Journal of Statistics on Health Decision

P7

Prediction of the number of hospitalizations due to diabetes in Portugal: a time series analysis

Mónica Fialho[1,2], Marília Antunes[1,3]

[1] Centro de Estatística e Aplicações da Universidade de Lisboa, Faculdade de Ciências, 1749-016 Lisboa, Portugal
[2] Instituto de Saúde Ambiental (ISAMB), Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal
[3] Departamento de Estatística e Investigação Operacional, Faculdade de Ciências, 1749-016 Lisboa, Portugal

Introduction

Diabetes is a chronic multifactorial disease characterized by high blood sugar levels, as a result of the body's incapacity to produce or make proper use of insulin [1]. This disease affects over 460 million adults worldwide, representing an important risk factor for vascular diseases and early mortality [1–4]. In Portugal, diabetes affects 9.8% of the population aged between 20 and 79 years, having been the cause of 3.8% of all deaths in 2018 [1,5]. Diabetes prevention should take place, essentially, at primary health care, in order to avoid or delay vascular complications. Due to inadequate glycaemic control, hospital admissions are frequent and hence deserve careful analysis, with particular emphasis on hospitalizations, to which most costs with diabetes are attributable [5]. A better understanding of the epidemiology of the disease is crucial to support medical decisions and allocate health resources [5]. This work focus on the temporal evolution of a particular type of hospital admissions, namely hospitalizations lasting at least one day caused by diabetes in Portugal, with an emphasis on prediction.

Methods

Records of hospital admissions with main diagnosis of diabetes (coded 250, according to International Classification of Diseases (ICD), 9th Revision, Clinical Modification (ICD-9-CM), E10, E11 or E13, based on ICD10CM/PCS) were extracted from the Hospital Morbidity Databases provided by the Central Administration of the Health System, I.P. (ACSS) of the Portuguese Ministry of Health, covering a period of nine years (2010-2018). A time series analysis on monthly hospitalizations in Portugal between January 2010 and December 2018 was conducted, using data from 2010 to 2016 (84 observations) as training set for model identification and estimation, and the remaining observations for model validation. Following the Box-Jenkins approach for Seasonal Autoregressive Integrated Moving Average (SARIMA) modelling [6], several models were identified as suitable. Akaike information criterion (AIC) was used to select the best model, with further analysis of residuals for diagnostic purposes. Concerning the predictive ability of the model, 1, 3, 6 and 12-month forecasts were computed based on a rolling-origin evaluation and compared using the Mean Absolute Error (MAE), the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) [7]. All statistical analyses were performed using R version 3.6.3 and RStudio version 1.3.618.

Results

From 2010 to 2018, there were 73 050 hospitalizations (35% of the 208 882 admissions due to diabetes; mean of 676 cases per month, 8 117 cases per year), with a decreasing trend. The number of hospitalizations dropped by 45%, with the highest number of records in 2010 (10 011) and a minimum in 2018 (5 530 cases). Besides trend, also seasonality can be observed in the series. A lower number of cases typically occurs in Summer months, as opposed to Winter, the season that records the greatest number of hospitalizations (Figure 1). The procedures of model identification and selection led to the choice of the model SARIMA$(1,1,2)\times(0,1,1)_{12}$ (AIC=10.647), for which rolling-origin-update and rolling-origin-recalibration forecasts were calculated, given fixed and rolling windows. Overall, rolling-origin-recalibration performs better, with the best results being obtained by the 1-month horizon rolling-window forecast (MAPE=7.8; Table 1, Figure 2). As the forecast timespan increases, the predictive accuracy of the model worsens.

Keywords:
SARIMA, Forecasting, Rolling-origin, Diabetes, Hospitalizations

Corresponding author:
Mónica Fialho
monicafialho15@gmail.com

First published: 23 OCT 2020

Table 1 - Accuracy measures of rolling-origin-update and rolling-origin-recalibration forecasts.

	Rolling-origin-update			Rolling-origin-recalibration					
	Fixed window			*Fixed window*			*Rolling window*		
	MAE	RMSE	MAPE	MAE	RMSE	MAPE	MAE	RMSE	MAPE
l = 1	44.9	54.8	9.1	41.1	48.7	8.2	39.5	47.4	7.8
l = 3	40.4	49.1	8.3	41.3	48.4	8.4	40.8	47.8	8.3
l = 6	45.0	56.8	9.8	44.0	55.9	9.5	44.9	56.4	9.7
l = 12	69.1	81.4	16.0	68.2	79.8	15.8	70.0	81.5	16.2

MAE, Mean Absolute Error; RMSE, Root Mean Square Error; MAPE, Mean Absolute Percentage Error; l, lead time (forecast horizon)

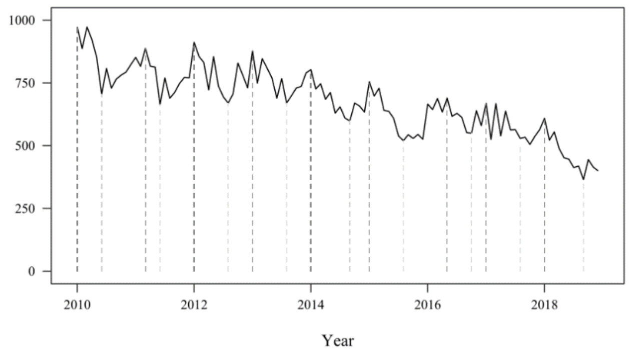

Figure 1 - Monthly hospitalizations due to diabetes from 2010 to 2018. Reference lines demonstrate seasonality, with peaks apparent between January and March, and troughs mainly between June and August.

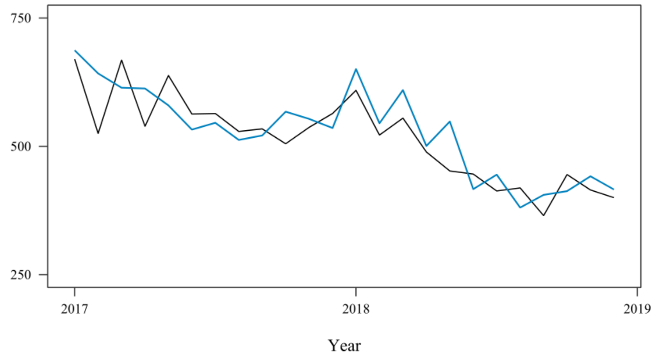

Figure 2 - 1-month forecasts for SARIMA(1,1,2)×(0,1,1)$_{12}$ model. Observed values (black line) and forecasts for years 2017 and 2018 based on rolling-origin-recalibration with rolling window procedure (blue line).

Conclusion

Despite the increasing trend of hospital admissions due to diabetes, hospitalizations have been decreasing since 2010. Results from time series analysis suggest that the SARIMA(1,1,2)×(0,1,1)$_{12}$ model can be used to predict future cases with good accuracy, provided that environmental and individual risk factors remain relatively stable over time, as sudden and dramatic changes could result in an increased demand for health care services. Further disaggregation of data by geographical region, with parameter re-estimation or identification of a new model, would potentiate the applicability and usefulness of such model, so that patients' needs can be anticipated, and medical resources assigned in a more effective way.

References:

1. IDF Diabetes Atlas. 9th ed. International Diabetes Federation; 2019.
2. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. Lancet. 2010;375(9733):2215–22. https://doi.org/10.1016/S0140-6736(10)60484-9
3. Raghavan S, Vassy JL, Ho YL, Song RJ, Gagnon DR, Cho K, et al. Diabetes mellitus–related all-cause and cardiovascular mortality in a national cohort of adults. J Am Heart Assoc. 2019;8(4). https://doi.org/10.1161/JAHA.118.011295
4. Lee Y Bin, Han K, Kim B, Lee SE, Jun JE, Ahn J, et al. Risk of early mortality and cardiovascular disease in type 1 diabetes: A comparison with type 2 diabetes, a nationwide study. Cardiovasc Diabetol. 2019;18(1):1–17. https://doi.org/10.1186/s12933-019-0953-7
5. Ministério da Saúde. Direção-Geral da Saúde. Programa Nacional para a Diabetes 2019. Desafios e Estratégias. Lisboa: Direção-Geral da Saúde; 2019.
6. Box GEP, Jenkins GM, Reinsel GC, Ljung GM. Time series analysis: forecasting and control. 5th ed. Hoboken, New Jersey: Wiley; 2016.
7. Tashman LJ. Out-of-sample tests of forecasting accuracy: An analysis and review. Int J Forecast. 2000;16(4):437–50. https://doi.org/10.1016/S0169-2070(00)00065-0