Impact of oral anticoagulant therapy on the occurrence of atrial fibrillation-related stroke in Portugal

Madalena Ribeiro Gil
madalena.gil@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

December 2020

Abstract

Atrial fibrillation is the most prevalent arrhythmia in clinical practice. Usually, the diagnosis of this pathology occurs following a complication. The main complications of atrial fibrillation are thromboembolic events, in particular, the stroke. In order to prevent the occurrence of such events, patients suffering from this arrhythmia are advised to follow an anticoagulant drug therapy. In Portugal, the increased uptake of oral anticoagulants and its impact on the occurrence of stroke in patients with atrial fibrillation remains uncertain, so the need to study this association is urgent. An aged Portuguese society leads to an increase in the prevalence of cardiovascular diseases, in particular, of atrial fibrillation and, consequently, of the thromboembolic events's risk. This paper aims to build a model that estimates the impact of oral anticoagulant therapy on the occurrence of stroke in patients with atrial fibrillation, in continental Portugal, between 2012 and 2018. The created model proves that the uptake of oral anticoagulation is associated with a decline in the number of stroke episodes in patients with atrial fibrillation. The developed model comprises a linear regression with correlated errors that describes the number of stroke episodes in patients with atrial fibrillation depending on the national uptake of oral anticoagulants and its clinical and demographic indicators.

Keywords: Atrial Fibrillation, Stroke, Oral Anticoagulation, Time Series, Regression Models

1. Introduction

Atrial fibrillation (AF) is the most prevalent arrhythmia in clinical practice. In 2010, the number of individuals with AF was estimated at nine million in the European Union and 33.5 million worldwide (Gouveia et al., 2015). This type of cardiac arrhythmia is directly associated with age, so it is to be expected that with the aging of the population, the number of individuals with AF may increase exponentially in the coming decades (Bonhorst, 2018). Estimates for the European Union predict this figure will double to 18 million by 2060 (Gouveia et al., 2015).

Often, the diagnosis of atrial fibrillation occurs following a complication. The main complication of AF is the systemic thromboembolism, in particular, the stroke. The risk of developing an ischemic stroke is about three to five times higher in patients with AF compared to patients without this arrhythmia (Gouveia et al., 2015). In Portugal, it is estimated that 14% of patients with atrial fibrillation have already developed a stroke (Gouveia et al., 2015). Thus, AF represents a significant cause of morbidity and mortality, mainly due to the associated risk of developing an ischaemic stroke.

Oral anticoagulation is an effective therapy in the prevention of stroke associated with atrial fibrillation. There is strong evidence that oral anticoagulation effectively reduces the risk of thromboembolic events.

Although there are some studies on the hospitalized atrial fibrillation-related stroke and its association with uptake of oral anticoagulants, the conclusions described in these publications vary. Data from the U.S. Medicare health insurance system showed that, between 1992 and 2007, there was a decrease in stroke rates in AF patients, which coincided with the doubling of oral anticoagulant uptake (Shroff et al., 2014). Conversely, another study conducted in the USA concluded that the incidence rates of stroke or transient ischaemic attack (TIA) in patients with atrial fibrillation remained unchanged between 2000 and 2010 (Chamberlain et al., 2016). This occurrence was attributed to the stabilization of the uptake rates of oral anticoagulants that occurred during the study period (Chamberlain et al., 2016). In turn, Suzuki et al. (2016) showed that, although the uptake of oral anticoagulants increased in Japan between 2004 and 2013,

there was a progressive increase in the incidence of thromboembolic events in AF patients.

Two types of oral anticoagulants are marketed in Portugal: vitamin K antagonists (VKA) and non-vitamin K antagonists (NOAC). From 2010 to the first quarter of 2016, the number of packages prescribed for NOAC increased. On the other hand, it was found that the number of prescribed packages of VKA, which in 2014 reached their peak prescription, are currently decreasing (Caldeira et al., 2017). However, despite this change in the prescription pattern, the national uptake of oral anticoagulants has increased substantially. This increase follows a significant increase in the use of NOAC (Caldeira et al., 2017).

In Portugal, the effect of increased uptake of oral anticoagulants and its impact on the occurrence of stroke in patients with AF remains uncertain, so the need to study this association is urgent. An aged Portuguese society leads to an increase in the prevalence of cardiovascular diseases, in particular, of atrial fibrillation and, consequently, of the thromboembolic events's risk. This paper aims to build a model that estimates the impact of oral anticoagulant therapy on the occurrence of stroke in patients with atrial fibrillation, in continental Portugal, between 2012 and 2018.

1.1. State of the Art

The present paper is motivated by Cowan et al. (2018), whose goal is to investigate whether changing patterns of anticoagulant use in atrial fibrillation have impacted on stroke rates in England. In this article, a series of Poisson regression models were fitted to model the rate of hospitalized AF-related stroke and its association with the use of oral anticoagulants or anti-platelet drugs between 2006 and 2016. The authors included, in addiction to the use of oral anticoagulants among patients with AF and CHA_2DS_2-VAS_c score ≥ 2, sine curves to model seasonality, AF prevalence and patients demographics including sex, age in years and Charlson Comorbidity Index categories I to III (Cowan et al., 2018).

Similarly, Yiin et al. (2017) also fitted Poisson regression models to calculate the relative incidence of AF-related ischaemic events that occurred in patients with AF from 2002 to 2012. The authors used demographic indicators (gender and age) of patients, $CHADS_2$ score and $CHA_2DS_2VAS_c$ score for risk of embolic ischaemic events, number of anticoagulated patients with known prior AF and number of non-anticoagulated patients with known prior AF.

Nevertheless, other authors chose to follow a different strategy. Shroff et al. (2014) aimed to evaluate ischemic stroke rates in demographic subsets (age, sex and race) of U.S. Medicare beneficiaries and to assess warfarin use for 1992-2010. Additionally, they sought to examine ischemic stroke risk among warfarin-treated and non-treated Medicare beneficiaries with AF to better understand the contribution of warfarin to ischemic stroke reduction.

Most of these articles arise from research groups in the areas of health and medicine. Therefore, more emphasis is placed on the epidemiological relevance of the study. Hence, the statistical methods and procedures used to obtain the estimates are rarely described.

1.2. Contributions

The goal of the study presented in this paper is to build a model that estimates the impact of oral anticoagulation on the occurrence of stroke in AF patients, in continental Portugal, between 2012 and 2018. In order to obtain satisfactory results, firstly, time series analysis was performed to describe the data. Then, time series methods were used to fit several models to the data.

In this study, **linear regression models with correlated errors** were built. The final model describes the monthly number of stroke episodes in AF patients depending on the national uptake of oral anticoagulants and its clinical and demographic indicators. The chosen model confirms that the use of oral anticoagulants is associated with a decline in the number of stroke episodes in AF patients.

2. Sample and Methods

2.1. Sample

To study the impact of oral anticoagulant therapy on the occurrence of stroke in AF patients in Portugal, a data set was provided by the Centre for Evidence Based Medicine (CEMBE), a structural unit of Lisbon School of Medicine. Two response variables were measured. The first one refers to monthly observations of the number of stroke episodes (ischaemic or hemorrhagic) in AF patients in Portuguese public hospitals between January 2012 and December 2018. The second one refers to the rate of stroke episodes (ischaemic or hemorrhagic) in AF patients. This rate corresponds to the ratio between the number of stroke episodes in AF patients and the number of AF patients. It is expressed in permille (‰). The data corresponds to adult patients (aged 18 years or older). The two reveal dependencies overtime and because of that are model as time series, with monthly dependencies.

Since we are dealing with linear regression mod-

els with correlated errors, the following time series data were used as **explanatory variables**: (i) number of patients treated with oral anticoagulants, (ii) number of AF patients, (iii) number of AF female patients, (iv) age of AF patients, (v) severity and (vi) Charlson Comorbidity Index of stroke episodes in patients with AF. The frequency of the number of AF patients, the number of AF female patients and the age of the AF patients is annual, while the frequency of the remaining variables is monthly. The annual data is between 2012 and 2018, while monthly data starts in January 2012 and ends in December 2018. Given the different frequencies of the variables used, monthly and annual, it was decided to repeat for all months of a given year, the annual estimates for that same year. Thus, several models of monthly frequency were obtained.

The data provided by CEMBE was extracted from two different databases: National Database on Hospital Morbidity (BDMH) and national statistics on the number of sold packages of oral anticoagulants. The BDMH includes all inpatient hospital episodes and outpatient clinics grouped into Homogeneous Diagnostic Groups (GDH) (ACSS, 2018). The monthly coded information grouped into GDH is subsequently forwarded to the BDMH located at the Central Administration of the Health System (ACSS). The BDMH includes all hospital and inpatient episodes coded into ICD-9-CM and ICD-10-CM/PCS from Portuguese public hospitals (ACSS, 2018).

On the other hand, the data on oral anticoagulants refers to national statistics on the number of sold packages, in each monthly period, of VKA and NOAC, in continental Portugal, between January 2012 and December 2018. In order to carry out this study, it was necessary to transform the number of sold packages of oral anticoagulants into an estimate of the number of patients undergoing treatment. For this purpose, the defined daily dose (DDD) was used. The DDD is the assumed average maintenance dose per day for a drug used for its main indication in adults. Thus, to estimate the number of patients receiving treatment each month, we started by calculating the number of milligrams sold, in each monthly period, for each of the six oral anticoagulants (acenocoumarol, warfarin, apixaban, dabigatran, edoxaban and rivaroxaban). DDD was then used to determine the number of days of treatment and, consequently, the number of months of treatment insured with the total milligrams sold per month, which is equivalent to calculating the number of patients undergoing treatment in one month. The number of months of treatment corresponds to the ratio between the number of days of treatment and the number of days in a month.

2.2. Methodology

A **time series** can be defined as a collection of random variables indexed according to the order they are obtained in time. For example, we may consider a time series as a sequence of random variables $X_1, X_2, ...$, where the random variable X_1 denotes the value taken by the series at the first time period, the variable X_2 denotes the value for the second time period and so on. In general, a collection of random variables $\{X_t\}$ indexed by t is referred as a stochastic process. In this paper, t will be discrete, i.e., the observations are made at fixed time intervals.

Classical regression models are often incapable of incorporating all of the interesting dynamics of a time series. In this domain, it is desirable to allow the response variable to be explained by the present and past values of the covariates and possibly by its own past values. There are two widely used time series models: ARMA and ARIMA models.

The integrated ARMA, or ARIMA model, is a broadening of the class of ARMA models to include differencing. When data are stationary, we fit ARMA models. However, when data show evidence of non-stationarity, we fit ARIMA models, where an initial differencing step can be applied to reduce the non-stationarity. The AR (autoregressive) component of the models is based on the idea that the current value of the series, X_t, can be explained as a function of p past values $X_{t-1}, X_{t-2}, ..., X_{t-p}$, where p determines the number of steps into the past needed to forecast the current value. On the other hand, the MA (moving average) component of the models assumes the white noise, ϵ_t, are combined linearly to form the observed data. A time plot of the data will typically suggest whether any differencing is needed. We need to be careful not to overdifference because this may introduce dependence where none exists. Generally, non-seasonal ARIMA models are denoted ARIMA(p, d, q) where parameters p, d, and q are non-negative integers, p is the non-seasonal AR order, d is the degree of non-seasonal differencing (usually $d \leq 2$) and q is the non-seasonal MA order.

In order to account for seasonal and non-stationary behavior, several modifications were made to the ARIMA model. Usually, the dependence on the past tends to occur most strongly at multiples of some underlying seasonal model by the function, $lag(s)$. Seasonal non-stationarity can occur, for example, when the process is nearly periodic in the season. Seasonal ARIMA models are commonly denoted SARIMA$(p, d, q) \times (P, D, Q)_s$ where parameters s, P, D, and Q are non-negative

integers, s is the number of periods in each season, P is the seasonal AR order, D is the degree of seasonal differencing (usually $D \leq 1$), and Q is the seasonal MA order.

Classical regression models allow for the inclusion of a lot of relevant information from predictor variables, but do not allow for the subtle time series dynamics that can be handled with SARIMA models. So, SARIMA models were extended in order to allow other information to be included in the models. The **linear regression model with SARIMA errors** performs time series regression by assuming that the errors may have a time series structure. Let $\{Y_t : t \in \mathbb{Z}\}$ be the time series response variable, $\{X_{1,t}, ..., X_{k,t} : t \in \mathbb{Z}\}$ be the set of time series predictors and $\{N_t : t \in \mathbb{Z}\}$ be the errors of the regression. In this case, $\{N_t : t \in \mathbb{Z}\} \sim$ SARIMA$(p, d, q) \times (P, D, Q)_s$. The linear regression model with SARIMA errors can be written as follows,

$$Y_t = \beta_0 + \beta_1 X_{1,t} + ... + \beta_k X_{k,t} + N_t, \quad t \in \mathbb{Z}, \quad (1)$$

with

$$\Phi_P(B^s)\phi(B)\nabla_s^D \nabla^d N_t = \Theta_Q(B^s)\theta(B)\epsilon_t, \quad t \in \mathbb{Z}. \quad (2)$$

The ordinary autoregressive and moving average components are represented by polynomials $\phi(B)$ and $\theta(B)$ of orders p and q, respectively, and the seasonal autoregressive and moving average components by $\Phi_P(B^s)$ and $\Theta_Q(B^s)$ of orders P and Q and ordinary and seasonal difference components by $\nabla^d = (1 - B)^d$ and $\nabla_s^D = (1 - B^s)^D$ (Shumway and Stoffer, 2006). The model has two error terms: the error from the regression model, $\{N_t : t \in \mathbb{Z}\}$, and the error from the SARIMA model, $\{\epsilon_t : t \in \mathbb{Z}\}$. Only the SARIMA model errors are assumed to be white noise. The white noise process is a random process of random variables that are uncorrelated, have mean zero, and finite variance.

After the estimation of the parameters of the regression model, the next step in model fitting is diagnostic. This investigation includes the analysis of the residuals of the model. If the model fits well, the residuals should be white noise. It was performed a statistical test, the Ljung-Box test, on the residuals to test this hypothesis (Ljung and Box, 1978).

When we have estimated a set of models that fit well the data, we need to take into account some performance metrics, in order to select the model that best explains the series under study. The performance metrics considered were the corrected Akaike information criterion (AICc) (Cavanaugh, 1997), the mean squared error (MSE), the mean absolute percentage error (MAPE) and the adjusted coefficient of determination (R_a^2). The measures that evaluate the quality of the forecasts are given as follows:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2, \quad (3)$$

where n is the number of observations, y_t is the observed value at time t, and $\hat{y}_t$ is the forecasted value also at time t.

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \times 100 \quad (\%), \quad (4)$$

where n is the number of observations, y_t is the observed value at time t, and $\hat{y}_t$ is the forecasted value also at time t.

The coefficient of determination, denoted R^2, is a statistical measure that represents the proportion of the variance in the dependent variable that is explained by the estimated regression line. Usually, $0 \leq R^2 \leq 1$. The closer R^2 is to 1, the better the estimated regression equation fits the data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \quad (5)$$

The adjusted coefficient of determination, denoted R_a^2, is a modification of R^2 that adjusts for the number of explanatory variables in the model.

$$R_a^2 = 1 - \frac{(n-1)}{(n-k)}\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}, \quad (6)$$

where k is the number of explanatory variables in the regression model.

In order to select the model that best fits the data, we chose the one that minimizes the value of the AICc, MSE, and MAPE and that maximizes the value of R_a^2.

The statistical analyses were performed in R (R Development Core Team, 2020).

3. Descriptive Data Analysis

3.1. Atrial Fibrillation-Related Stroke

The time plots of each of the response variables are shown in Figure 1 and Figure 2, respectively.

The plotted data shows a clearly seasonality pattern. According to some authors, the presence of these seasonal patterns may be due to environmental phenomena, in particular: temperature, humidity, and hours of sunshine (Christensen et al., 2012).

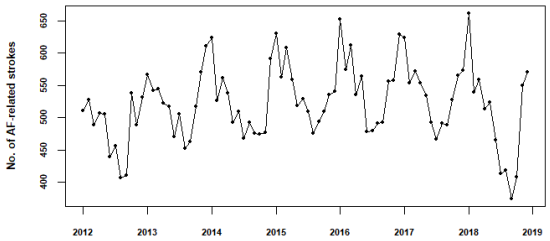

Figure 1: Number of stroke episodes in AF patients in Portuguese public hospitals between January 2012 and December 2018.

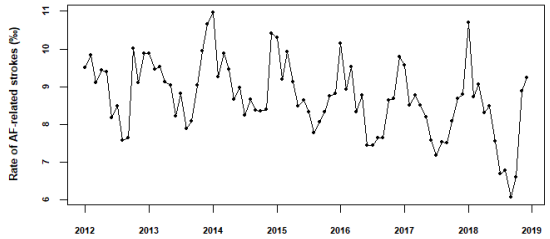

Figure 2: Rate of stroke episodes in AF patients in Portuguese public hospitals between January 2012 and December 2018.

The trend plot in Figure 3 shows that, between January 2012 and November 2013, there was an increase in the number of stroke episodes in patients with atrial fibrillation, followed by a decreasing period until July 2014. Nevertheless, this trend is reversed between July 2014 and July 2015. Again, there is a decrease in the number of stroke episodes in AF patients between July 2015 and January 2016. Between January 2016 and July 2016 there was a slight increase in the number of stroke episodes in patients with AF. However, from July 2016 to December 2018, the number of episodes of AF-related stroke decreases.

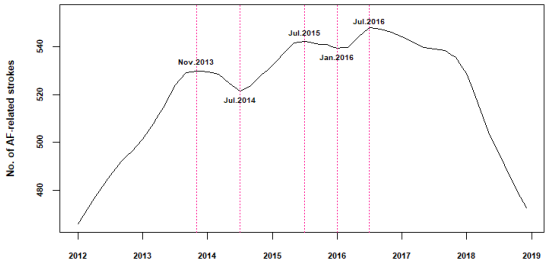

Figure 3: Trend in the number of stroke episodes in AF patients in Portuguese public hospitals between January 2012 and December 2018.

3.2. Oral Anticoagulation

The time plot of the number of patients treated with oral anticoagulants (VKA or NOAC) is presented in Figure 4. These data corresponds to an estimate of the number of patients treated with oral anticoagulants. As mentioned in Subsection 2.1, this estimate was calculated from the number of sold packages of oral anticoagulants, in continental Portugal, between January 2012 and December 2018.

The plotted data (see Figure 4) shows a gradual increase in the number of anticoagulated individuals from January 2012 to December 2018. This increase is due to the fact that there has been a significant increase in the use of NOAC (see Figure 5). Caldeira et al. (2017) also noted this.

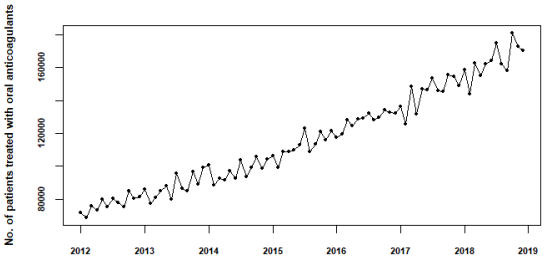

Figure 4: Number of patients treated with oral anticoagulants in continental Portugal between January 2012 and December 2018.

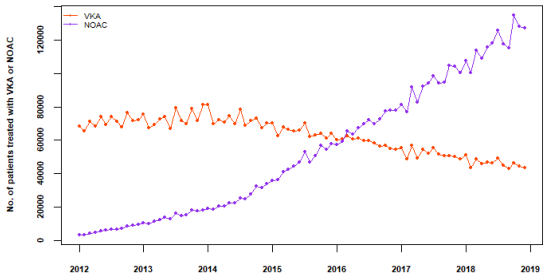

Figure 5: Number of patients treated with VKA (orange) or NOAC (purple) in continental Portugal between January 2012 and December 2018.

From 2012 to 2018, the use of VKA decreased, as opposed to the use of NOAC, whose number of anticoagulated patients increased significantly. Since 2014, anticoagulants composed by apixaban, dabigatran and rivaroxaban have begun to be reimbursed at 69% of their market price by the National Health System (NHS) for stroke prevention in adults with atrial fibrillation.

3.3. Atrial Fibrillation

The time plot of the number of AF patients registered in Portuguese public hospitals from 2012 to 2018 is shown in Figure 6. The data corresponds to patients with atrial fibrillation aged 18 years and older.

The proportion of patients with atrial fibrillation was calculated. The time plot is presented in Figure 7. This measure corresponds to the ratio between the number of AF patients and the total number of patients registered in the BDMH, in each year. This proportion represents a satisfactory proxy for atrial fibrillation prevalence in Portugal.

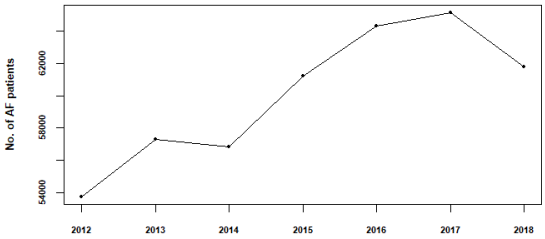

Figure 6: Number of AF patients in Portuguese public hospitals between 2012 and 2018.

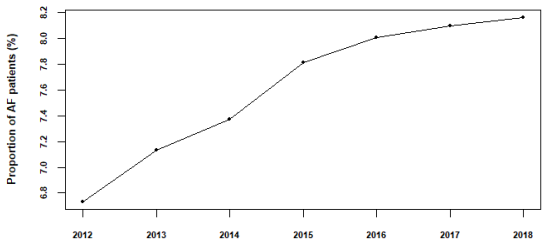

Figure 7: Proportion of AF patients in Portuguese public hospitals between 2012 and 2018.

The time plot in Figure 7 exhibits an increasing trend. It is known that, in the last 20 years, hospital admissions for AF increased by 66% due to the aging of the population, the increase in the prevalence of chronic heart diseases and the greater use of ambulatory electrocardiographic monitoring (Coordenação Nacional para as Doenças Cardiovasculares, 2009). Thus, it is also expected that the proportion of patients with a record of AF (for all patients registered in the BDMH) has, for the same reasons, increased in the period under analysis.

3.4. Additional Explanatory Variables

In order to create an appropriate model to describe the series under study, other time series data were added to the linear regression models with SARIMA errors. Besides the number of patients treated with oral anticoagulants and the number of AF patients, other time series like, the number of AF female patients, the age of AF patients, severity and Charlson Comorbidity Index of stroke episodes in AF patients were also added to the time series regression models. In addiction, a categorical variable, denoted $(periodo_{2016-2018})_t$, was also added. This variable takes value equal to 1 for observations of 2016, 2017, or 2018. Otherwise, it takes 0. The reason for the inclusion of this variable in the regression models is supported by Figures 3 and 5. Figure 3 shows a downward trend in the number of stroke episodes in patients with atrial fibrillation since 2016. In its turn, Figure 5 shows that at the beginning of 2016, the use of NOAC equals the use of VKA, however, the number of anticoagulated patients with NOAC eventually exceeds the number of anticoagulated patients with AVK, resulting in an increase in the total number of patients treated with oral anticoagulants. Once the number of anticoagulated patients between January 2016 and December 2018 is somewhat higher than the number of anticoagulated patients between January 2012 and December 2015 (see Figure 4) and, the trend in the number of stroke episodes is essentially decreasing since 2016, it is important to understand the impact on the number of episodes of AF-related stroke in the transition from one period to the other. Thus, in order to analyze this effect, it was decided to add this categorical variable to the regression models in two different ways: to add this variable as a simple explanatory variable or to add an interaction term between this variable and the logarithmic series of the number of patients treated with oral anticoagulants, i.e., $log\,(doentes_{aco})_t \times (periodo_{2016-2018})_t$.

4. Modeling

Firstly, it was necessary to standardize the series involved in the regression models. Thus, **logarithmic transformation** was applied to each time series of variables.

The *auto.arima()* function from the *forecast* package (Hyndman and Khandakar, 2008) from R (R Development Core Team, 2020) was used to automatically adjust the best linear regression model with SARIMA errors to the data.

In order to assess the quality of the forecasts of the potential models, a training period was established between January 2012 and July 2018 and a test period between August 2018 and December 2018. Thus, for each potential model, the forecasts for the test period will be presented, as well as, the prediction errors for the training and test sets. However, the estimation of the final model is made with all the data.

4.1. Number of episodes of AF-related stroke

Several models were built in order to describe the logarithmic series of the number of stroke episodes in AF patients in Portuguese public hospitals between January 2012 and December 2018.

The model given by equation (7) was the best model achieved that includes the variable $(periodo_{2016-2018})_t$ as a simple explanatory variable. The fitted model is given by,

$$\begin{aligned}\hat{E}_{avc,t} = &-0.0419\,X_{1,t} - 0.1438\,X_{2,t} + \\ &+1.2172\,X_{3,t} + \hat{N}_{t-12} + \hat{\epsilon}_t - 0.5151\,\hat{\epsilon}_{t-12}.\end{aligned} \tag{7}$$

where $E_{avc,t} = log\,(epi_{avc})_t$, i.e., the logarithmic series of the number of stroke episodes in AF patients, $X_{1,t} = (periodo_{2016-2018})_t$, $X_{2,t} = log\,(doentes_{aco})_t$, i.e., the logarithmic series of the number of patients treated with oral anticoagulants,

and $X_{3,t} = log\,(doentes_{fa})_t$, i.e., the logarithmic series of the number of AF patients.

Moreover, it was fitted a SARIMA$(0,0,0) \times (0,1,1)_{12}$ model to the regression model errors.

For the given model, AICc=-174.79, MSE=0.004, MAPE=0.701%, and $R_a^2 = 0.703$.

The forecasts of the logarithmic series of the number of stroke episodes in AF patients made by the model for the last 5 months are shown in Figure 8. The MSE and the MAPE for the training set are 0.003 and 0.609%, respectively. For the test period, the MSE and the MAPE of the forecasts are 0.030 and 2.257%, respectively, which confirm the quality of the predictions. Despite this, the first three months of the test period (August 2018, September 2018, and October 2018) are outside the forecast range. Such behavior may be due to the sharp drop in the number of events in these months.

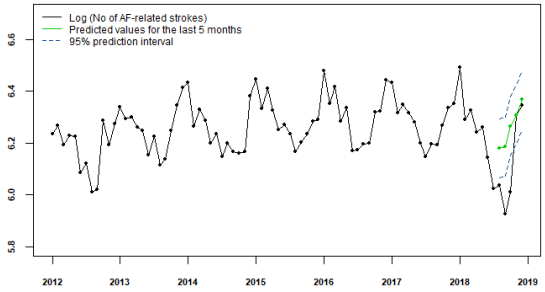

Figure 8: Forecasts for the last 5 months (green) made by model (7) with 95% prediction interval (dashed blue line).

The next presented model also describes the logarithmic series of the number of stroke episodes in AF patients. However, unlike the previous model, it includes the interaction term between the categorical variable and the logarithmic series of the number of patients treated with oral anticoagulants, i.e., $log\,(doentes_{aco})_t \times (periodo_{2016-2018})_t$. The fitted equation of the model with best performance is given by,

$$\begin{aligned}\hat{E}_{avc,t} = &-0.1376\,X_{1,t} - 0.0036\,X_{2,t} + \\ &+1.2137\,X_{3,t} + \hat{N}_{t-12} + \hat{\epsilon}_t - 0.5169\,\hat{\epsilon}_{t-12}.\end{aligned} \tag{8}$$

where $E_{avc,t} = log\,(epi_{avc})_t$, i.e., the logarithmic series of the number of stroke episodes in AF patients, $X_{1,t} = log\,(doentes_{aco})_t$, i.e., the logarithmic series of the number of patients treated with oral anticoagulants, $X_{2,t} = log\,(doentes_{aco})_t \times (periodo_{2016-2018})_t$, and $X_{3,t} = log\,(doentes_{fa})_t$, i.e., the logarithmic series of the number of AF patients.

Again, it was fitted a SARIMA$(0,0,0) \times (0,1,1)_{12}$ model to the regression model errors.

For the given model, AICc=-174.87, MSE=0.004, MAPE=0.700%, and $R_a^2 = 0.703$.

The forecasts of the logarithmic series of the number of stroke episodes in AF patients made by the model for the last 5 months are shown in Figure 9. The MSE and the MAPE for the training set are 0.003 and 0.609%, respectively. For the test period, the MSE and the MAPE of the forecasts are 0.031 and 2.268%, respectively, which confirm the quality of the predictions. Once again, the first three months of the test period (August 2018, September 2018, and October 2018) are outside the forecast range.

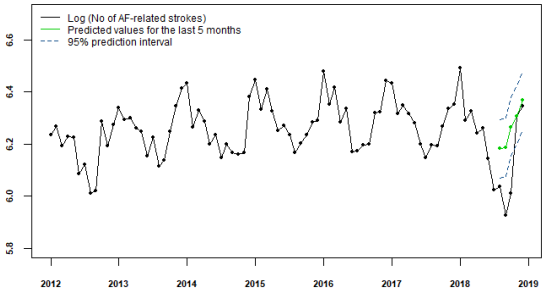

Figure 9: Forecasts for the last 5 months (green) made by model (8) with 95% prediction interval (dashed blue line).

4.2. Rate of episodes of AF-related stroke

In order to describe the logarithmic series of the rate of stroke episodes in AF patients in Portuguese public hospitals between January 2012 and December 2018, some models were built.

The following model presented was the best model achieved that includes the variable $(periodo_{2016-2018})_t$ as a simple explanatory variable. The fitted model equation is given by,

$$\begin{aligned}\hat{T}_{avc,t} = &-0.0331\,X_{1,t} - 0.1104\,X_{2,t} + \\ &+\hat{N}_{t-12} + \hat{\epsilon}_t - 0.4872\,\hat{\epsilon}_{t-12}.\end{aligned} \tag{9}$$

where $T_{avc,t} = log\,(taxa_{avc})_t$, i.e., the logarithmic series of the rate of stroke episodes in AF patients, $X_{1,t} = (periodo_{2016-2018})_t$, and $X_{2,t} = log\,(doentes_{aco})_t$, i.e., the logarithmic series of the number of patients treated with oral anticoagulants.

Furthermore, it was fitted a SARIMA$(0,0,0) \times (0,1,1)_{12}$ model to the regression model errors.

For the given model, AICc=175.87, MSE=0.005, MAPE=2.029%, and $R_a^2 = 0.708$.

The forecasts of the logarithmic series of the rate of stroke episodes in AF patients made by the model for the last 5 months are shown in Figure 10. The MSE and the MAPE for the training set are 0.003 and 1.763%, respectively. For the test period, the MSE and the MAPE of the forecasts are 0.030 and 7.236%, respectively. Despite this, the first three months of the test period (August 2018, September 2018, and October 2018) are outside the forecast range.

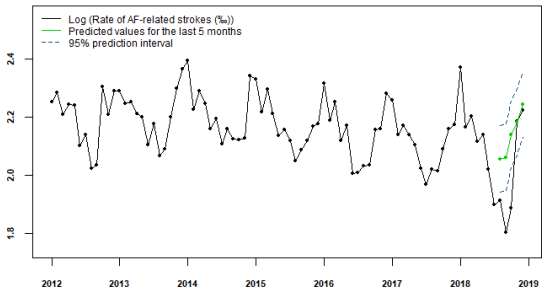

Figure 10: Forecasts for the last 5 months (green) made by model (9) with 95% prediction interval (dashed blue line).

The next presented model also describes the logarithmic series of the rate of stroke episodes in AF patients. However, unlike the previous model, it includes the interaction term between the categorical variable and the logarithmic series of the number of patients treated with oral anticoagulants, i.e., $log\,(doentes_{aco})_t \times (periodo_{2016-2018})_t$. The fitted equation of the model with best performance is given by,

$$\begin{aligned}\hat{T}_{avc,t} = -0.1051\,X_{1,t} - 0.0029\,X_{2,t} + \\ +\hat{N}_{t-12} + \hat{\epsilon}_t - 0.4905\,\hat{\epsilon}_{t-12}.\end{aligned} \tag{10}$$

where $T_{avc,t} = log\,(taxa_{avc})_t$, i.e., the logarithmic series of the rate of stroke episodes in AF patients, $\{X_{1,t}\} = log\,(doentes_{aco})_t$, i.e., the logarithmic series of the number of patients treated with oral anticoagulants, and $\{X_{2,t}\} = log\,(doentes_{aco})_t \times (periodo_{2016-2018})_t$.

Again, it was fitted a SARIMA$(0,0,0) \times (0,1,1)_{12}$ model to the regression model errors.

For the given model, AICc=-175.97, MSE=0.004, MAPE=2.027%, and $R_a^2 = 0.709$.

The forecasts of the logarithmic series of the rate of stroke episodes in AF patients made by the model for the last 5 months are shown in Figure 11. The MSE and the MAPE for the training set are 0.003 and 1.761%, respectively. For the test period, the MSE and the MAPE of the forecasts are 0.031 and 7.273%, respectively.

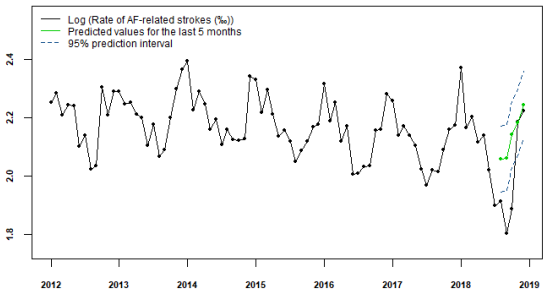

Figure 11: Forecasts for the last 5 months (green) made by model (10) with 95% prediction interval (dashed blue line).

Again, the first three months of the test period (August 2018, September 2018, and October 2018) are outside the forecast range. As stated, such behavior may be due to the sharp drop in the number of events in these months.

4.3. Final Model

In clinical terms, the models that use as a response variable, the number of stroke episodes in AF patients, are more significant for this study, once they incorporate, besides the number of patients treated with oral anticoagulants, the number of patients with atrial fibrillation as an explanatory variable. From the clinician's perspective, it is important not only to understand the impact of the effect of oral anticoagulation on the occurrence of stroke in adults with atrial fibrillation, but also to understand the effect of this cardiac arrhythmia on the occurrence of stroke.

According to the clinician's opinion, the model given by **equation** (**8**) presents a clearer and more objective interpretation of the problem, so it was defined as the **final model** to describe the impact of oral anticoagulant therapy on the occurrence of episodes of AF-related stroke, in continental Portugal, from 2012 to 2018. The model was estimated with all the data.

Considering the signs of the estimates of the regression coefficients of each of the explanatory series, it is observed that the variables $log\,(doentes_{aco})_t$ and the interaction term $log\,(doentes_{aco})_t \times (periodo_{2016-2018})_t$ have a negative weight on the logarithmic series of the number of stroke episodes in AF patients, whereas the variable $log\,(doentes_{fa})_t$ exerts a positive weight. Therefore, the chosen model can be interpreted as follows: if the number of patients treated with oral anticoagulants increases 1% between January 2012 and December 2015, i.e., when $periodo_{2016-2018} = 0$, then the number of stroke episodes that occur in patients with atrial fibrillation is expected to drop 0.1376%. On the other hand, if the number of patients treated with oral anticoagulants increases 1% between January 2016 and December 2018, i.e. when $periodo_{2016-2018} = 1$, then the number of stroke episodes that occur in patients with atrial fibrillation is expected to drop 0.1412%. However, if the number of patients with this cardiac arrhythmia increases 1%, then the number of stroke episodes is expected to grow 1.2137%.

Thus, the **adjusted final model** (**8**) indicates that the increase in the nationwide uptake of oral anticoagulants between 2012 and 2018, in continental Portugal, is associated with a decrease in the number of stroke episodes in patients with atrial fibrillation. In fact, it should be noted that the increase in NOAC uptake, from 2016, was substantial and contributed to a significant reduction in the

number of stroke events in AF patients. Thus, a decrease of 4.2% of stroke episodes in patients with AF between 2016 and 2018 is expected, due to increased uptake of NOAC in this period. On the other hand, the referred cardiovascular disease is associated with an increase in the number of stroke episodes, as expected.

Nevertheless, it should be noted that the four best models identified and analyzed in Subsections 4.1 and 4.2 point in the same direction, i.e., they indicate that oral anticoagulant therapy is associated with a decline in the number/rate of stroke episodes in patients with AF, in Portugal.

Figure 12 shows the logarithmic series of the number of stroke episodes in AF patients in Portuguese public hospitals between January 2012 and December 2018 (black) and the selected model to fit the data (magenta).

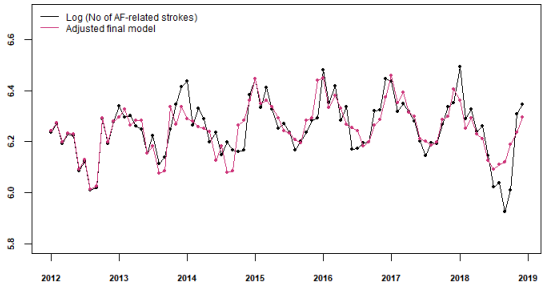

Figure 12: Log series of the number of stroke episodes in AF patients in Portuguese public hospitals between January 2012 and December 2018 (black) and adjusted final model (magenta).

5. Conclusions

In this analysis, it was found that an increase in the nationwide uptake of oral anticoagulants between 2012 and 2018, in Portugal, was remarkably associated with a decline in the number of episodes of AF-related stroke.

The model given by equation (8) was defined as the final model to estimate the impact of oral anticoagulation on the occurrence of AF-related stroke, in continental Portugal, between 2012 and 2018. The model describes the number of episodes of AF-related stroke using the following time series as explanatory variables: number of patients treated with oral anticoagulants, interaction term between the previous variable and a categorical one (see Subsection 3.4), and number of patients with AF. Logarithmic transformation was applied to each time series of variables in order to estimate the model.

The development of this model aims to illustrate the Portuguese panorama on how the use of oral anticoagulants is related to the occurrence of stroke episodes amongst AF patients, thus allowing to support health decisions in the cardiovascular field. The existence of this model also awakens to the constant need to identify and manage the therapy used in the control of atrial fibrillation, in order to prevent complications associated with this arrhythmia.

5.1. Achievements

The chosen model is totally satisfactory, since the values achieved by AICc, MAPE, and R_a^2 reveal a model well adjusted to the data. The adjusted final model shows that an increase in the nationwide uptake of oral anticoagulants from 2012 to 2018, in continental Portugal, is associated with a decrease in the number of stroke episodes in adults with AF. Thus, the results obtained by the model corroborate the results achieved in clinical trials, i.e., using real data it was possible to demonstrate that oral anticoagulation is an effective therapy in the prevention of AF-related stroke.

5.2. Limitations of the Study

Despite the success in obtaining a model with a good fit and easy interpretation capable of describing the effect of oral anticoagulant therapy on the occurrence of AF-related stroke, the limitations of this study are acknowledged. The data was aggregated and it was not possible to track individual patients. This aggregation could lead to a possible 'ecological fallacy' in which the findings are attributed incorrectly to the individual patient level, i.e., one cannot presume that the findings are applicable to individual patients.

Another limitation to consider in this study was the use of two different data sources, namely, the BDMH and a data source with the national statistics on the number of sold packages of oral anticoagulants, since it was not possible to obtain the data from a single information system.

The use of an estimate for the number of patients treated with oral anticoagulants also represents a limitation of this research. As mentioned in Subsection 2.1, this estimate was calculated from the national statistics on the number of sold packages of oral anticoagulants. Nevertheless, it is assumed that the vast majority of these patients are undergoing oral hypocoagulation due to AF, since it is considered that the main therapeutic indication for the use of oral anticoagulants is for the prevention of AF-related stroke.

5.3. Future Work

With the strengthening and improvement of the health information systems, it would be interesting to replicate this study, since with more advanced health information systems, the data would be extracted directly from each patient's clinical record. In this way, it would no longer be necessary to use

aggregated data.

Thus, the estimates of the impact of oral anticoagulation on the occurrence of stroke in adults with atrial fibrillation would be more accurate, allowing to support and implement, with greater safety, health decisions in the cardiovascular scope.

References

ACSS. ACSS apresenta o novo ecossistema sobre Morbilidade Hospitalar. URL: http://www.acss.min-saude.pt/2018/06/04/acss-apresenta-o-novo-ecossistema-sobre-morbilidade-hospitalar/, 2018. (accessed on 04/2020).

D. Bonhorst. Um novo olhar sobre a prevalência da fibrilhação auricular em Portugal - O Estudo Safira. *Revista Portuguesa de Cardiologia*, 37 (4):315–317, 2018.

D. Caldeira, M. Barra, C. David, J. Costa, J. J. Ferreira, and F. J. Pinto. Prevalência da anticoagulação oral em doentes com fibrilhação auricular em Portugal: revisão sistemática e meta-análise de estudos observacionais. *Revista Portuguesa de Cardiologia*, 33(9):555–560, 2014.

D. Caldeira, J. J. Ferreira, and F. J. Pinto. A era dos novos anticoagulantes orais em Portugal. *Revista Portuguesa de Cardiologia*, 36(7-8):577–578, 2017.

J. Cavanaugh. Unifying the derivations of the Akaike and corrected Akaike information criteria. *Statistics and Probability Letters*, 31:201–208, 1997.

A. M. Chamberlain, R. D. Brown, A. Alonso, B. J. Gersh, J. M. Killian, S. A. Weston, and V. L. Roger. No decline in the risk of stroke following incident atrial fibrillation since 2000 in the community: A concerning trend. *Journal of the American Heart Association*, 5:1–9, 2016.

A. L. Christensen, L. H. Rasmussen, M. G. Baker, G. Y. H. Lip, C. Dethlefsen, and T. B. Larsen. Seasonality, incidence and prognosis in atrial fibrillation and stroke in Denmark and New Zealand. *BMJ Open*, 2:1–8, 2012.

Coordenação Nacional para as Doenças Cardiovasculares. Terapêutica Antitrombótica da Fibrilhação Auricular. URL: http://www2.portaldasaude.pt/NR/rdonlyres/6B7EB2F9-CF0B-4368-A073-01C0D72998E1/0/TerapeuticaAntitromboticaFibrilhacaoAuricular.pdf, 2009. (accessed on em 04/2020).

J. Costa, F. Fiorentino, D. Caldeira, M. Inês, C. L. Pereira, L. Pinheiro, A. Vaz-Carneiro, M. Borges, and M. Gouveia. Custo-efetividade dos novos anticoagulantes orais na fibrilhação auricular em Portugal. *Revista Portuguesa de Cardiologia*, 34 (12):723–737, 2015.

J. C. Cowan, J. Wu, M. Hall, A. Orlowski, R. M. West, and C. P. Gale. A 10 year study of hospitalized atrial fibrillation-related stroke in England and its association with uptake of oral anticoagulation. *European Heart Journal*, 39:2975–2983, 2018.

M. Gouveia, J. Costa, J. Alarcão, M. Augusto, D. Caldeira, L. Pinheiro, A. V. Pinheiro, and M. Borges. Carga e custo da fibrilhação auricular em Portugal. *Revista Portuguesa de Cardiologia*, 34(1):1–11, 2015.

R. Hyndman and Y. Khandakar. Automatic Time Series Forecasting: the forecast Package for R. *Journal of Statistical Software*, 26(3):1–22, 2008.

G. Ljung and G. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303, 1978.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.r-project.org/, 2020. (accessed on 04/2020).

G. R. Shroff, C. A. Solid, and C. A. Herzog. Atrial Fibrillation, Stroke and Anticoagulation in Medicare Beneficiaries: Trends by Age, Sex, and Race, 1992-2010. *Journal of the American Heart Association*, 3:1–10, 2014.

R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications With R Examples*. Springer, 2^{nd} edition, 2006.

S. Suzuki, T. Otsuka, K. Sagara, H. Semba, H. Kano, S. Matsuno, H. Takai, Y. Kato, T. Uejima, Y. Oikawa, K. Nagashima, H. Kirigaya, T. Kunihara, J. Yajima, H. Sawada, T. Aizawa, and T. Yamashita. Nine-year trend of anticoagulation use, thromboembolic events, and major bleeding in patients with non-valvular atrial fibrillation - Shinken database analysis. *Circulation Journal*, 80:639–649, 2016.

G. S. C. Yiin, D. P. J. Howard, N. L. M. Paul, L. Li, Z. Mehta, and P. M. Rothwell. Recent time trends in incidence, outcome and premorbid treatment of atrial fibrillation-related stroke and other embolic vascular events: a population-based study. *Journal of Neurology, Neurosurgery and Psychiatry*, 88:12–18, 2017.