2000, 2005 and 2008. They observed negative correlation in all cases, the strongest ($R \approx -0.65$) between the PCPI in 1970 and the AADRs in 2005.

Also, an evaluation of the disease management in function of different places where AD patient sojourn was investigated by **(Xu et al., 2020);** they performed a $\chi 2$test to assess the significant changes in the proportion of dementia death at difference places (decedent's home, hospital, nursing home/long term care, and other places) across state and over time, and studied the association between the proportion of dementia decedents in various places and determinants of dementia care. The data included the nationwide death certificates between 2000 and 2014. The results show a large inter-state and temporal difference in the proportion of deaths at each of the four places of death ($\chi 2p < 0,01$), there was a significant association between the access to care facility resource in hospitals and nursing home/long term care and dementia decease at respective places (all $p < 0,001$); and between Medicare and Medicaid investments in various health agencies/services and proportion of dementia death at various places (nursing home/long term care deaths, hospital deaths and home deaths) (all $p < 0.005$).

In addition, AD progression is far to be neglected, and a way to predict this is a great addition to the disease management.Following this ideology, **(Fisher et al., 2019)** Applied a Conditional Restricted Boltzmann Machine to forecast Alzheimer's disease progression with a sample of 1909 patients with Mild Cognitive Impairment Alzheimer's Disease; the training data comprised 18-month trajectories of 44 clinical variables; the model performance was evaluated using 5-fold cross validation which results over were either aggregated or averaged.

Furthermore, noticing the significant increase in disease's deaths record in the country, a projection of what the it might face in the future is far unnecessary. In line with this, **(Hebert et al., 2013)**Performed a weighted logistic regression to estimate AD dementia incidence in United State with a sample of 1913 individuals; the study estimated that out of 4.7 million individuals with AD dementia in 2010, 0,7 million were 65 to 74 years, 2,3 were 75 to 84, and

1,8 were 85+ with all 95% confidence interval not including 0 ; Also, the study projected the total number of peoples with AD dementia to be 13,8 million in 2050, with 7,0 million aged 85+ years.Similarly, **(Taylor et al., 2017)** after analysis of state-level and county-level death certificate data from the National Vital Statistics System from 1999 to 2014, highlighted that age-adjusted rates of Alzheimer's death-rate have significantly increased in 41 states, with highest counties in the Southeast, the Midwest and the West.

Many ideas online describe AD death in the United State taking into account variables such as Age, Gender, Hispanic origin, Races, and States (Link, Accession date: 14/04/2021). However, few are those that are based on solid statistical tests, and for those who are, an update sees all its importance to highlight any potential changes since the past results. Therefore, our study finds its place in this issue solving.

3. Material and Method

3.1. Data source description

WONDER online database of the Centers for Disease Control and Prevention **(CDC)** contains open access data about natality, vaccination, tuberculosis, cancer incidence, HIV and AIDS, mortality, and many others topic related to diseases of United State residents. Data for this work were obtained after filling the Underlying Cause of Death data, 1999-2019 request of CDC WONDER.The data involve national mortality and population data at county level covering the years 1999-2019, and are acquired after filling a request constituted of 7 sections.

- The first section is the table layout organization:Here we have to specify the criteria by which we want to group our request's results.We give a title to our request, we choose the additional rate and the age adjusted rate option, and we specify: the year of data collection, the gender, the Hispanic origin, and the race that interest us.
- The second section is the location selection:We choose the location option by which we want our dataset to have been grouped; either by State, by Census Regions, or by HHS (Health and Human Service) Regions.Also, we choose the year of NCHS Urban-Rural Scheme for Counties that we prefer.
- The third section is the demographics criteria's selection:In this section we specify the Age-group interval, the Gender, the Race, and the Hispanic origin by which we want to group our data.
- The fourth section is the year and month selection:Here we select the year and/or month of data's collection that interest us.
- The fifth section is the place of death's, autopsy's, and weekday's selection: here we specify if in our dataset we are interested in knowing the weekday the death has occurred, the autopsy status, and the place of date.
- The sixth section is where we select the cause of death that interest us between severalcoded according to the International Classification of Disease Tenth Revision (ICD-10), and grouped in five categories: ICD-10 Codes (which groups all cause of death), ICD-10 113 Cause List (which groups 113 selected cause of death), ICD-10 130 Cause List (Infants) (which groups 130 Selected Causes of Infant Death), Drug/Alcohol Induced Causes (which groups death caused by Drug/Alcohol), and Injury Intent and Mechanism.

- The seventh section is for additional option in our request; here we precise if we want to export our result; to show: total, zero values, or suppressed values in our data. We choose the number of decimals in our values; and set the timeout for data access.

3.2. Data acquisition

We filled 2 successive requests to acquire our data. For the first request, in the first section we grouped the data by “5-Year age group”, “Gender”, “Hispanic origin”, “Race”, by States, and “Place of death”. In the second section we choose our data location by state, and selected “all the United States” and the “2013 Urbanization”. In the third section, we selected the “Five-Year age group”, “All Genders”, “All Hispanic Origin”, and “All Races”. In the fourth section we set up 2019 as our interested year.In the fifth section we selected “All Weekdays”, “All Autopsy values”, and “All Place of Death”.In the sixth section we clicked on the “ICD-10 113 Cause List” and selected “# Alzheimer disease (G30).In the seventh section we ticked the “show total” and the “show zero values”, we precise 3 for our “decimal places”, and set the “Data Access Timeout” to 15 minutes. The second request was similar to the first, except that the data were grouped only by “Month” and the interested years was the interval “2015 -2019”.

3.3. Dataset description

The dataset was constituted of 15 columns representing the different variables: States, State Code, Five Year Age Groups, Five Year Age Groups Code, Gender, Gender Code, Hispanic Origin, Hispanic Origin Code, Race, Race Code, Place of Death, Place of Death Code, Deaths, Population, Crude Rate.

- ‘State’ is constituting of 51 categories representing different States in the US.

- ‘Five-year age groups’ represents the classification of deaths by age within a 5-year interval.Our dataset distinguishes 15 categories from the age of 35 to 100+.

- ‘Gender’ categorizes deaths by sex and differentiate 2 groups: Female and Male.

- ‘Hispanic Origin’ categorizes deaths according to the population’s origin. Hispanic/Latino, are distinguished from Not Hispanic/Latino.

- ‘Race’ differentiates the deaths according to the different population Race in the US. 4 categories are to distinguish in our dataset: American Indian or Alaskan Native, Asian or Pacific Islander, Black or African American, and White.

- ‘Place of Death’ groups deaths according to theplace they have occurred.The different categories in the data are: Medical Facility - Inpatient, Medical Facility - Outpatient or ER, Medical Facility - Dead on Arrival, Decedent’s home, Hospice Facility, Nursing home/long term care, an Other.

- ‘Deaths’ comprises the number of deaths.

- ‘Population’ is US population estimate in the different States.

- ‘Crude Rate’ expresses the number of deaths reported per 100,000 population of a selected factor (State and 5-Years-Age-Group for our case).

- ‘State Code’, ‘Five Year Age Groups Code’; ‘Gender Code’; ‘Hispanic Origin Code’; ‘Race Code’; and ‘Place of Death Code’ represent the code corresponding to each category in the given variable name.

3.4. Data analysis

Our data were analyzed using IBM SPSS Statistics version 28. The Time Series Analysis (TSA) was done using the 2019 deaths record only; and the Forecasting was done using 2015-2019 records.

For the TSA, the descriptive statistic frequency was used to access the number of deaths in each category of our different variables; histogram and pie chart were plotted to illustrate the difference between categories; and line plot was used to describe Alzheimer’s deaths tendency according to specific criteria.

Three research questions were to investigate: is there a difference in number of Alzheimer’s deaths between gender at equal age group? is there a difference in the number of Alzheimer’sdeaths between Races and places of deaths? is there a correlation between age and the number of Alzheimer’s deaths? The first and the second questions were tested with $\chi 2$, and forward method was applied on the different 5-years-age groups when testing the first question. The third question was evaluated with ρtest.

For the Time Series Forecasting, we visualized the series with a run sequence plot; the stationarity was tested with the Augmented Dickey-Fuller Test, and the ACF and PACF plots of the original series was used to identify trend or seasonality in the series. Seasonal difference was applied to make the series stationary, ACF and PACF plots of the differenced serieswere used to identify the model with which we will fit our series. Three different models were tested: SARIMA (1,0,1) (1,1,0), SARIMA (0,0,2) (2,1,0), SARIMA (0,0,1) (2,1,0) all with constant. The models were compared using the Stationary R-squared, the Normalized BIC, and the significance of the components.The best model among the three was used for our forecasting, and the forecast was done from January 2020 to December 2025.

4. Finding.

4.1. Time series analysis

Alzheimer's deaths by gender

Our data recorded a total of 121.499 Alzheimer's deaths in 2019. The frequency distribution between genders shows a higher number of deaths among females than males. 83.516 in the first group compared to 37.983 in the second group; and corresponding respectively to 69% and 31% of the total number of deaths. By differentiating, the number of deaths in women represents an increase of 119.87% compared to the number in men.

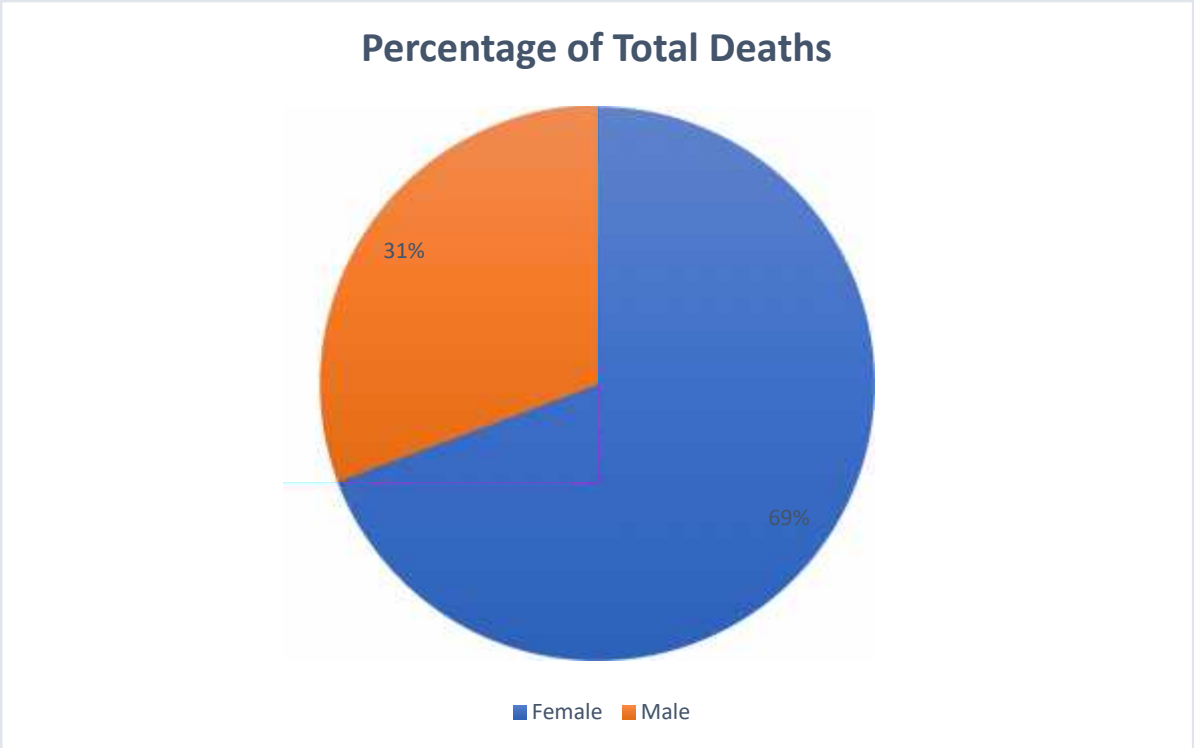


Figure 2: Alzheimer's deaths by gender.

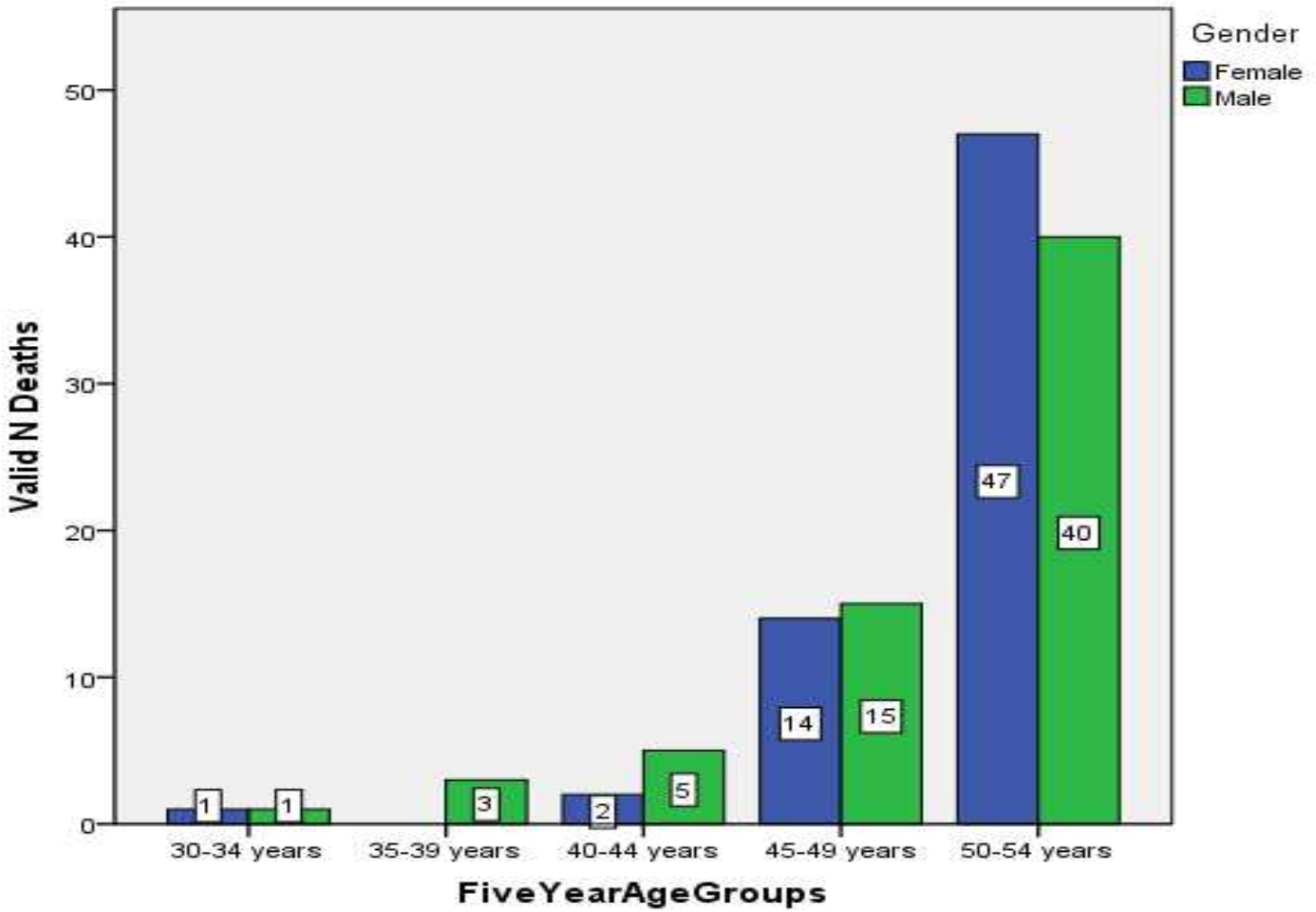
The figure is a Pie chartpresenting the number of deaths and the percentage of deaths occurrence by gender. The blue color characterizes the female percentage of deaths from the total, and the orange color the male percentage.

For the same age group, from 30 to 59 years, the number of deaths between gender varied lightly; We counted 1 death in male and 1 In female for age group 30-34 years; 3 deaths in male compared to 0 in female for age group 35-39 years; 5 deaths in male compared to 2 in female for age group 40-44 years; 15 deaths in male compared to 14 in female for age

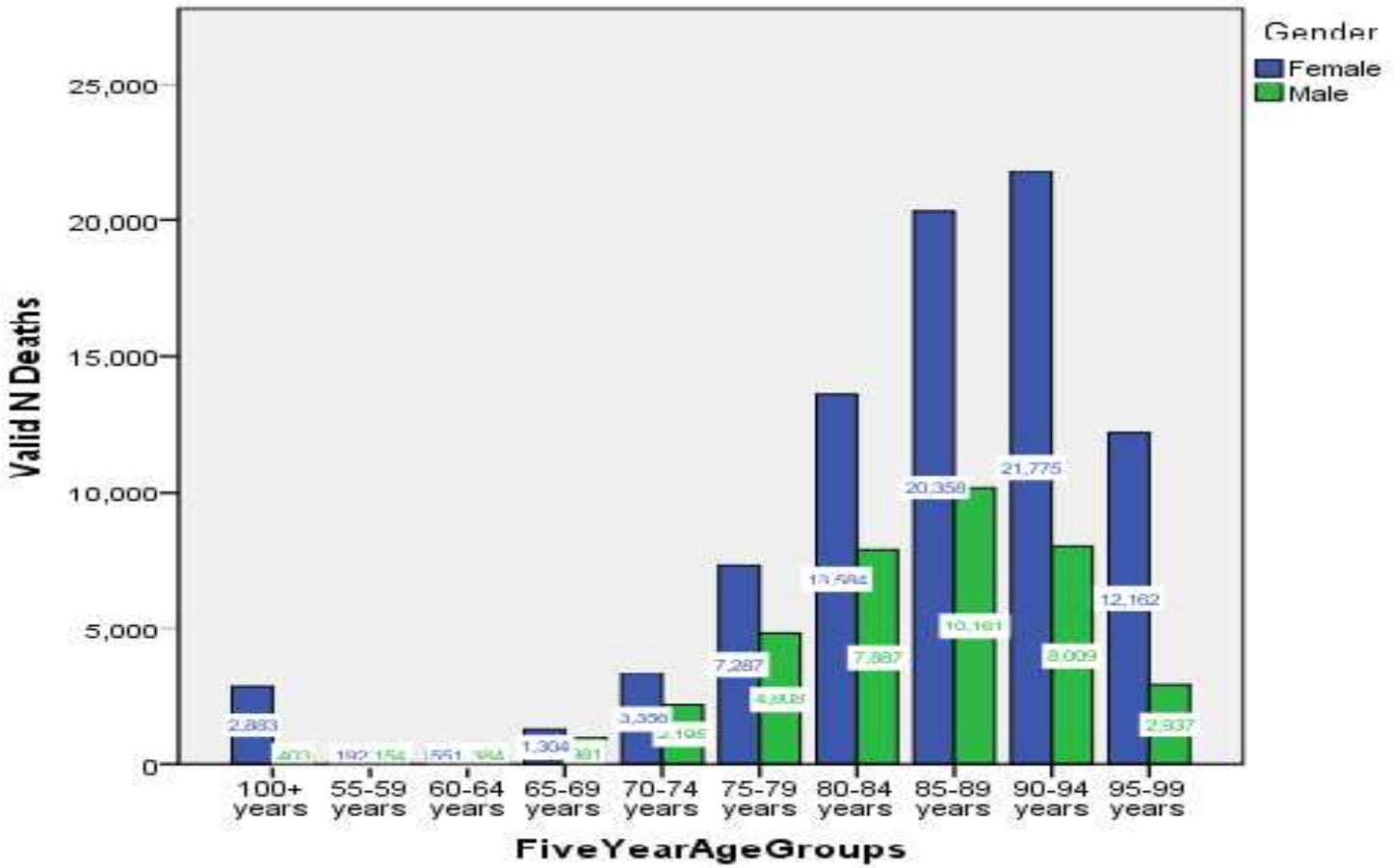
group 45-49 years; 40 deaths in male compared to 47 in female for age group 50-54 years; 154 deaths in male compared to 192 in female for age group 55-59 years. From 60 to 100+ years, the difference in deaths occurrence between men and women became high; 384 deaths in males compared to 551 in female for age group 60-64; 981 deaths in male compared to 1.304 in female for age group 65-69 years; 2.195 deaths in male compared to 3.356 in female for age group 70-74 years; 4.808 deaths in male compared to 7287 in female for age group 75-79 years; 7.887 deaths in male compared to 13.584 in female for age group 80-84 years; 10.161 deaths in male compared to 20.358 in female for age group 85-89 years; 8.009 deaths in male compared to 21.775 in female for age group 90-94 years; 2.937 deaths in male compared to 12.162 in female for age group 95-99 years; 403 deaths in male compared to 2.883 in female for age group 100+ (**Figure 3**).

Furthermore, the variation in number of deaths in Female from the number in male at equal age group was calculated and illustrated in **Figure 4**. From 30 to 49 years, there was either a decrease or a non-variation in the number of deaths in female compared to the number in male. The results show 0% among 30–35-year-old records; -100% among 35–39-year-old records; -60% among 40–44-year-old records; -6,66% among 45–49-year-old records. From 50 to 69 years, number of deaths in female increased slightly from the number in male. We noted 17,5% increase among 50–54-year-old records. 24,67% increase among 55–59-year-old records; 43,48% increase among 60–64-year-old records; 32,92% increase among 65–69-year-old records. From 70 to 100+ years, this increase becomes higher; respectively 52,89% among 70–74-year-old records; 51,55% among 75–79-year-old records; 72,23% among 80–84-year-old records; 100,37% among 85–89-year-old records; 171,88% among 90–94-year-old records; 314,09% among 95–99-year-old records; and 615,38% among 100+ year-old records.

In addition, Pearson $\chi 2$ results of difference in Alzheimer's deaths between genders at equal age group showed a non-significance difference ($p = 0,580$) at equal five years interval from 45 to 69 years old; and a significance difference ($p = 0,0001$) at equal five years interval from 70 to 100+ years old.



3-A) Alzheimer's deaths by gender at equal age group (age 30 to 54)



3-B)Alzheimer's deaths by gender at equal age group (age 55 to 100+)

Figure 3: Sum of Alzheimer's deaths by gender at equal age group.

The figure presents the total count of deaths between genders at equal age group; From age 30 to 54 (A); and from age 55 to 100+ (B). The blue color represents the female group and the green color represents the male group. Age is grouped in 5 years old interval, and the graph distinguish 15 groups from age 30 to 100+. The number in colors (blue and green) are the valid number/total count of deaths in the respective gender (female and male).

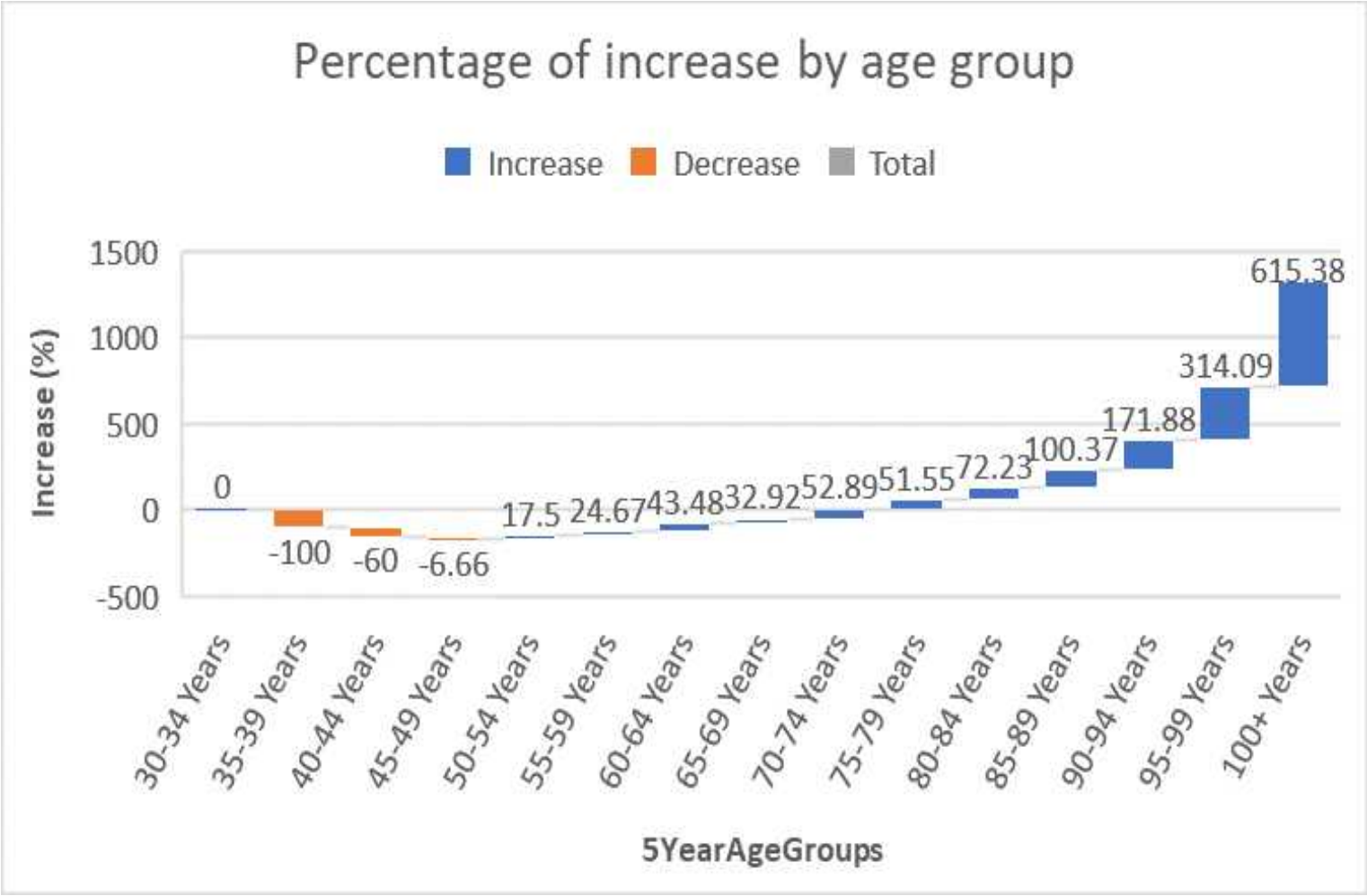
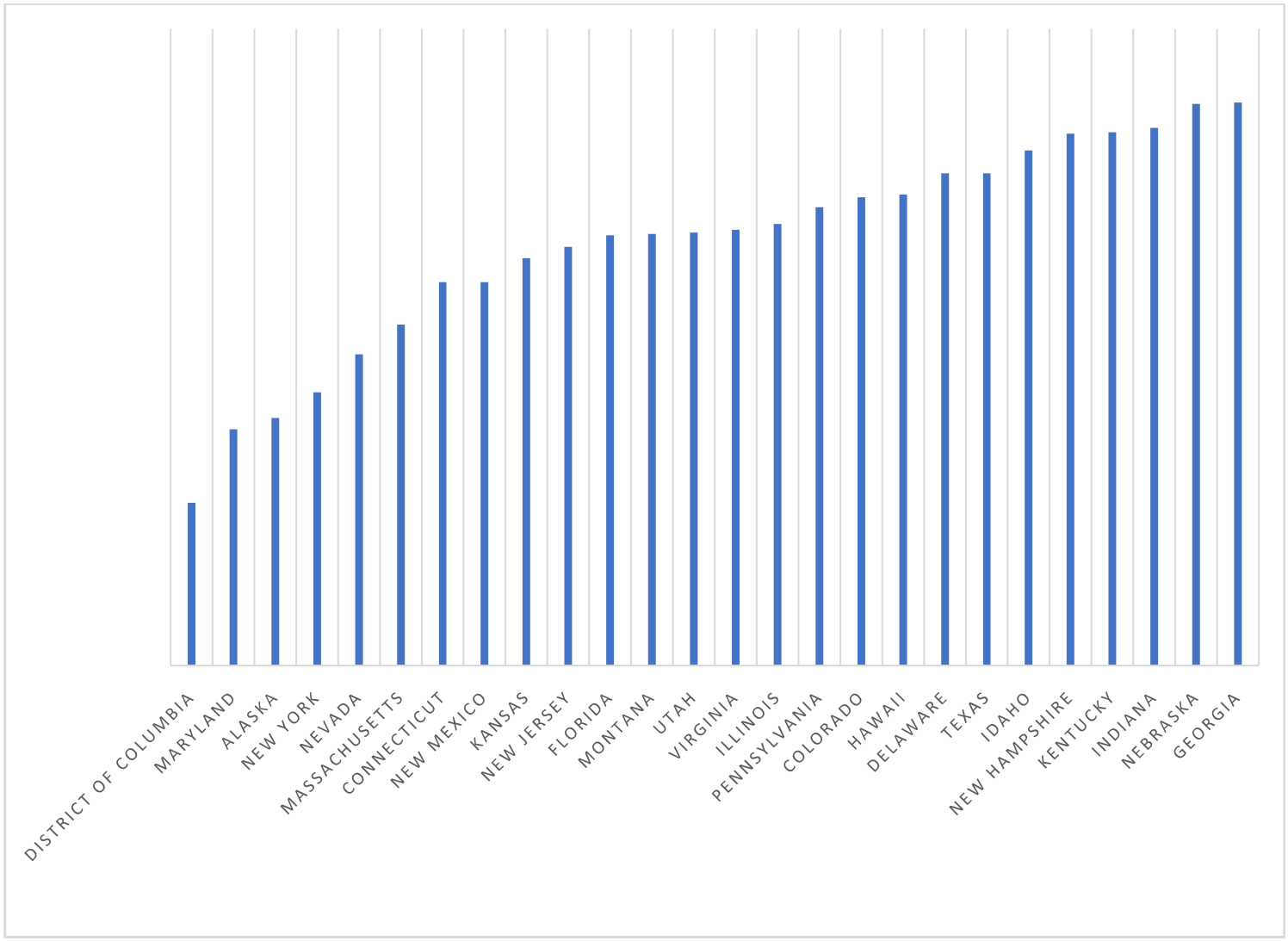


Figure 4: Percentage of difference in number of deaths between gender at equal age group.

The figure illustrates the percentage of difference in number of deaths in female from the number in male at equal age group. The blue color represents the increase; and the orange color represent the decrease. On the X axis we have the age grouped in 5 years intervals from 30-100+ years; and on the Y axis we have the difference in percentage.

- **Alzheimer's deaths by States**

The plot of the histogram of Alzheimer's deaths Crude Rate per 100,000 populationsof each States **(Figure 5)** showed a variation from 11.5 to 56. The lowest value was seen in the District of Columbia (11.5) and the highest in South Dakota (56). There was similar crude rate between Connecticut and New Mexico (27,1); Delaware and Texas (34,8); Minnesota and Missouri (45,3); and between North Carolina and Rhode Island (43). Florida, Montana, Utah, and Virginia had almost the same crude rate (between 30,4 and 30,8); same as Colorado and Hawaii (33,1 and 33,3); New Hampshire and Kentucky (37,6 and 37,7); Nebraska and Georgia (39,7 and 39,8); Wisconsin, Wyoming and Arizona (between 41 to 41,9); Iowa and California (42,6 and 42,7); Michigan, Ohio and Oklahoma (between 44,7 and 44,9); West Virginia and Louisiana (46,4 and 46,6); Washington, Oregon and Tennessee (from 47,1 to 47,6).


DISTRICT OF COLUMBIA
MARYLAND
ALASKA
NEW YORK
NEVADA
MASSACHUSETTS
CONNECTICUT
NEW MEXICO
KANSAS
NEW JERSEY
FLORIDA
MONTANA
UTAH
VIRGINIA
ILLINOIS
PENNSYLVANIA
COLORADO
HAWAII
DELAWARE
TEXAS
IDAHO
NEW HAMPSHIRE
KENTUCKY
INDIANA
NEBRASKA
GEORGIA

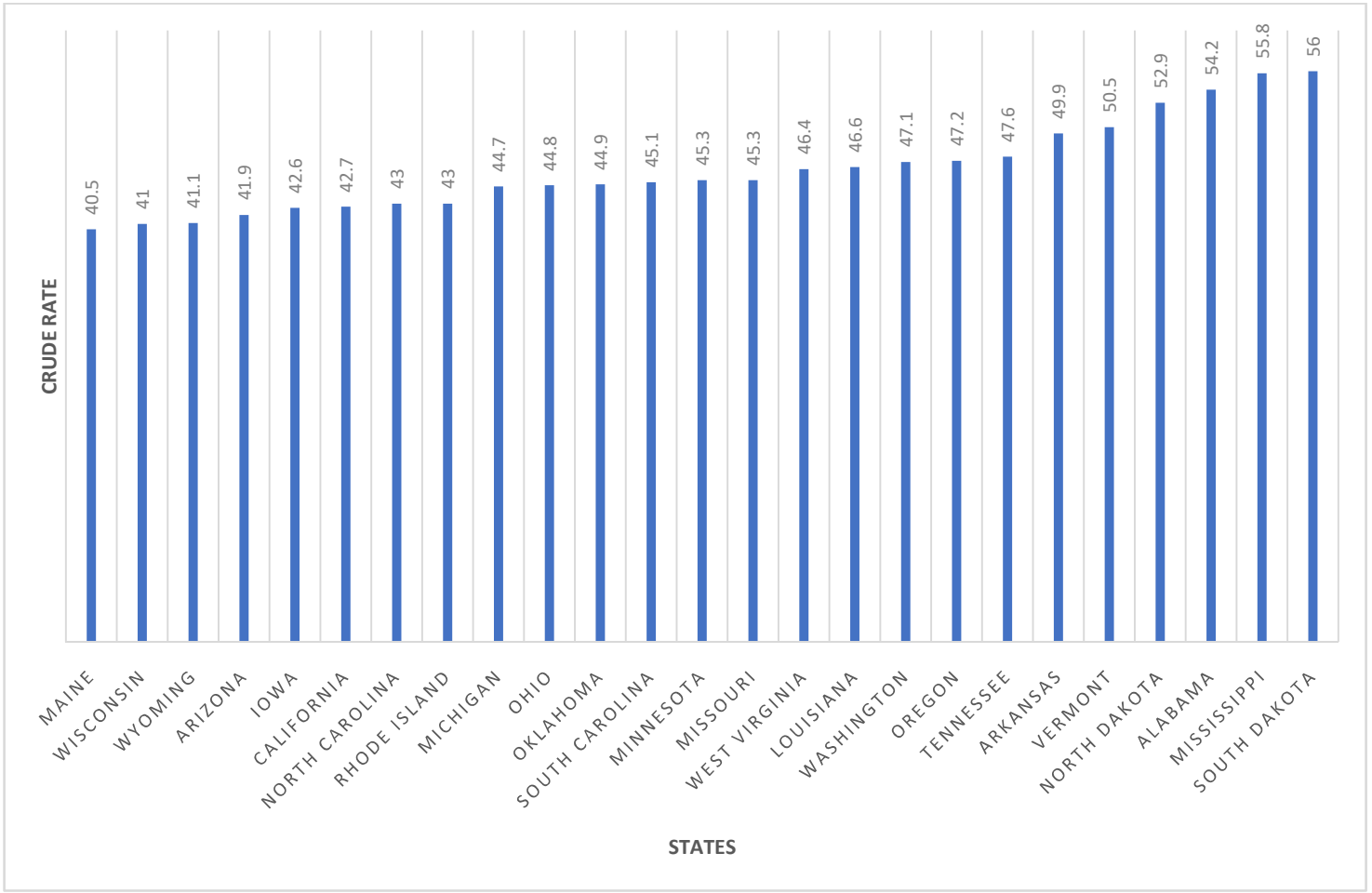


Figure 5: Alzheimer's Deaths Crude Rate by Sates.

The figure presents the crude rate of Alzheimer's deaths in the different States of the US. The crude rate is calculated as the number of deaths per 100,000 persons in each state

Alzheimer's deaths by race

Races' comparison showed a difference in deaths occurrence in the 4 groups. The highest number of deaths was in the White race with 10.8683 deaths corresponding to 89% of the total, and the lowest number in the American Indian or Alaska Native race with 381 deaths corresponding to almost 0% of the total. Between this two, Black or African American and Asian or Pacific Islander races counted 9.369 deaths (corresponding to 8% of the total) and 3.066 deaths (corresponding to 3% of the total) respectively (**Figure 6**).

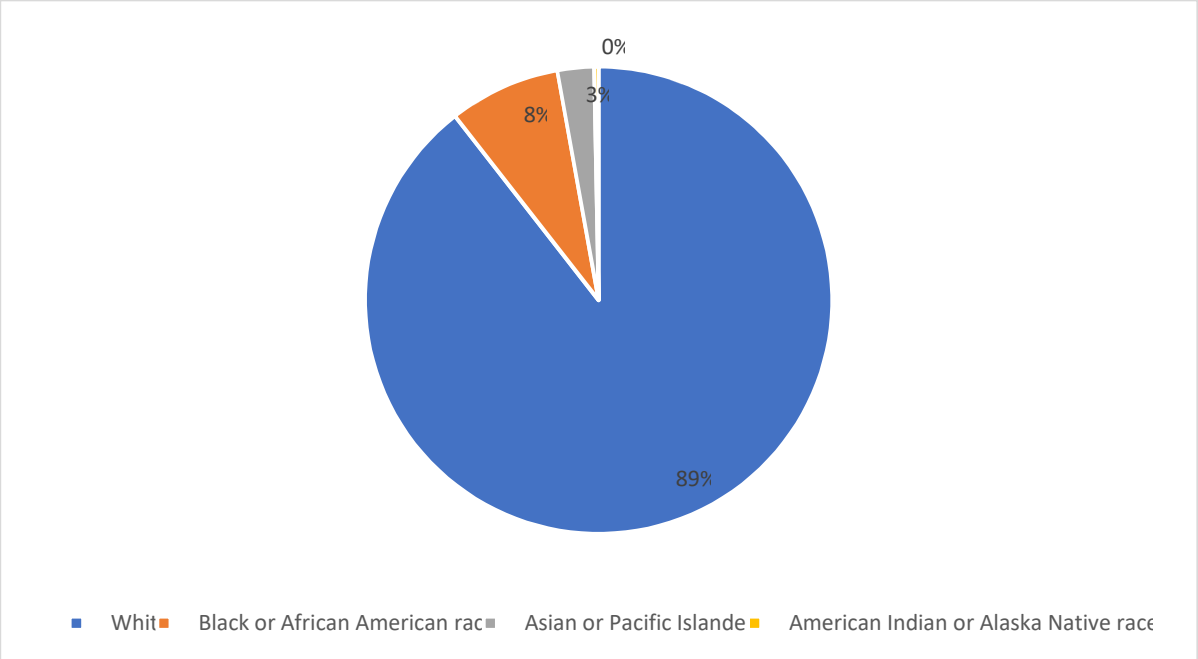


Figure 6: Number and Percentage of deaths by Races.

The figure illustrates the percentage of deathsfrom the totalfor each race. The 4 races are differentiated by color, the blue color corresponding to the White, the orange color representing the Black or African American, the grey color corresponding to the Asian or Pacific Islander, and the yellow color representing the American Indian or Alaska Native.

Alzheimer's deaths by Hispanic Origin

When compared deaths according to the Hispanic origin, the number of deaths in the Not Hispanic or Latino group (113.146 deaths) was 13,76 times the number in the Hispanic or Latino group (8.221 deaths). This was representing exactly 99,46% and 0,54% of the total deaths in the country as the **figure 7** illustrates.

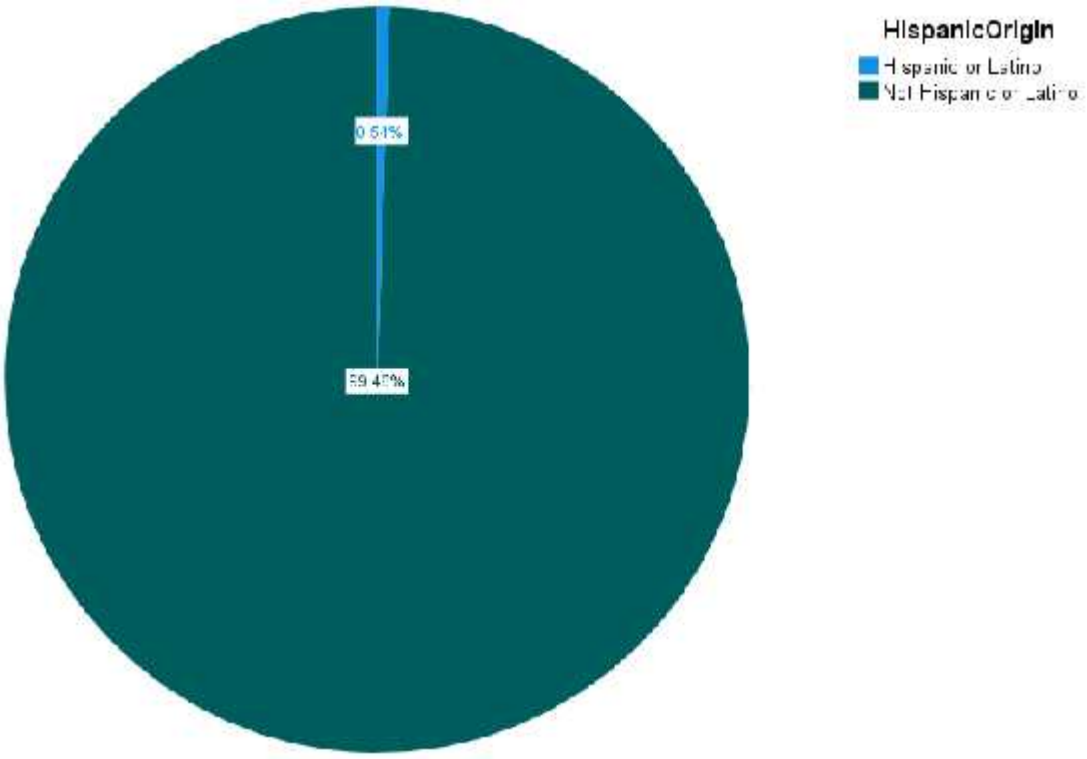


Figure 7: Deaths distribution between Hispanic Origin.

The figure presents the number of deaths (A) and the percentage from total deaths (B) between Hispanic origin groups. In (A) the X-axis represents the Hispanic origin (either Hispanic or Latino or Not Hispanic or Latino); with omission of deaths for which Hispanic origin was not stated in the data; and the Y-axis represents the valid number of deaths. In (B) the blue color represents the Hispanic or Latino group, and the dark green color represents the Not Hispanic or Latino group. The values in the pie chart are the corresponding percentage of deaths in each group relative to all deaths; with the color of each value matching with the color of the Hispanic group.

- **Alzheimer's deaths by places of deaths.**

For deaths occurrence by places, we highlighted that there were 60.844 deaths in Nursing Home/Long term Care, 34.503 deaths in decedents home, 12.001 deaths in other places, 7.928 deaths in Hospital facility, 5283 deaths in Medical Facility-inpatient, 873 deaths in Medical Facility-Outpatient or ER, 51 deaths in Medical Facility-Deaths on Arrival, and 16 deaths with unknown place of deaths. These represented respectively 50%, 28%, 10%, 7%, 4%, 1%, and 0% of the total deaths **(Figure 8)**.

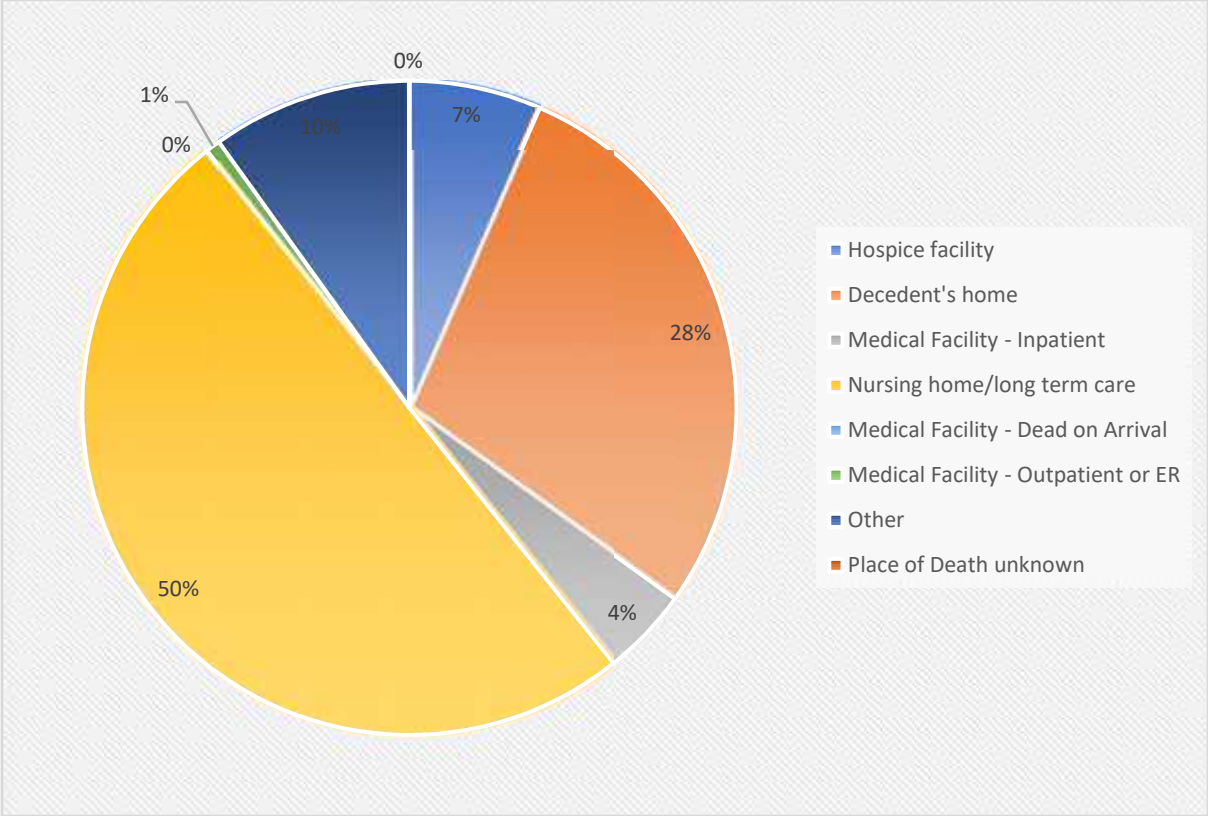


Figure 8: Alzheimer's deaths by place of death.

The figure classifies deaths by the places it has occurred, and give the percentage of each place from the total number of deaths. The yellow color represents deaths that has occurred in Nursing home/Long Term Care; the orange color those in Decedent's Home; the dark blue color those in Hospice facility; the light blue color those in Medical Facility-Dead on Arrival; they gray color those in Medical Facility-Inpatient; they green color those in Medical Facility-Outpatient or ER.

Furthermore, there was a large inter-races difference in the frequency of Alzheimer's death in each of the 8 places of deaths involved in our dataset ($\chi 2$ p = 0,0001). In every place of deaths, the White race had the highest frequency of deaths, in front of the Blacks or African American race, the Asian or Pacific Islander race, and the American Indian or Alaska Native race; in descending order.

Alzheimer's deaths by age

In our dataset Alzheimer's deaths crude rate by population age group were unreliable for age group 30-34 years, 35-39 years, 40-44 years; and not applicable for age group 85-89 years, 90-94 years, 95-99 years, and 100+ years. ρ test showed a perfect positive correlation (correlation coefficient (r) = 1) at the 0,01 level between deaths crude rate and age. The plot of the deaths crudes rates against the five years age group interval aligned with the correlation's results with a curve showing an increasing shape over the entire age group interval (**Figure 9**).

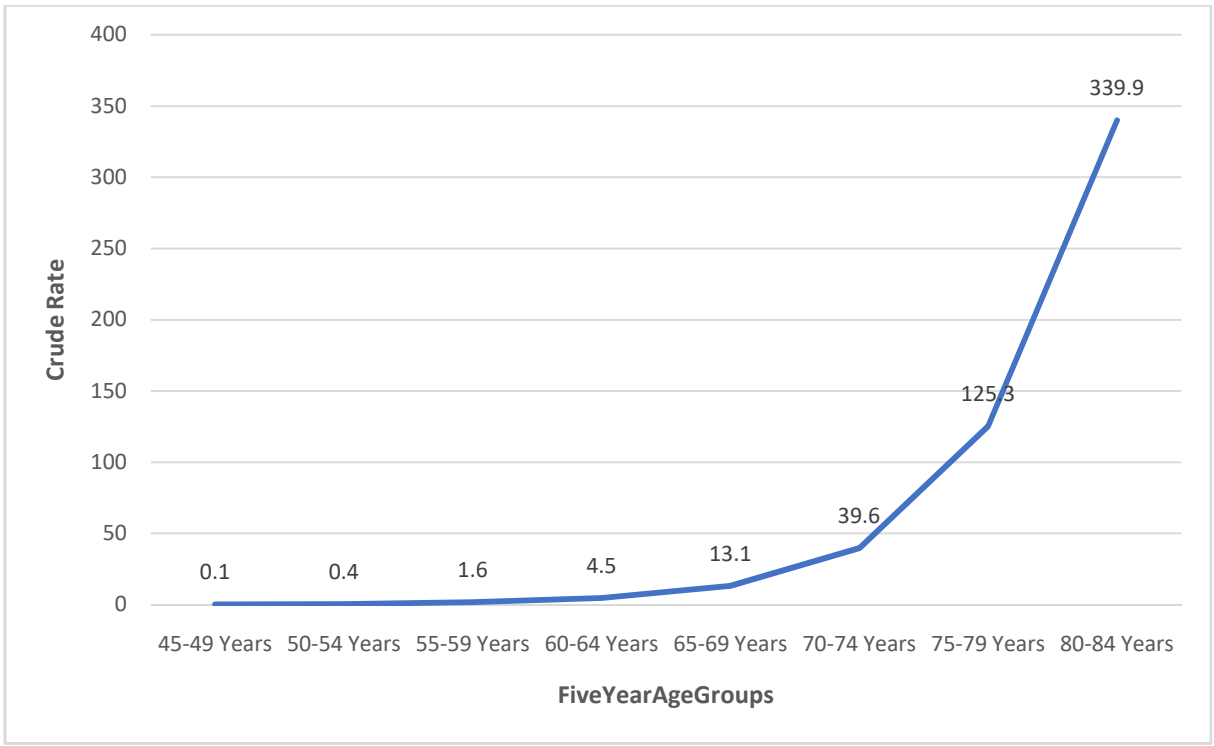


Figure 9: Line plot of Alzheimer's deaths crude rate by age.

The figure shows the relationship between Alzheimer's death crude rate and age. The X-axis represents the age interval (5 years) of Alzheimer's decedents; and on the Y-axis is the crude rate expressing the number of deaths per 100.000 person of the given population age interval.

4.2. Time series forecasting of Alzheimer's deaths records.

The time series run sequence plot (**Figure 10**) showed phases of increase and decrease in the number of Alzheimer's deaths by months. this gradually increases from July to January then decrease slightly in February and increase back in March before dropping down from April to June.For every year interval, the highest number of deaths was observed in January, and the lowest number in June.

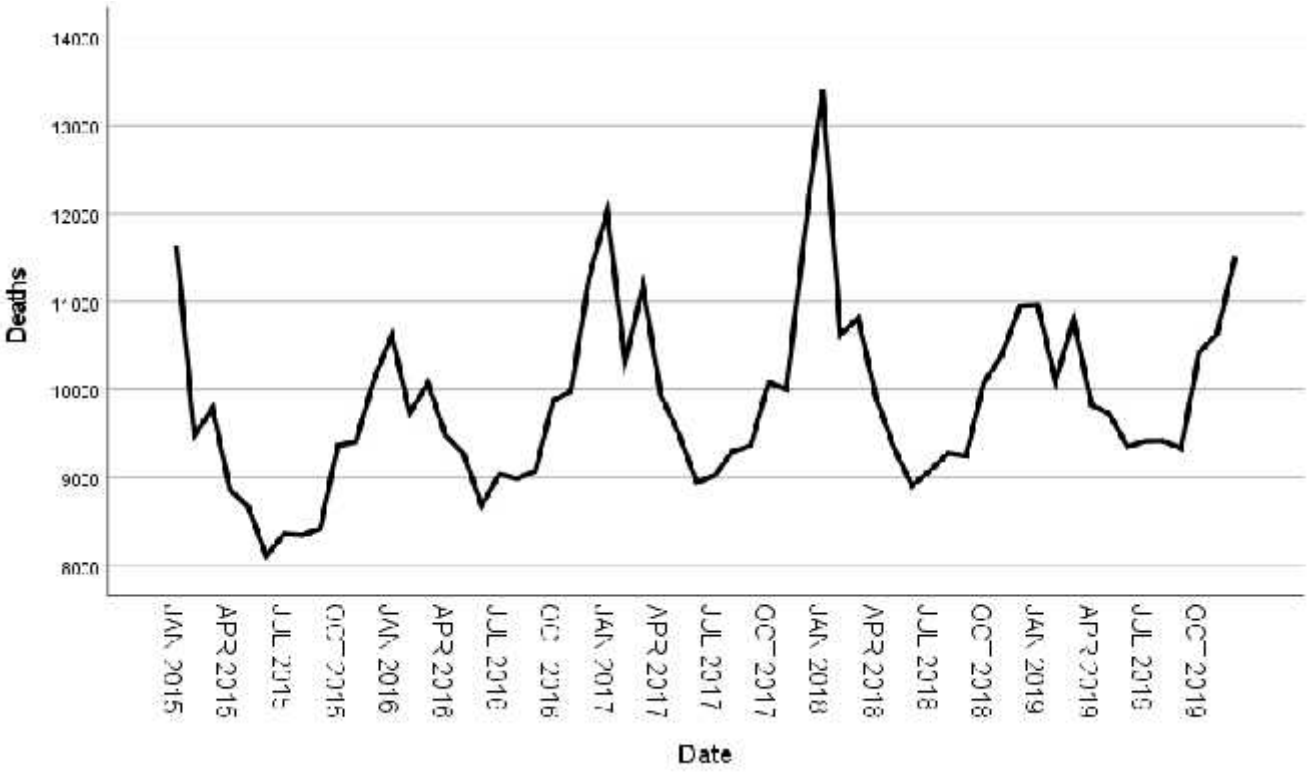
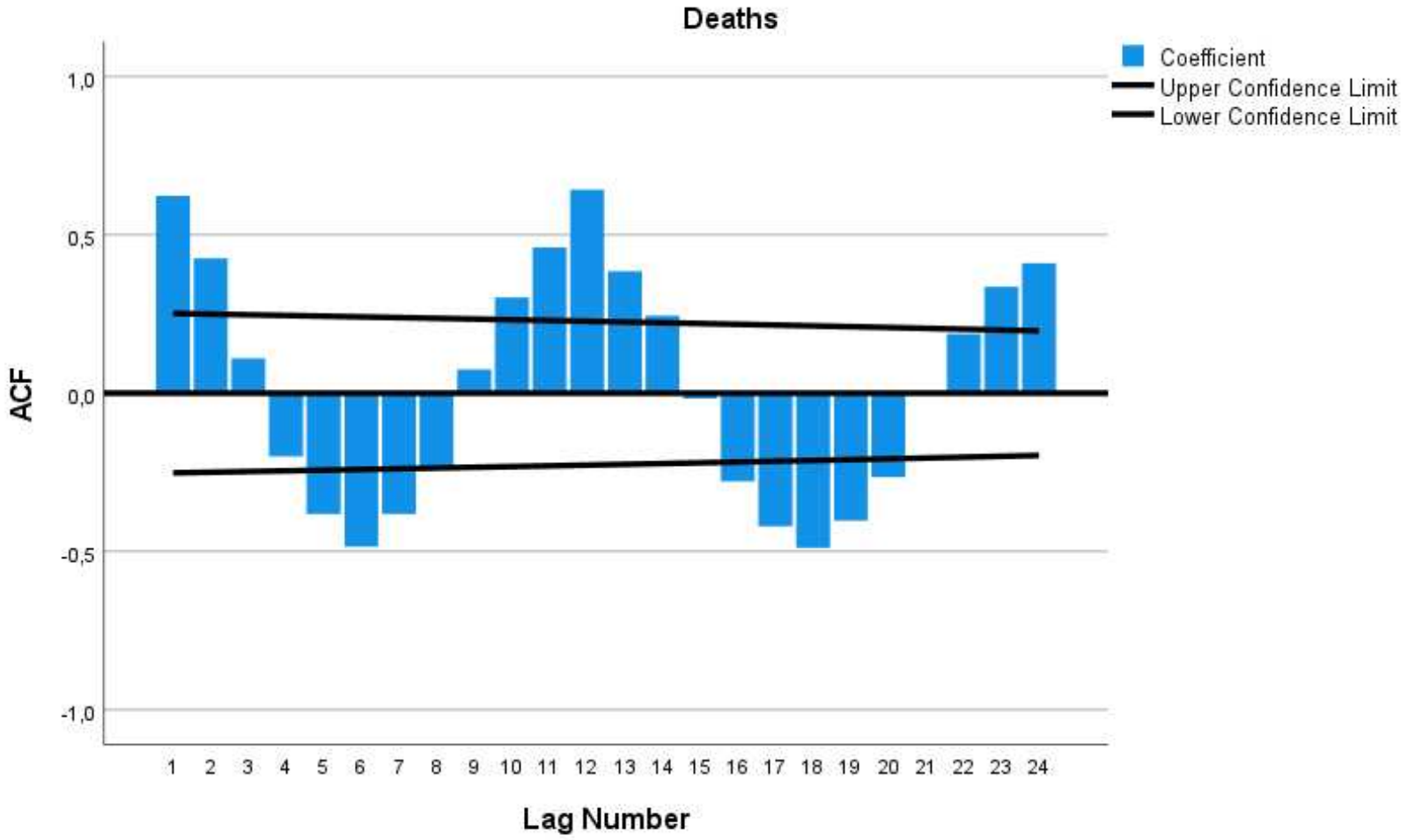


Figure 10: Run sequence plot of Alzheimer's deaths from 2015 to 2019.

The figure presents the number Alzheimer's Deaths by months. It consists of 5 years of monthly data collected from January 2015 to December 2019. On the Y-axis is the valid number of deaths records; and on the X-axis is the date in the format month and year.

The ACF and PACF plots of the original (figure 11) series revealed a seasonality pattern in the occurrence of Alzheimer's deaths, both showed significant peak at lag 1 and at lag 12. Also, the ACF showed repeated pattern after each 12 lags, indicating a yearly seasonality.


Deaths
Coefficient
Upper Confidence Limit
Lower Confidence Limit
ACF
1,0
0,5
0,0
-0,5
-1,0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
Lag Number

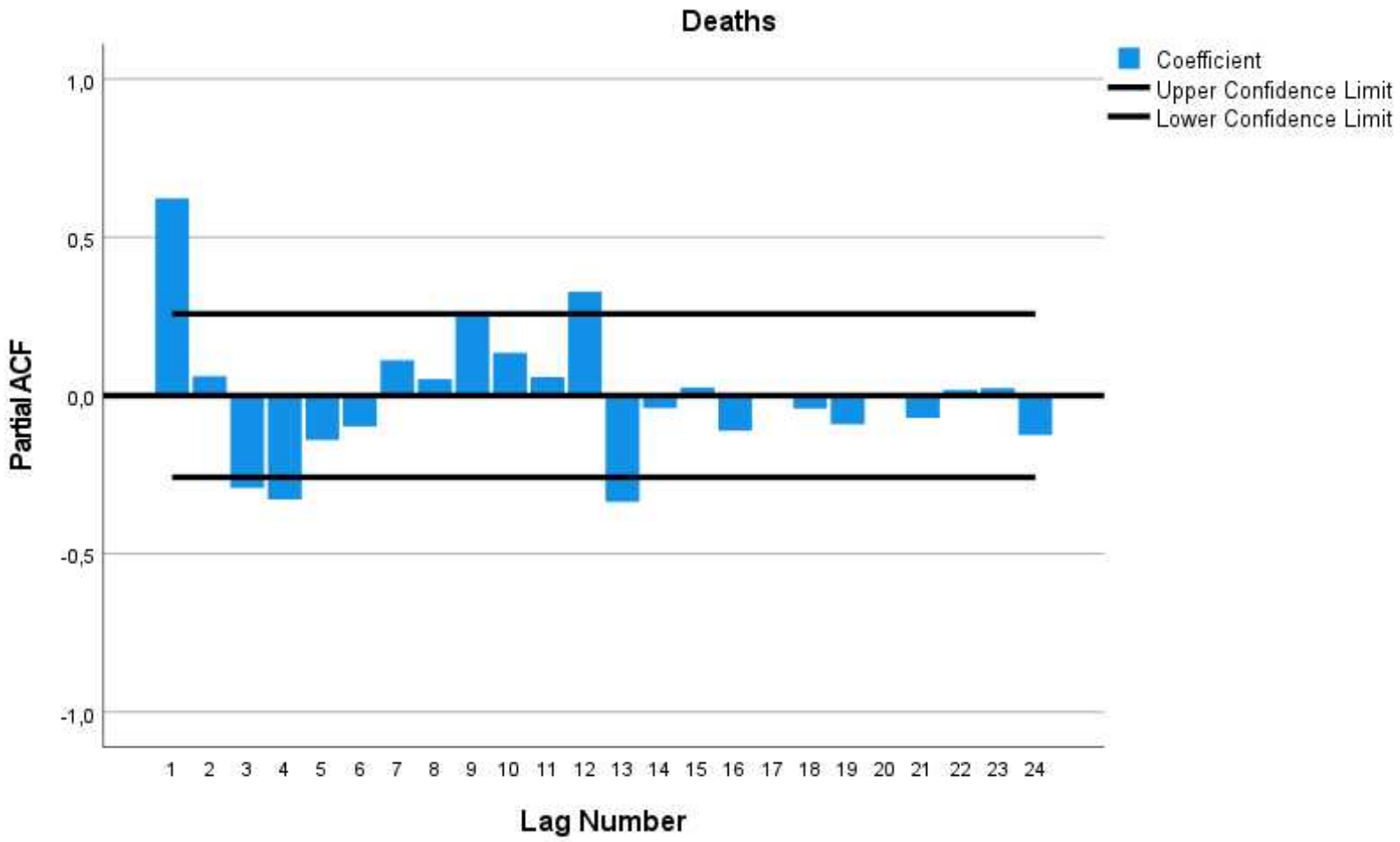


Figure 11: ACF and PACF plot of the original series.

The figures present the plot of the Autocorrelation (figure above) and Partial Autocorrelation (figure below) function of Alzheimer deaths. The blue color represents the significant coefficient at each lag; and the black line the upper (line superior to 0) and lower (line inferior to 0) confidence limit.

The Augmented Dickey-Fuller Test resultshowed that our original series was not stationary (p-value = 0,367) with a p-value higher than 0,05. Seasonal difference was applied to the series to achieve stationarity, and the Augmented Dickey-Fuller Test result of the seasonal differenced series confirmed its stationarity (p-value = 0,001).

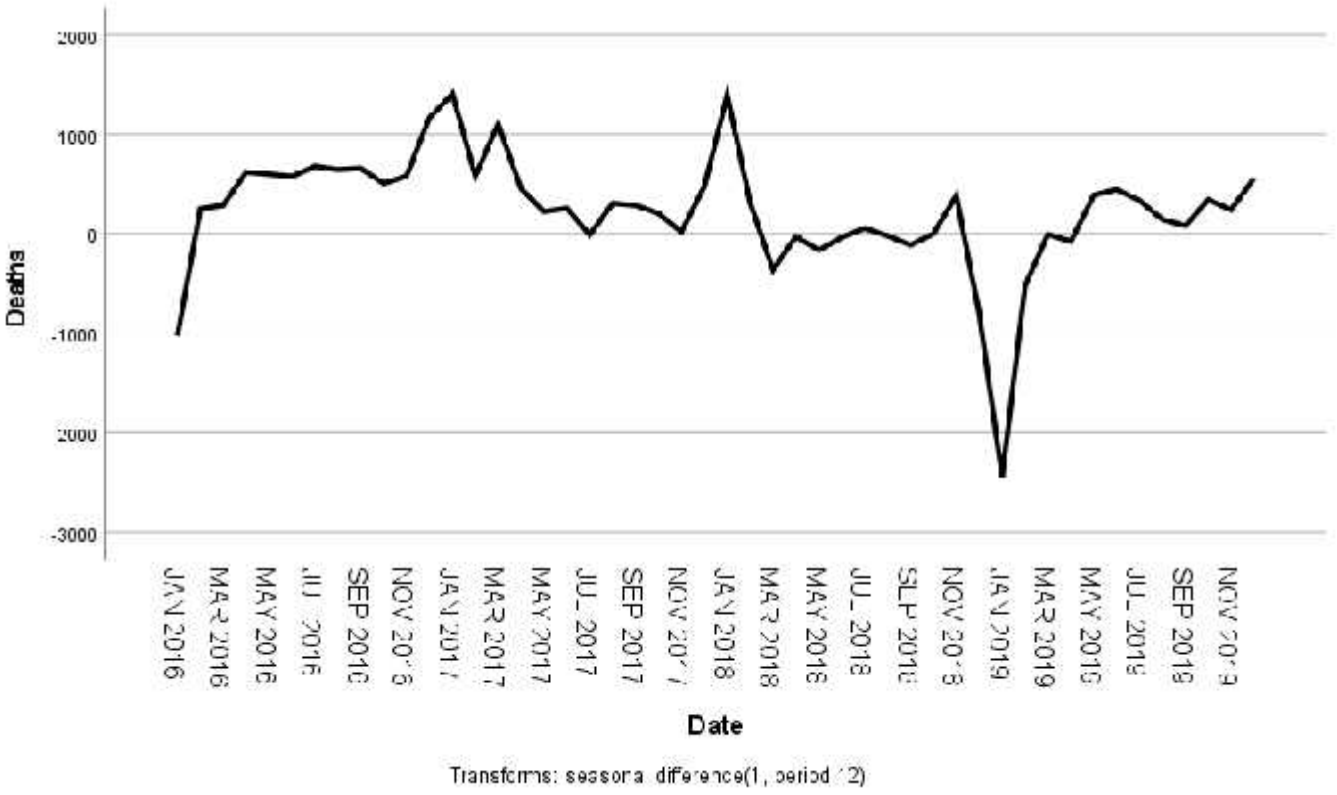
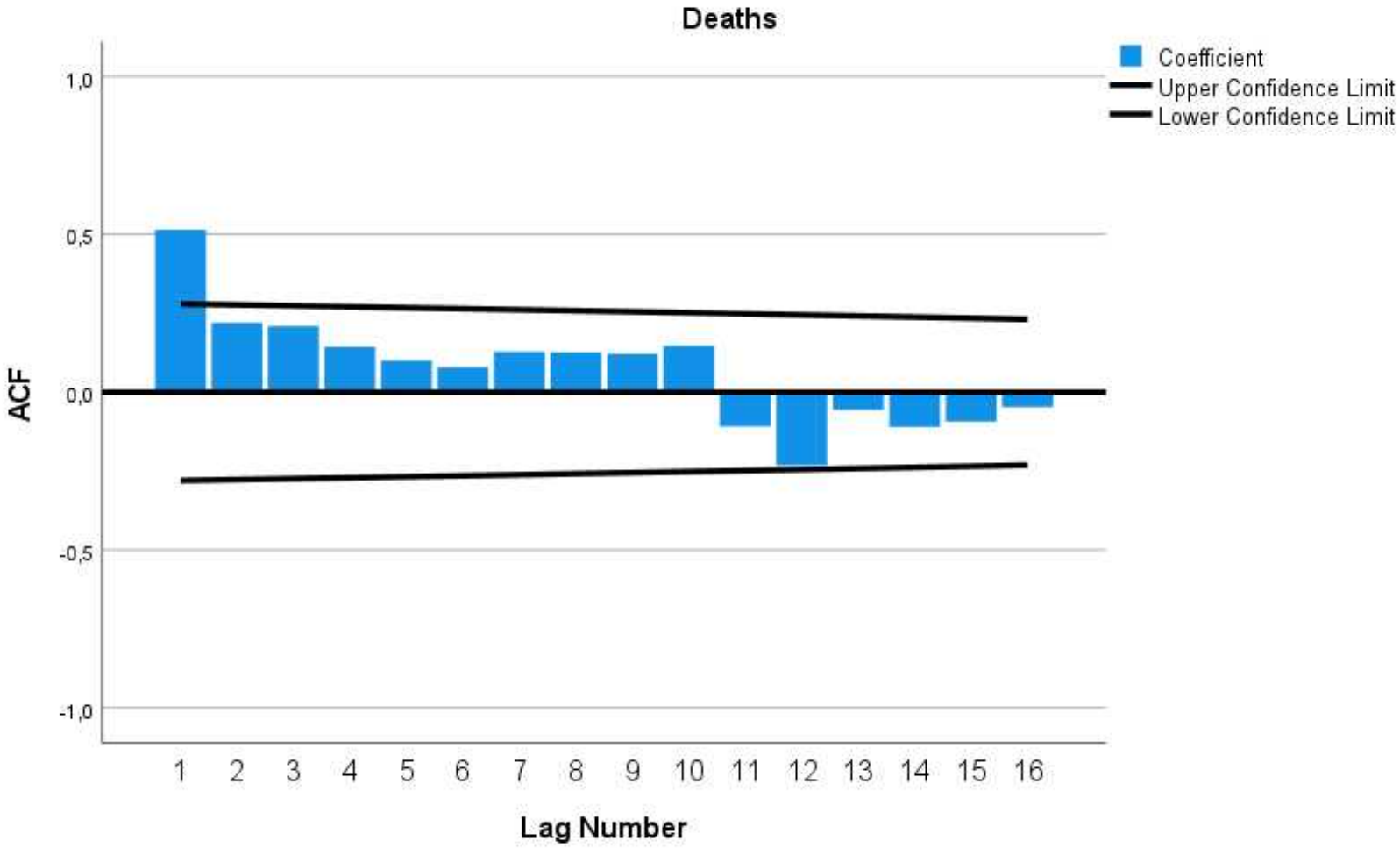


Figure 12: Run sequence plot of the seasonal differenced series.

The figure presents the number of Alzheimer's deaths by month from which have been subtracted the number recorded in the same month of the previous year.

The ACF of the seasonal differenced series showed only one significant positive pic at lag 1, and the PACF a significant positive pic at lag 1 and at lag 11**(figure 13)**. In reference to Table 1, SARIMA (1,0,1) (1,1,0), SARIMA (0,0,2) (2,1,0), and SARIMA (0,0,1) (2,1,0) all with constant weretested for our model fitting.


Deaths
Coefficient
Upper Confidence Limit
Lower Confidence Limit
ACF
1,0
0,5
0,0
-0,5
-1,0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Lag Number

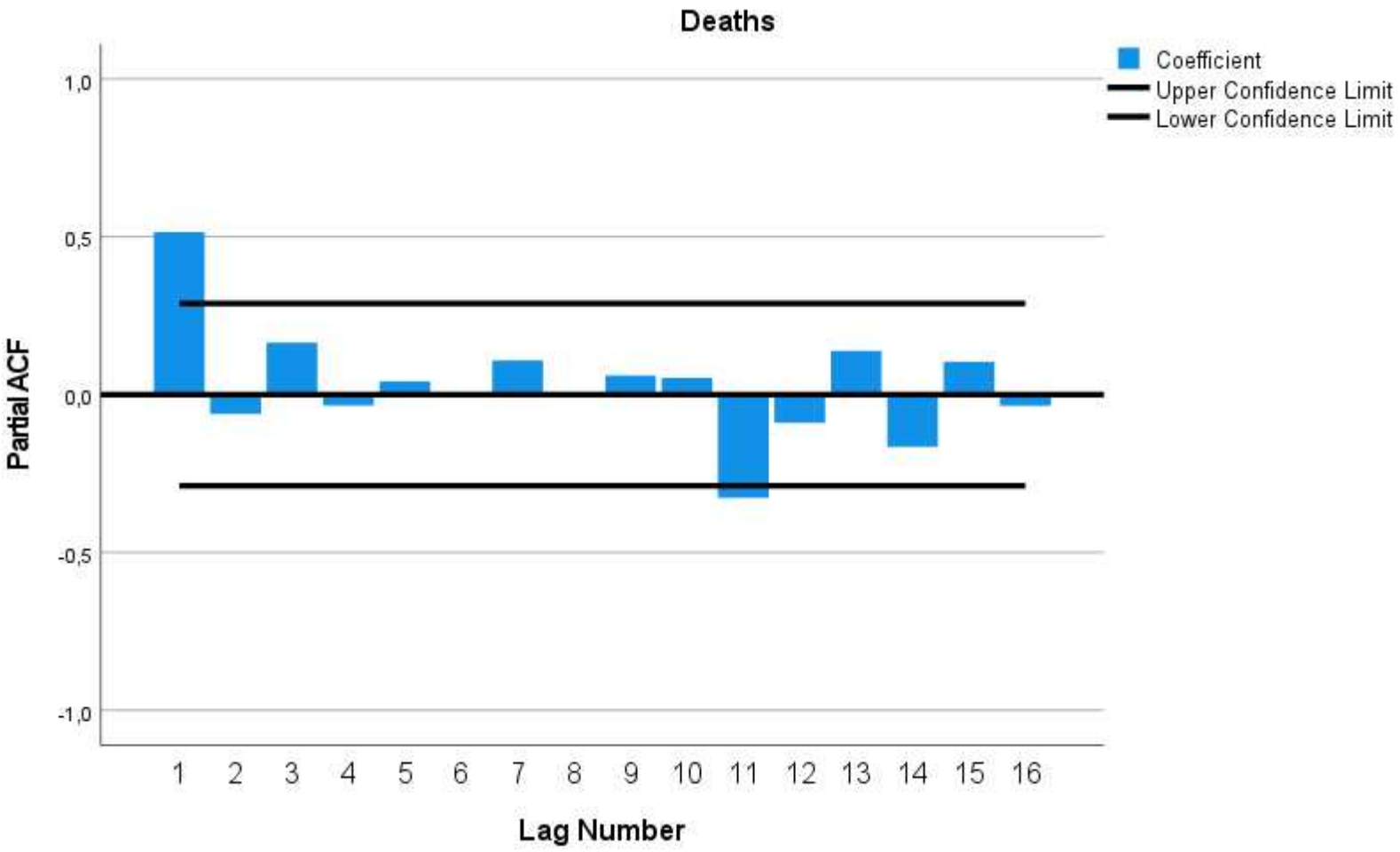


Figure 13: ACF and PACF of the seasonal differenced series.

The figures present the Autocorrelation and Partial Autocorrelation plot of the seasonal differenced series. The blue color represents the significant coefficient at each lag; and the black line the upper (line superior to 0) and lower (line inferior to 0) confidence limit.

The Stationary R-square, the Normalized BIC, and the significance of the components were used to compare the different models (**Table 2**). Among all, SARIMA (0,0,1) (2,1,0)yielded the lowest Normalized BIC, the highest Stationary R-Square, and the Highest ratio of significant components (**Figure 14**). Therefore, it was chosen as the best model to fit our data.

Table 2: SARIMA's models' statistics.

The table presents the results of the different SARIMA models tested. The first column represents the different models; c = with constant. The first line represents the criteria by which the models were compared; SRS = Stationary R-Square of the given model, N-BIC = Normalized BIC of the given model, Sig = Number of significant components in the given model (p, q, d, c, P, Q, D,).

SARIMA Models	SRS	N-BIC	Sig
(1,0,1) (1,1,0)c	0,421	12,675	4/5
(0,0,2) (2,1,0)c	0,656	12,258	3/4
(0,0,1) (2,1,0)c	0,656	12,154	4/4

The plot of the ACF and PACF of the residual of the chosen SARIMA model (0,0,1) (2,1,0)c showed that almost all the lags were not within the 95% CI (**Figure 14**). The lines plots of the fitted against the observed values (**Figure 15**) were matching at almost all of the values; highlightinga pretty good model. Although somevalues did not match, we took parsimony as the watchword and the model was used forforecasting.

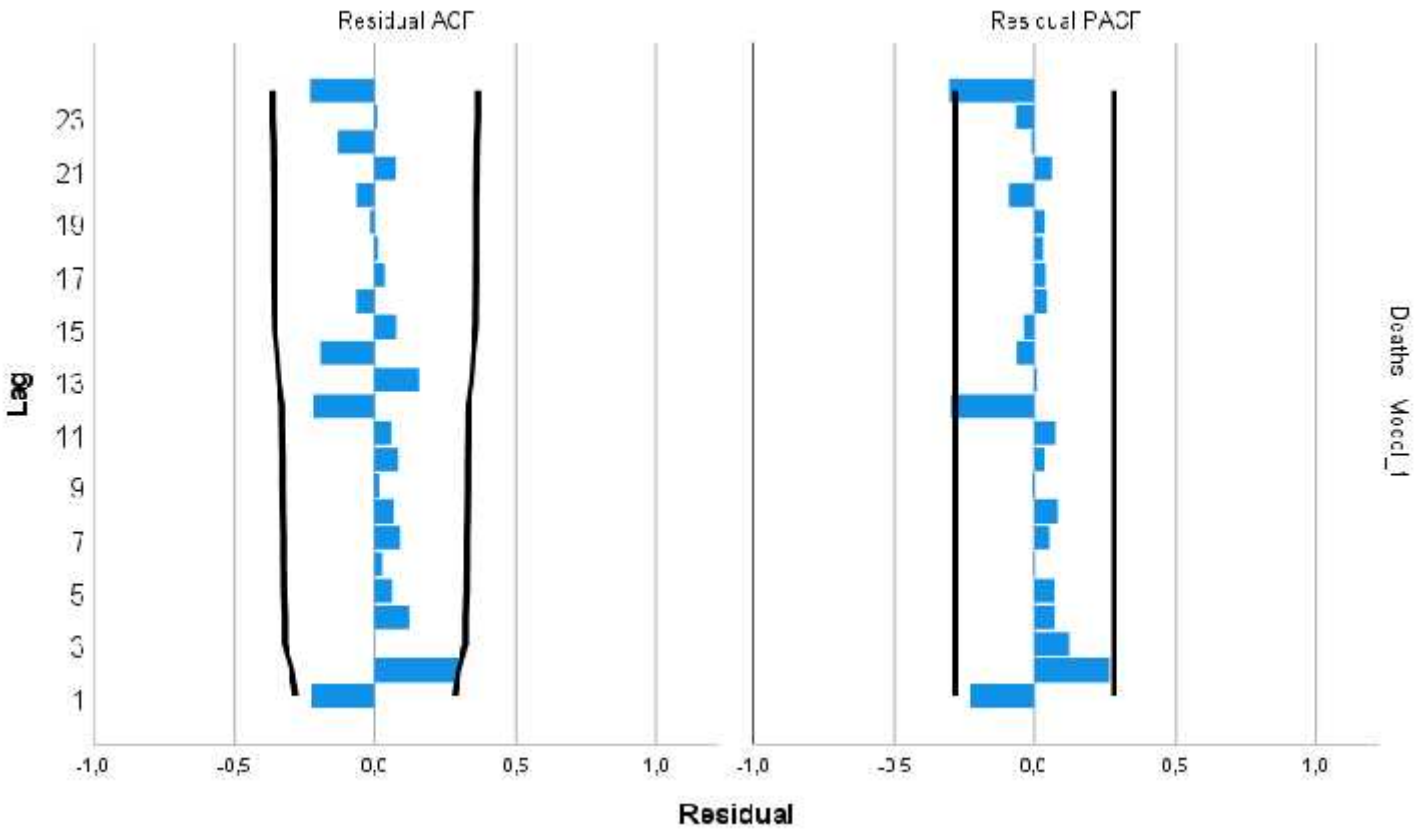


Figure 14: ACF and PACF plots of the residual of SARIMA model (0,0,1) (2,1,0)c.

The figures present the Autocorrelation and Partial Autocorrelation plot of the residual of SARIMA model (0,0,1) (2,1,0)c. The blue color represents the significant coefficient at each lag; and the black line the upper (line superior to 0) and lower (line inferior to 0) confidence limit.

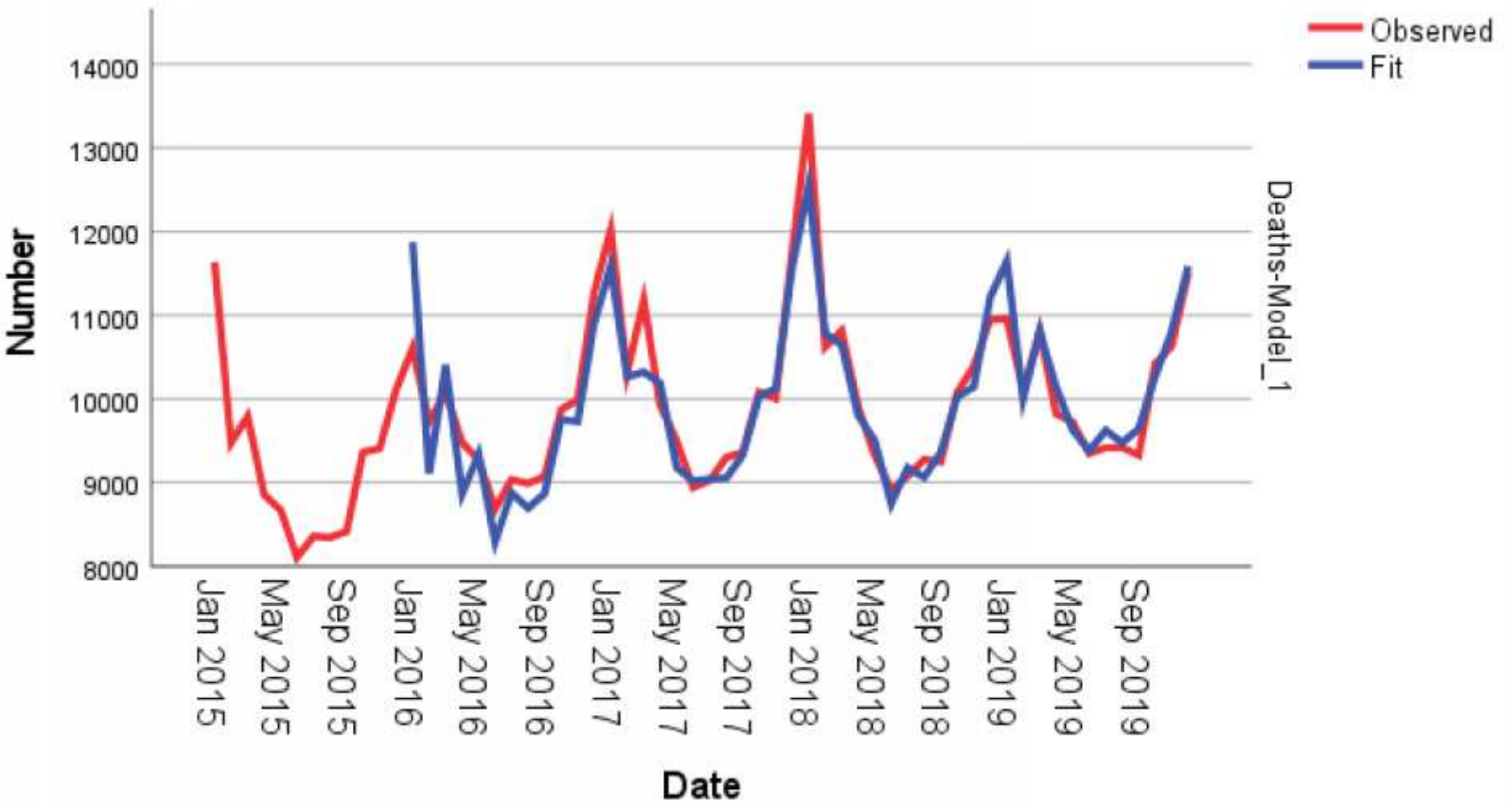


Figure 15: Plot of the fitted SARIMA (0,0,1) (2,1,0)c model against the observed series.

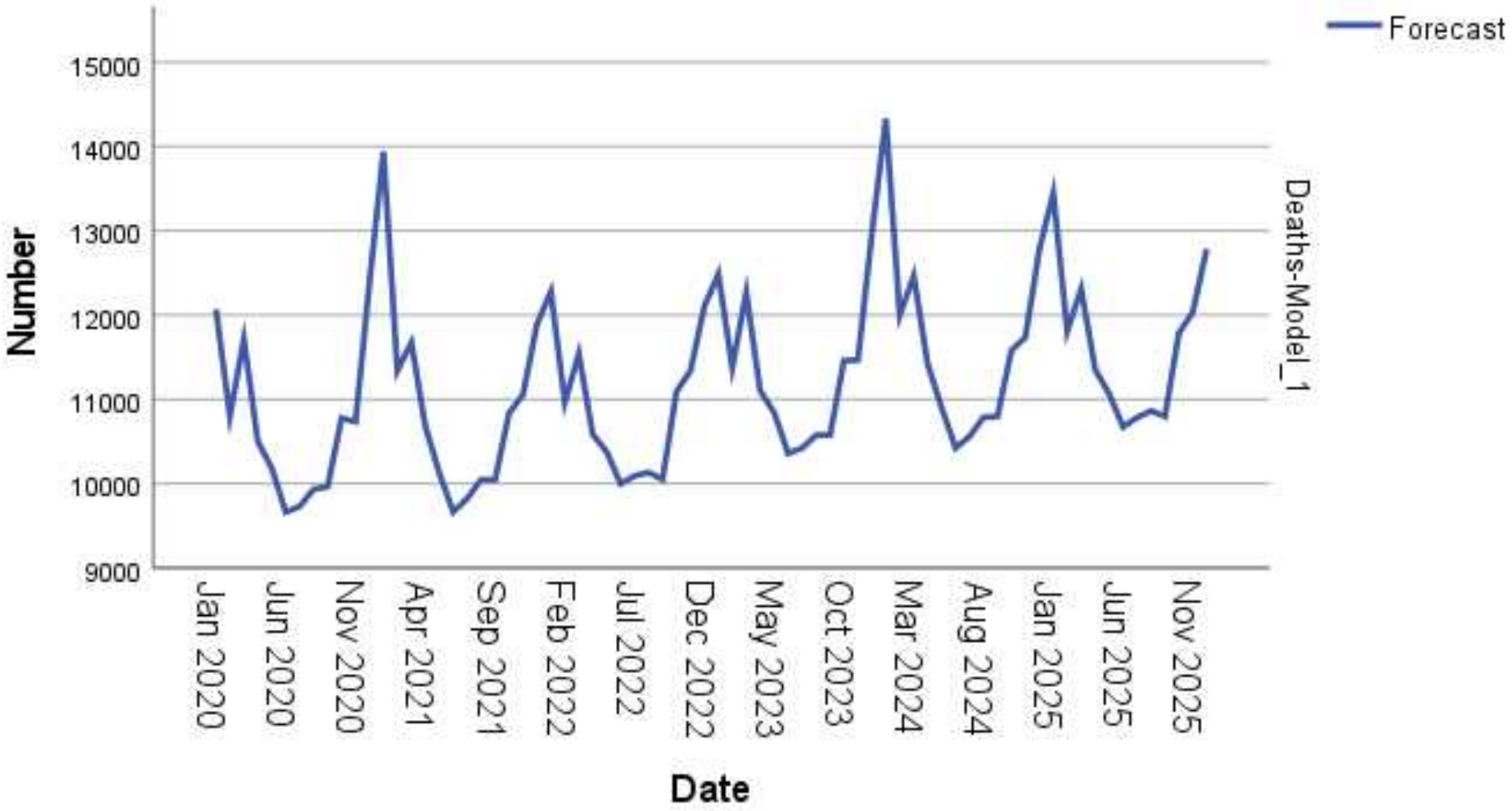
The figure presents the plot of the fitted against the observed series. The X-axis represents the number of deaths and the Y-axis the time period in month and year. The red line represents the observed values, and the blue line represents the model's fitted values.

Our forecast exercise was done 72 months ahead from December 2019 (**Figure 15**). The forecasted values for the first three years with the upper and lower confidence limit were extracted and presented in Table 3.

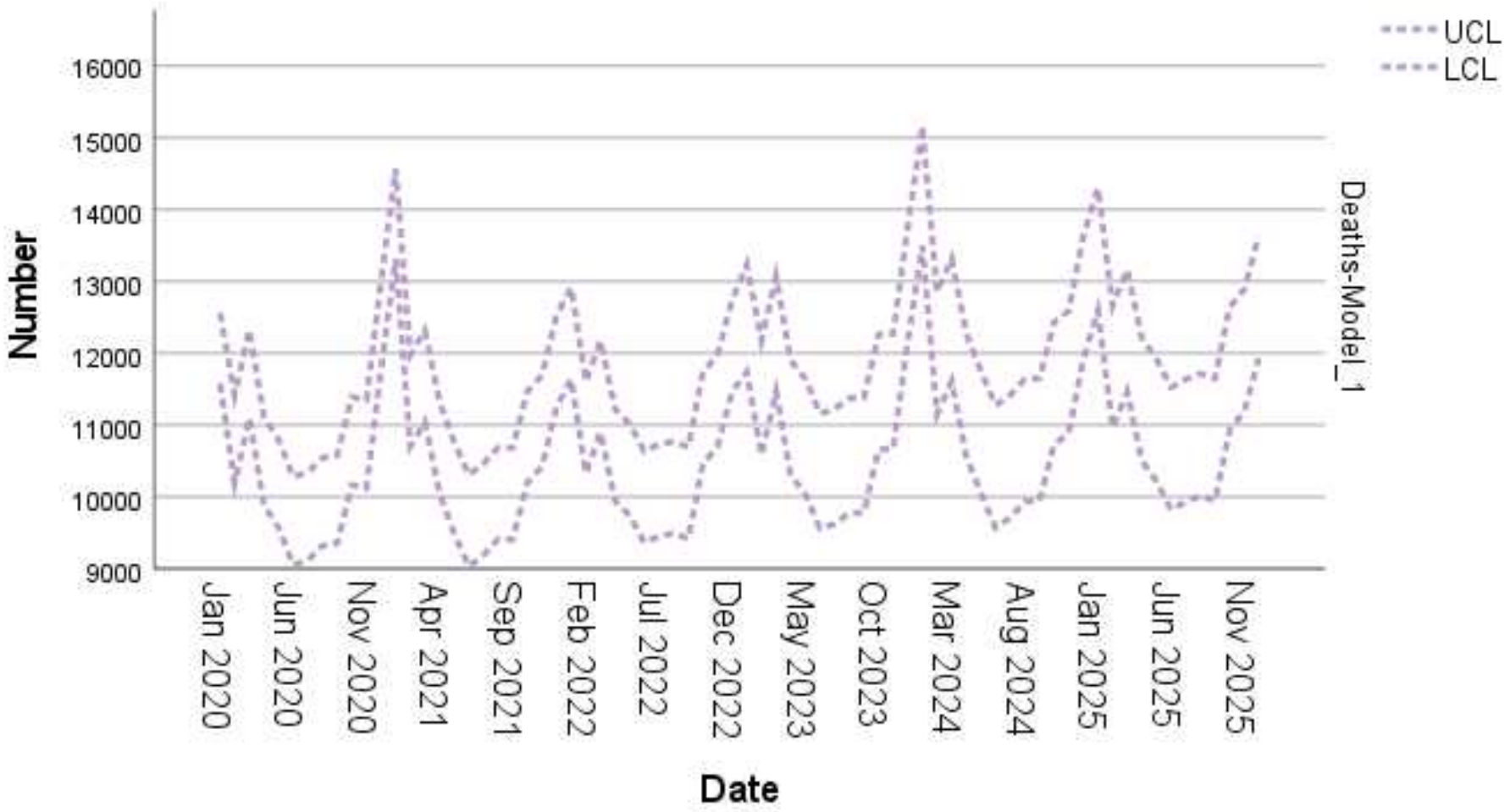
Table 3: Alzheimer's deaths Forecasting.

The table presents the predicted number of Alzheimer's death with upper and lower confidence limit by month from January 2020 to December 2022.

		2020	2021	2022			2020	2021	2022
Jan	**Forecast**	12.072	13.938	12.280	**Jul**	**Forecast**	9.732	9.825	10.094
	UCL	12.564	14.567	12.918		**UCL**	10.344	10.463	10.732
	LCL	11.578	13.310	11.643		**LCL**	9.120	9.188	9.456
Feb	**Forecast**	10.805	11.343	10.966	**Aug**	**Forecast**	9.934	10.048	10.134
	UCL	11.417	11.980	11.604		**UCL**	10.547	10.685	10.772
	LCL	10.193	10.705	10.328		**LCL**	9.322	9.410	9.496
Mar	**Forecast**	11.717	11.676	11.541	**Sep**	**Forecast**	9.966	10.046	10.053
	UCL	12.329	12.314	12.179		**UCL**	10.578	10.684	10.692
	LCL	11.105	11.039	10.903		**LCL**	9.354	9.409	9.415
Apr	**Forecast**	10.503	10.685	10.584	**Oct**	**Forecast**	10.778	10.836	11.095
	UCL	11.115	11.323	11.222		**UCL**	11.390	11.474	11.733
	LCL	9.891	10.048	9.946		**LCL**	10.166	10.199	10.457
May	**Forecast**	10.183	10.128	10.378	**Nov**	**Forecast**	10.736	11.058	11.341
	UCL	10.795	10.765	11.016		**UCL**	11.348	11.696	11.979
	LCL	9.571	9.490	9.740		**LCL**	10.124	10.421	10.702
Jun	**Forecast**	9.665	9.667	10.000	**Dec**	**Forecast**	12.384	11.892	12.106
	UCL	10.277	10.304	10.638		**UCL**	12.996	12.530	12.744
	LCL	9.053	9.029	9.362		**LCL**	11.772	11.255	11.468



16-A) Forecasted values of Alzheimer's deaths



16-B) Upper and Lower Confidence Limit of the forecasted values.

Figure 16: Plot of the forecasted Alzheimer's deaths with UCL and LCL.

The figure illustrates in (A) the plot of the forecasted values of Alzheimer's deaths from January 2020 to December 2025; in (B) the plot of the UCL (broken line above) and LCL (broken line below) of our forecasted values. On the X-axis are the dates coded as month and year, on the Y-axis are the number of deaths.

5. Discussion

We extracted from our results that AD can occur at early age (30-35 years); and that it affects more women than men in general. This aligns with **(Z. Yang & Levey, 2015)** results of 15.5% AD risk in female vs. 13.1% AD risk in male obtained after a Life Time Analysis of the Nationally representative Medicare Current Beneficiary Survey data from age 65 to death.However, we highlighted that this difference starts to be statistically significant from 70 years old to above; under the age of 70 both women and men are likely at the same risk of dying from the disease or developing it.

The significant increases in number of Alzheimer's deaths with age direct to two ideas:

- First, it can be the result of an increasing risk of developing the disease with age; as sustained by**(Farfel et al., 2019).**They authors applied logistic and linear regression to examine two relationships; one between age and log odds (logit) of AD dementia, and another between the age and the indices of AD dementia pathology (global cognition, episodic memory etc.).The results of the first relationship testedhighlighted a strong positive association betweenage andlog odds of AD dementia with 0,067 logit increases in number of AD dementia for each additional year of age; this corresponds to an OR of 1.070 (95% Confidence Interval [CI]: 1.051–1.088, $p < 0.001$. The results of the second relationship tested showed a strong linear association of age with global cognition and with episodic memory indices (all $p < 0{,}001$), and a non-linear association of age with binary AD diagnosis, β-amyloid load, and PHF tau tangle density indices (all $p < 0.05$).
- Second, it can be the result of a long survival time for those who manifested the disease at their early age as showed by **(Helzner et al., 2008)**.They authors obtained 9,9 years (95% CI 6,8 - 13,0) postdiagnosis survival for Alzheimer's patients aged less than 75 at diagnosis, 6,9 years (95% CI 6,8 - 13,0) for those aged 75-84 years at diagnosis, and 4,4 years (95% CI 3,5 – 5,3) for those aged 85+ (p-value < 0,0001; log-rank statistic = 20,8); the data included 2.125 individuals sampled between the 155th and 181st Streets in Manhattan.

Furthermore, Alzheimer's deaths showed a wide distribution between States. The lowest crude rate was observed in the District of Columbia followed by Maryland, Alaska, New York, and Nevada in ascending order. The highest crude rate was observed in South Dakota followed by Mississippi, North Dakota, Vermont, and Arkansas in descending order.

From these, it can be seen that the Southeast, the Midwest, and the West are the region most affected by the disease as sustained by **(Taylor et al., 2017).**This result direct to two ideas; either there was a very unbalance number of AD patients in reference to the population between states in 2019, or there was a difference in the level of disease management between States during that period.

In addition, we retained that Whites Americans are most affected by AD than none-White Americans (Black or African American, Asian or Pacific Islander, American Indian or Alaska Native); with a 748,025% increase in numbers of deaths in the first group compared to the number in the second group. When grouping the population in accordance to Hispanic Origin, Not Hispanic or Latino died more from Alzheimer's than Hispanic or Latino; the number of deaths in the first group was 13,763 times the number of deaths in the second group. These show an increase in Alzheimer's mortality trend since the study of **(Steenland et al., 2009).** who sustained a 56% increase in White death rate compared to non-whites and a 72% increase in non-Hispanic compared to Hispanic after analysis of US Alzheimer disease (AD) mortality rates from 1999 to 2004.More specifically, the same trait was observed when considering places of deaths. We further highlight that within the non-white population, Blacks or Africans Americans are the most affected by the disease; before Asian or Pacific Islander, and American Indian or Alaska Native who are the less affected.

Also, we noticed a higher number of Alzheimer's deaths in facilities than in decedent's home. Between facilities, medical facility counted the lowest number of deaths aligning behind Hospice facility and Nursing Home/Long term care in ascending order. These can be the results of either an unbalance in the number of Alzheimer's patients in these places; or a difference in the quality of care giving to Alzheimer's patients in respective places in line with the study of **(Xu et al., 2020).**The study's results highlighted a significant association between the access to care facility resource in hospitals and nursing home/long term care and dementia decease at respective places (all p-values< 0,001), and between Medicare and Medicaid investments in various health agencies/services and proportion of dementia death at various places (nursing home/long term care deaths, hospital deaths and home deaths) (all $p < 0.005$).

Therewith, the seasonality pattern observed in our results directed us to deduct that people with Alzheimer in the US are less likely to die in spring (decreasing number of deaths from March to June) and more in summer, autumn and winter (increasing number of deaths from July to February); with a higher risk in winter (peak in January).

6. Conclusion

We retain from our study that as already known in the literature, Alzheimer affect more women than men; more Hispanic or Latino than Not Hispanic or Latino; and more Whites than non-white. Patients die more in nursing home/long term care and in their living houses because of access to care facility resource, and the risk of dying or developing the disease increases with age.

For an update, we highlighted that Alzheimer can occur at early adults age (around 35 years old); and the difference in risk between men and women start to be considerable from age 70 to above. Alzheimer's patients are more likely to die in summer, autumn and in winter (where we observe the higher risk); and less likely to die in in spring. White patients are the most dying in the different places of deaths (Nursing Home/Long term Care, Hospital facility, Medical Facility, Hospice facility, Decedents homes). Within the non-white population, Alzheimer affects more Blacks than Asians, and more Asians than American Indians.

The Government of the US could think about how to put an accent on the equipment and the quality of care in Nursing home, and create an educational system for family with Alzheimer disease patients. Future researchescould be directed on the analysis of survival time for Alzheimer's patients; on the evaluation of disease management between States in the US; and on the evaluation of the disease risk and its development according to meteorological seasons.

References

1. 2020 Alzheimer's disease facts and figures. (2020). *Alzheimer's and Dementia*, *16*(3), 391–460. https://doi.org/10.1002/alz.12068
2. Aditya Satrio, C. B., Darmawan, W., Nadia, B. U., & Hanafiah, N. (2021). Time series analysis and forecasting of coronavirus disease in Indonesia using ARIMA model and PROPHET. *Procedia Computer Science*, *179*, 524–532. https://doi.org/10.1016/j.procs.2021.01.036
3. Anderson, O. D. (1976). On the transformation of raw time series data: A review. *Statistische Hefte*, *17*(4), 285–289. https://doi.org/10.1007/BF02923038
4. ArunKumar, K. E., Kalaga, D. V., Sai Kumar, C. M., Chilkoor, G., Kawaji, M., & Brenza, T. M. (2021). Forecasting the dynamics of cumulative COVID-19 cases (confirmed, recovered and deaths) for top-16 countries using statistical machine learning models: Auto-Regressive Integrated Moving Average (ARIMA) and Seasonal Auto-Regressive Integrated Moving Average (SARIMA). *Applied Soft Computing*, *103*, 107161. https://doi.org/10.1016/j.asoc.2021.107161
5. Berchtold, N. C., & Cotman, C. W. (1998). Evolution in the conceptualization of dementia and Alzheimer's disease: Greco-Roman period to the 1960s. *Neurobiology of Aging*, *19*(3), 173–189. https://doi.org/10.1016/S0197-4580(98)00052-9
6. Centers for Disease Control and Prevention, National Center for Health Statistics. Underlying Cause of Death 1999-2019 on CDC WONDER Online Database, released in 2020. Data are from the Multiple Cause of Death Files, 1999-2019, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at http://wonder.cdc.gov/ucd-icd10.html on Apr 14, 2021 6:36:24 AM
7. Chawsheen, T. A., & Broom, M. (2017). Seasonal time-series modeling and forecasting of monthly mean temperature for decision making in the Kurdistan Region of Iraq. *Journal of Statistical Theory and Practice*, *11*(4), 604–633. https://doi.org/10.1080/15598608.2017.1292484
8. Cummings, J. L., Morstorf, T., & Zhong, K. (2014). Alzheimer's disease drug-development pipeline: Few candidates, frequent failures. *Alzheimer's Research and Therapy*, *6*(4), 37. https://doi.org/10.1186/alzrt269
9. Daniels, M. J., & Pourahmadi, M. (2009). Modeling covariance matrices via partial autocorrelations. *Journal of Multivariate Analysis*, *100*(10), 2352–2363. https://doi.org/10.1016/j.jmva.2009.04.015

10. Dickey, D. A., & Fuller, W. A. (1981). *Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root* (Vol. 49, Issue 4). https://about.jstor.org/terms
11. Diks, C., & Panchenko, V. (2007). *Institute of Statistical Science, Academia Sinica NONPARAMETRIC TESTS FOR SERIAL INDEPENDENCE BASED ON QUADRATIC FORMS*. *17*(1), 1–4. https://doi.org/10.2307/26432512
12. Domínguez, M. A., & Lobato, I. N. (2003). Testing the Martingale Difference Hypothesis. *Econometric Reviews*, *22*(4), 351–377. https://doi.org/10.1081/ETC-120025895
13. Farfel, J. M., Yu, L., Boyle, P. A., Leurgans, S., Shah, R. C., Schneider, J. A., & Bennett, D. A. (2019). Alzheimer's disease frequency peaks in the tenth decade and is lower afterwards. *Acta Neuropathologica Communications*, *7*(1), 104. https://doi.org/10.1186/s40478-019-0752-0
14. Fisher, C. K., Smith, A. M., Walsh, J. R., Simon, A. J., Edgar, C., Jack, C. R., Holtzman, D., Russell, D., Hill, D., Grosset, D., Wood, F., Vanderstichele, H., Morris, J., Blennow, K., Marek, K., Shaw, L. M., Albert, M., Weiner, M., Fox, N., … Kubick, W. (2019). Machine learning for comprehensive forecasting of Alzheimer's Disease progression. *Scientific Reports*, *9*(1). https://doi.org/10.1038/s41598-019-49656-2
15. Förstl, H., & Kurz, A. (1999). Clinical features of Alzheimer's disease. In *European Archives of Psychiatry and Clinical Neuroscience* (Vol. 249, Issue 6, pp. 288–290). Springer Verlag. https://doi.org/10.1007/s004060050101
16. Garre-Olmo, J. (2018). Epidemiology of alzheimer's disease and other dementias. In *Revista de Neurologia* (Vol. 66, Issue 11, pp. 377–386). Revista de Neurologia. https://doi.org/10.33588/rn.6611.2017519
17. Ghoudi, K., Kulperger, R. J., & Rémillard, B. (2001). A nonparametric test of serial independence for time series and residuals. *Journal of Multivariate Analysis*, *79*(2), 191–218. https://doi.org/10.1006/jmva.2000.1967
18. Ghoudi, K., & Rémillard, B. (2018). Serial independence tests for innovations of conditional mean and variance models. *Test*, *27*(1), 3–26. https://doi.org/10.1007/s11749-016-0521-3
19. Gubner, J. A. (2006). Probability and Random Processes for Electrical and Computer Engineers. In *Probability and Random Processes for Electrical and Computer Engineers*. Cambridge University Press. https://doi.org/10.1017/cbo9780511813610
20. Häkkinen, S., Chu, S. A., & Lee, S. E. (2020). Neuroimaging in genetic frontotemporal dementia and amyotrophic lateral sclerosis. In *Neurobiology of*

Disease (Vol. 145, p. 105063). Academic Press Inc. https://doi.org/10.1016/j.nbd.2020.105063

21. Hebert, L. E., Weuve, J., Scherr, P. A., & Evans, D. A. (2013). Alzheimer disease in the United States (2010-2050) estimated using the 2010 census. *Neurology*, *80*(19), 1778–1783. https://doi.org/10.1212/WNL.0b013e31828726f5
22. Helzner, E. P., Scarmeas, N., Cosentino, S., Tang, M. X., Schupf, N., & Stern, Y. (2008). Survival in Alzheimer disease: A multiethnic, population-based study of incident cases. *Neurology*, *71*(19), 1489–1495. https://doi.org/10.1212/01.wnl.0000334278.11022.42
23. Howard, R., McShane, R., Lindesay, J., Ritchie, C., Baldwin, A., Barber, R., Burns, A., Dening, T., Findlay, D., Holmes, C., Hughes, A., Jacoby, R., Jones, R., Jones, R., McKeith, I., Macharouthu, A., O'Brien, J., Passmore, P., Sheehan, B., … Phillips, P. (2012). Donepezil and Memantine for Moderate-to-Severe Alzheimer's Disease. *New England Journal of Medicine*, *366*(10), 893–903. https://doi.org/10.1056/nejmoa1106668
24. Jain, G., & Mallick, B. (2017). A Study of Time Series Models ARIMA and ETS. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2898968
25. Jalil, A., & Rao, N. H. (2019). Time Series Analysis (Stationarity, Cointegration, and Causality). In *Environmental Kuznets Curve (EKC)* (pp. 85–99). Elsevier. https://doi.org/10.1016/b978-0-12-816797-7.00008-4
26. John Wainwright, Mark Mulligan (Eds), Environmental Modelling: 1 Finding Simplicity in Complexity, Second Edition, pp. 133-149, 2013
27. Karch, J. (2020). Improving on Adjusted R-Squared. *Collabra: Psychology*, *6*(1). https://doi.org/10.1525/collabra.343
28. Khan, F. M., & Gupta, R. (2020). ARIMA and NAR based prediction model for time series analysis of COVID-19 cases in India. *Journal of Safety Science and Resilience*, *1*(1), 12–18. https://doi.org/10.1016/j.jnlssr.2020.06.007
29. Kuan, C.-M. (2008). *LECTURE ON TIME SERIES DIAGNOSTIC TESTS*.
30. Lancet, T. (2011). The three stages of Alzheimer's disease. In *The Lancet* (Vol. 377, Issue 9776, p. 1465). Elsevier B.V. https://doi.org/10.1016/S0140-6736(11)60582-5
31. Ledolter, J., & Abraham, B. (1981). Parsimony and Its Importance in Time Series Forecasting. *Technometrics*, *23*(4), 411. https://doi.org/10.2307/1268232
32. Li, B., & Babu, G. J. (2019). *A Graduate Course on Statistical Inference*. Springer New York. https://doi.org/10.1007/978-1-4939-9761-9

33. Lian, H. (2014). Semiparametric bayesian information criterion for model selection in ultra-high dimensional additive models. *Journal of Multivariate Analysis*, *123*, 304–310. https://doi.org/10.1016/j.jmva.2013.09.015
34. Liu, H., Zhang, Y., Zhao, H., Du, Y., & Liu, X. (2021). The heterogeneity among subgroups of haplogroup J influencing Alzheimer's disease risk. *Journal of Advanced Research*. https://doi.org/10.1016/j.jare.2021.02.003
35. Luetkepohl, H., & Xu, F. (2009). The Role of the Log Transformation in Forecasting Economic Variables. *CESifo Working Paper Series*. https://ideas.repec.org/p/ces/ceswps/_2591.html
36. Moffat, I. U., & Akpan, E. A. (2019). White Noise Analysis: A Measure of Time Series Model Adequacy. *Applied Mathematics*, *10*(11), 989–1003. https://doi.org/10.4236/am.2019.1011069
37. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Wiley: Introduction to Time Series Analysis and Forecasting - Douglas C. Montgomery, Cheryl L. Jennings, Murat Kulahci* (p. 472).
38. Mrna, L., & Hornik, P. (2016). Autocorrelation function for monitoring the gap between the steel plates during laser welding. *Physics Procedia*, *83*, 1223–1232. https://doi.org/10.1016/j.phpro.2016.08.129
39. Nankervis, J. C., & Savin, N. E. (2010). Testing for Serial Correlation: Generalized Andrews-Ploberger Tests. *Source: Journal of Business & Economic Statistics*, *28*(2), 246–255. https://doi.org/10.1198/]bes.2009.08115
40. Phillips, J. R., Matar, E., Ehgoetz Martens, K. A., Moustafa, A. A., Halliday, G. M., & Lewis, S. J. G. (2021). Evaluating a novel behavioral paradigm for visual hallucinations in Dementia with Lewy bodies. *Aging Brain*, *1*, 100011. https://doi.org/10.1016/j.nbas.2021.100011
41. Portet, S. (2020). A primer on model selection using the Akaike Information Criterion. *Infectious Disease Modelling*, *5*, 111–128. https://doi.org/10.1016/j.idm.2019.12.010
42. Prathipati, B., Rohini, P., Kola, P. K., & Reddy Danduga, R. C. S. (2021). Neuroprotective effects of curcumin loaded solid lipid nanoparticles on homocysteine induced oxidative stress in vascular dementia. *Current Research in Behavioral Sciences*, *2*, 100029. https://doi.org/10.1016/j.crbeha.2021.100029
43. ROSCA, E. R. (2010). Stationary And Non-Stationary Time Series. *The Annals of the "Stefan Cel Mare" University of Suceava. Fascicle of The Faculty of Economics and Public Administration*, *10*(1(11)), 177–186.

https://ideas.repec.org/a/scm/ausvfe/v10y2010i1(11)p177-186.html

44. Salisu, A. A., Oloko, T. F., & Oyewole, O. J. (2016). Testing for martingale difference hypothesis with structural breaks: Evidence from Asia–Pacific foreign exchange markets. *Borsa Istanbul Review*, *16*(4), 210–218. https://doi.org/10.1016/j.bir.2016.09.001
45. Schroeter, M. L., Stein, T., Maslowski, N., & Neumann, J. (2009). Neural correlates of Alzheimer's disease and mild cognitive impairment: A systematic and quantitative meta-analysis involving 1351 patients. *NeuroImage*, *47*(4), 1196–1206. https://doi.org/10.1016/j.neuroimage.2009.05.037
46. Shampo, M. A., Kyle, R. A., & Steensma, D. P. (2013). Alois Alzheimer–Alzheimer Disease. In *Mayo Clinic Proceedings* (Vol. 88, Issue 12, p. e155). Elsevier Ltd. https://doi.org/10.1016/j.mayocp.2013.01.031
47. Sharma, N., & Singh, A. N. (2016). Exploring biomarkers for Alzheimer's disease. In *Journal of Clinical and Diagnostic Research* (Vol. 10, Issue 7, pp. KE01–KE06). Journal of Clinical and Diagnostic Research. https://doi.org/10.7860/JCDR/2016/18828.8166
48. Shin, Y., & Schmidt, P. (1992). The KPSS stationarity test as a unit root test. *Economics Letters*, *38*(4), 387–392. https://doi.org/10.1016/0165-1765(92)90023-R
49. c Bisgaard, Murat Kulahci, (2011), Time Series Analysis and Forecasting by Example, opyright © 2011 by John Wiley & Sons, Inc. Hoboken, New Jersey.
50. Steenland, K., MacNeil, J., Vega, I., & Levey, A. (2009). Recent trends in Alzheimer disease mortality in the United States, 1999 to 2004. *Alzheimer Disease and Associated Disorders*, *23*(2), 165–170. https://doi.org/10.1097/WAD.0b013e3181902c3e
51. Stępkowski, D., Woźniak, G., & Studnicki, M. (2015). Correlation of Alzheimer's Disease Death Rates with Historical Per Capita Personal Income in the USA. *PLOS ONE*, *10*(5), e0126139. https://doi.org/10.1371/journal.pone.0126139
52. Tackenberg, C., Kulic, L., & Nitsch, R. M. (2020). Familial Alzheimer's disease mutations at position 22 of the amyloid β-peptide sequence differentially affect synaptic loss, tau phosphorylation and neuronal cell death in an ex vivo system. *PLoS ONE*, *15*(9 September 2020). https://doi.org/10.1371/journal.pone.0239584
53. Taylor, C. A., Greenlund, S. F., McGuire, L. C., Lu, H., & Croft, J. B. (2017). Deaths from Alzheimer's Disease — United States, 1999–2014. *MMWR. Morbidity and Mortality Weekly Report*, *66*(20), 521–526.

https://doi.org/10.15585/mmwr.mm6620a1

54. Tsay, R. S. (2001). Time series and forecasting: Brief history and future research. In *Statistics in the 21st Century* (pp. 121–131). CRC Press. https://doi.org/10.2307/2669408
55. Valipour, M. (2015). Long-term runoff study using SARIMA and ARIMA models in the United States. *Meteorological Applications*, *22*(3), 592–598. https://doi.org/10.1002/met.1491
56. Victor, O. O. (n.d.). *Martingale Theory*.
57. Weller, J., & Budson, A. (2018). Current understanding of Alzheimer's disease diagnosis and treatment. In *F1000Research* (Vol. 7). F1000 Research Ltd. https://doi.org/10.12688/f1000research.14506.1
58. Wimo, A., Jönsson, L., Bond, J., Prince, M., & Winblad, B. (2013). The worldwide economic impact of dementia 2010. *Alzheimer's and Dementia*, *9*(1), 1-11.e3. https://doi.org/10.1016/j.jalz.2012.11.006
59. Xie, X.-Y., Zhao, Q.-H., Huang, Q., Dammer, E., Chen, S., Ren, R.-J., & Wang, G. (2021). Genetic profiles of familial late-onset Alzheimer's Disease in China: The Shanghai FLOAD study. *Genes & Diseases*. https://doi.org/10.1016/j.gendis.2021.05.001
60. Xu, W., Wu, C., Fletcher, J., & Fletcher, J. (2020). Assessment of changes in place of death of older adults who died from dementia in the United States, 2000-2014: A time-series cross-sectional analysis. *BMC Public Health*, *20*(1), 765. https://doi.org/10.1186/s12889-020-08894-0
61. Yang, W., Woltjer, R. L., Sokal, I., Pan, C., Wang, Y., Brodey, M., Peskind, E. R., Leverenz, J. B., Zhang, J., Perl, D. P., Galasko, D. R., & Montine, T. J. (2007). Quantitative proteomics identifies surfactant-resistant α-synuclein in cerebral cortex of Parkinsonism-dementia complex of Guam but not Alzheimer's disease or progressive supranuclear palsy. *American Journal of Pathology*, *171*(3), 993–1002. https://doi.org/10.2353/ajpath.2007.070015
62. Yang, Z., & Levey, A. (2015). Gender Differences: A Lifetime Analysis of the Economic Burden of Alzheimer's Disease. *Women's Health Issues*, *25*(5), 436–440. https://doi.org/10.1016/j.whi.2015.06.001
63. Yu, F., Guo, J., Zhu, X., & Shi, G. (2015). Real time prediction of unoccupied parking space using time series model. *ICTIS 2015 - 3rd International Conference on Transportation Information and Safety, Proceedings*, 370–374.

https://doi.org/10.1109/ICTIS.2015.7232145