On the use of Statistical Process Control in Monitoring Mortality. An Application to European Countries

Ana Debón[1], Vicent Giner-Bosch[1], Majarlika Cabrerizo-Cabanos[2]

[1] *Centre for Quality and Change Management, Universitat Politècnica de València, Valencia, Spain;* *andeau@eio.upv.es, vigibos@eio.upv.es*

[2] *Departamento de Estadística e Investigación Operativa Aplicadas y Calidad,*Universitat Politècnica de València

Abstract
The evolution of mortality is a key global concern from both an economic and social point of view. In particular, being able to detect and predict changes in mortality compared to its expected behaviour as accurately as possible is a desirable goal. In this context, the standardized mortality ratio (SMR) is commonly used in order to measure the mortality of a country with regard to its neighbouring countries at a given moment in time.
In this work, we address the study of the evolution of the SMR of a country over time, modelled as a time series. The joint use of time series and statistical process control (SPC) techniques to model and monitor the behaviour of the SMR is explored. Both approaches are relevant and complementary. On one hand, time series are an appropriate tool to study and characterize the evolution of SMR over time and to forecast it. On the other hand, SPC permits detection of significant changes in the trend of the variable being monitored. More precisely, we suggest monitoring the residuals of the fitted time series model using control charts.We present and discuss the results of applying our proposal to mortality data for European countries in a 20-year period. These results show the relevance of our approach and delineate our next research steps.
Finally, the use of other approaches combining time series and SPC techniques is also outlined.

Keywords: mortality, standard mortality ratio, time series analysis, statistical process control, control charts.

1. Introduction

In recent decades we have witnessed remarkable progress in health throughout the world. Due to these advances, death rates from infectious diseases and cardiovascular diseases have decreased in overall terms.

The causes of death vary widely according to the country, but according to the World Health Organization (WHO) (2015), in the period from 2010 to 2012 the leading causes of mortality worldwide were ischemic heart disease, strokes, infections of the lower respiratory tract and chronic obstructive pulmonary disease. However, historically, not only diseases have influenced mortality (Lancaster, 2012). In developed countries, people mainly die of chronic diseases, namely cardiovascular disease, cancer, dementia, chronic obstructive pulmonary disease or diabetes. Infections of the lower respiratory tract are the only leading infections of note that cause of death. A high infant mortality rate is also unusual. On the other hand, in low-income countries, the leading causes of death are infectious diseases. These same countries are also where there are major childbirth complications, leading to more deaths of newly born children and those under 1 year of age (according to WHO).

Therefore, it is clear that the rates have not changed only as a function of time (historical evolution), but also due to the geographical area, since not all countries have the same health and economic conditions.

Mortality in all European countries has dropped sharply in the twentieth century and what we have seen of this century. All countries have experienced similar patterns of change. In the first half of the twentieth century mortality decline was due to a sharp decrease in infectious diseases; however, at the same time, cardiovascular disease and cancer increased, becoming the predominant causes of death in the second half of the twentieth century. Despite these similar trends in European countries, there are still considerable differences in mortality levels (EUROSTAT, 2009).

This work focuses on Europe. In short, it aims to model and monitor the variability of a parameter as important as the standardized mortality ratio (SMR) through the combined use of appropriate time series and control chart tools. The use of combined control charts with time

series was previously studied by Knoth and Schmid (2004). However, as far as we can tell, the study of mortality time series using control charts is a field which is yet to be explored.
This paper is structured as follows. In Section 2 basic concepts such as SMR are introduced to measure the phenomenon of demographic mortality. The following section briefly explains the statistical techniques used to analyze the data: time series and control charts. Later, in Section 4, the techniques mentioned are applied to mortality data from European countries. In this section, each country is studied individually and as part of one of two groups based on their geographical location, which constitutes the main contribution of this work. Finally, in Section 5, the main conclusions arising from this work are presented.

2. Data and statistics

This study deals with mortality data for European countries for the period 1990 to 2009 with an age range from 0 to 110+. The data was taken from the Human Mortality Database (2015) for a total of 26 countries: Austria, Belarus, Belgium, The Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Netherlands, Norway, Poland, Portugal, Slovakia, Slovenia, Spain, Sweden, Switzerland, The United Kingdom and Ukrainia.

The statistic used to study the mortality of European countries is the Standardized Mortality Ratio (SMR). The SMR is defined as the ratio of the actual number of deaths in studied population A to the number of deaths that would be expected in population A if it experienced the age-specific death rates of the standard population (Hinde, 2014). If the SMR is greater than 1, there are "excess deaths" in the studied population. The SMR is used to compare the mortality risk of a studied population and is defined as:

$$SMR_{i,t} = \frac{O_{i,t}}{E_{i,t}} \quad \text{if } i \in \{1, \dots, N\} \text{ and } t \in \{1, \dots, T\} \quad (1)$$

where i is the European country and t the moment of time (in our data t is expressed in years). $O_{i,t}$ represents the number of observed deaths for each European country at the moment of time t and $E_{i,t}$ the expected deaths for each European country at the moment of time t, if all countries had the same mortality as the whole of Europe.

3. Statistical Methods

Statistical analysis was performed using the R environment for statistical computing (R Core Team, 2015) together with several R-packages. tseries (Trapletti y Hornik, 2015), qcc (Scrucca, 2004).

3.1. Time Series

Prediction of the mortality index requires the previous adjustment of a time series to SMR, which is carried out by means of the Box–Jenkins methodology using an R-package forecast developed by Hyndman (2015) (Hyndman and Khandakar, 2008). The standard autoregressive integrated moving average (ARIMA) models are denoted ARIMA(p,d,q) where p, d and q are autoregressive, difference and moving average orders, respectively. The time series are extensively described in Peña (2005).

3.1. Control Charts

Statistical process control (SPC), whose basic theory was developed by Walter Shewhart in late 1920, is a set of tools used to achieve stability in a process by reducing variability. The key tool in SPC is the control chart, and this is what will be used in this work. Montgomery (2012) can be consulted for more information about this and other SPC tools. A summary of EWMA charts is presented below, as they will be used in this work due to the data being individual observations. EWMA graphics or exponentially weighted moving averages are usually performed on individual observations. This chart also gathers the values of past observations in each period. The variable that is represented in each period is an average of present observations and past observations, where more weight is given to the most recent observations. In general, this type of average, where new information is being incorporated constantly and weighting is subtracted from the historical information is called moving average. EWMA charts use a very specific way of creating moving averages, which is to weight the historical information which decays exponentially with time. Overall, the EWMA chart is less sensitive to small deviations than the CUSUM chart (another commonly used control chart which is not described in this section as it is not used in this paper).

4. Application of SMR in Europe

4.1. Individual results for each country

The process under control is mortality in Europe in terms of the SMR index, calculated according to expression (1). The SMR was obtained for 26 European countries from 1990 to 2009 from data of Section 2. Figure 4 shows the values of the SMR for each European country studied between 1990 and 2009.

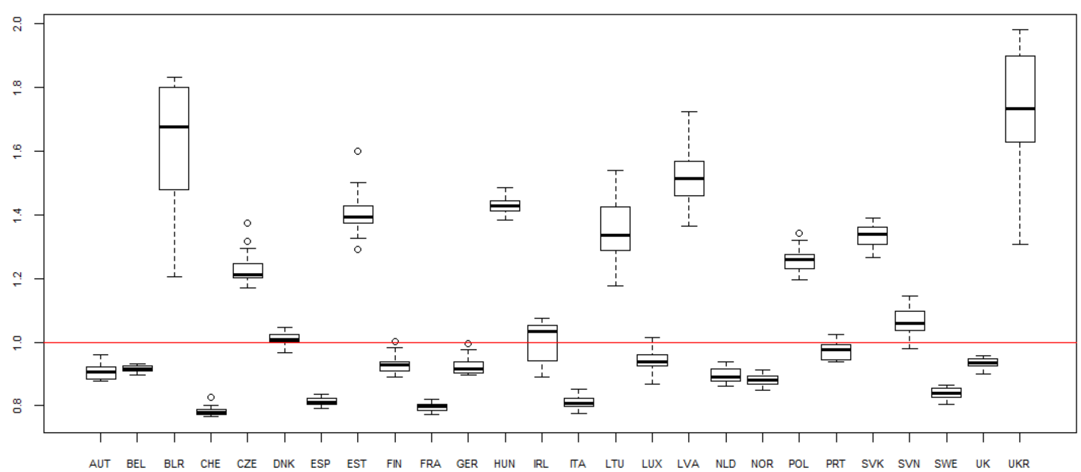

Figure 1: SMR dispersion chart for the countries studied

We differentiate between countries with SMR values greater or less than 1:

- Countries with SMR less than 1: Germany, Austria, Belgium, Denmark, Spain, Finland, France, Holland, Ireland, Italy, Luxembourg, Norway, Portugal, The United Kingdom, Sweden and Switzerland. An SMR less than 1 means that the deaths observed in these countries are below what would be expected, if they behaved like Europe as a whole, therefore privileged countries, as their "status" is better overall. This SMR below 1 was maintained throughout the study period.

- Countries with SMR greater than 1: Belarus, Estonia, Hungary, Latvia, Lithuania, Poland, The Czech Republic and Ukrainia. Deaths observed in these countries are above what would be expected, if they behaved like the whole of Europe, and therefore not privileged countries, as their "status" is worse in general. The levels of SMR above 1 were maintained during the study period.
- Countries with SMR about 1: Slovenia. This country has similar deaths to what would be expected if it behaved like the whole of Europe and, therefore, a similar overall "status". For the studied years, the SMR has both fallen below and risen above 1.

Furthermore, in Figure 1 the variability of SMR can be seen for each country. For example, Belarus and Ukraine are the countries with most variability in SMR; while Belgium and Spain are two of the countries with the lowest variability in the SMR.

The SMR index for a year has autocorrelation which is why using the SPC directly on these correlated data will produce tight control limits and cause a substantial increase in false alarms thereby decreasing the ability to detect changes in the process (Psarakis and Papaleonida, 2007). One possible strategy is to fit a time model to them, and then apply the traditional control charts to the residuals. The reason for using control charts on the residuals is that, assuming a correct time series model fitted to the data, the residuals are independent and identically distributed random variables, thus fulfilling the assumptions necessary for using control charts. If a change is detected in the average of the residuals, it means that the process mean has changed. Thus, presenting residuals on a control chart provides a mechanism to detect changing trends or behaviour in the time series.

Firstly, the series of 20 data points is divided into two groups: a training set (first 15 observations) and a validation set (last five observations). Then using the methodology of Box and Jenkins (1976) the predictive model (AR, MA, ARMA or ARIMA) which best fits the training set is chosen (Peña, 2005). After testing the model and validating it, residuals are monitored with a EWMA chart. With these charts it is possible to discern in which years and in which countries there was a level of mortality outside the expected.

The analysis was performed for each of the countries, for brevity and since the methodology is repetitive, the analyses are not presented, although the results of all findings will be commented on. To improve the properties of homocedasticity we will work with the logarithm of the SMR time series.

4.1. Results for countries in the east and west of Europe

History, culture, religion, economy, social development and especially traditions in Eastern Europe have had a different evolution to the countries of Western Europe as indicated in Herţa-Gongola (2004). An example of the contrast in these two regions is the level of SMR. Figure 1 shows that countries in Eastern Europe such as Ukrainia, Hungary and Poland have a value above 1; while countries with lower values such as France, Switzerland and the United Kingdom form part of Western Europe. For this reason, it is convenient to group, on the one hand, countries with high mortality, and on the other hand, countries with low mortality.

The geographical grouping is as follows:

- Western Europe: Portugal, Spain, France, Switzerland, Italy, Austria, Luxembourg, Belgium, Germany, Holland, Denmark, Sweden, The United Kingdom, Ireland, Norway and Finland.
- Eastern Europe: Estonia, Latvia, Lithuania, Poland, Belarus, Ukrainia, The Czech Republic, Slovakia, Hungary, Slovenia.

The control chart for the mean and range of the residuals of the ARIMA models fitted to the log(SMR) for the countries in Eastern Europe and Western Europe are shown in Figure 2 and 3, respectively. On the left side (Figure 2), there are two points outside the control limits, corresponding to the last two years; this indicates that the deaths observed were less than expected deaths in those years. On the left side (Figure2) there is an observation outside the control limits which indicates that during the last year there were large differences in the SMR values between the countries that constitute the region of Eastern Europe. Considering the control for the range of residuals for the countries in Western Europe (Figure 3, rigth), more abnormalities are seen than in Eastern Europe. In this chart there is an observation outside the control limits. This point indicates that during the last year there were large differences in the values of SMR between the countries that form the region of Western Europe. In addition, there is a series of 7 observations (between 1997 and 2003), in which the differences between the countries is less than the average.

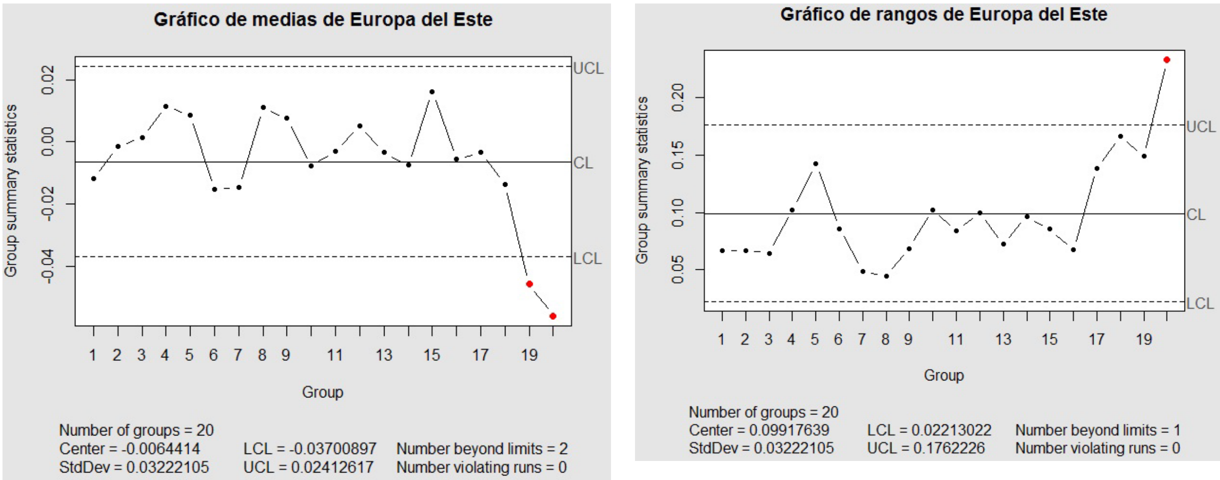

Figure 2: Control chart for the average (left) and range (right) of the residuals of the ARIMA models fitted to the log(SMR) for countries in Eastern Europe

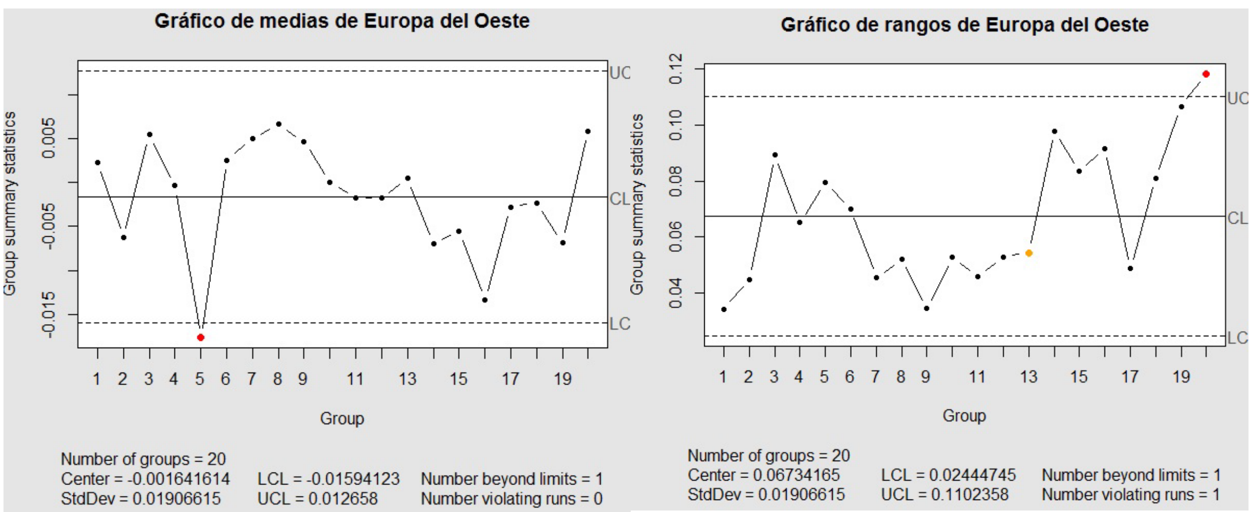

Figure 3: Control chart for the average (left) and range (right) of the residuals of the ARIMA models fitted to the log(SMR) for countries in Western Europe

5. Conclusions

In this work, monitoring mortality through control charts was not carried out directly on the SMR values, as the index for a particular year has temporal autocorrelation, i.e. a connection with past observations. Instead, the SMR series was modelled for each country through a time series prediction model, with the residuals generated by each model (normal and independent), values which were subsequently monitored with EWMA control charts.

Of the total of 26 countries considered, those exhibiting behaviour outside the expected range according to the methodology described are: Slovenia in 2009, Estonia in 2009, France in 2006, 2007, 2008, Holland in 2007, 2008 and 2009, Hungary in 2008 and 2009, Ireland in 2008 and 2009, Italy in 2009, Latvia in 2007, 2008 and 2009, Lithuania in 2009, The United Kingdom in 2009, The Czech Republic in 2008 and 2009 and Ukrainia in 2008 and 2009.

According to our methodology, the alarms in these years are symptoms of a possible change of behaviour in mortality in these countries.

In our study, we observed the existence of two types of very different country: countries with an SMR greater than one (which, by the definition of SMR indicates the observation of more deaths than would be expected by comparison with the rest of Europe) and countries with an SMR below 1 (fewer deaths than expected). Geographically, this distribution generally corresponds to countries in Eastern Europe and Western Europe, respectively. For this reason it is interesting to compare the mortality of these two regions. Eastern Europe has higher average residuals than Western Europe. Having higher average values indicates that in Eastern Europe countries there is less fit in the time series models, which is due to a greater range of residuals. In this regard, Western Europe has an average range of residuals lower than Eastern Europe, which means that mortality is much more homogeneous and stable in Western Europe.

6. Acknowledgments

This work was supported by a grant from Ministerio de Economía y Competitividad (MTM2013-45381)

7. References

EUROSTAT (2009). Health statistics-Atlas on mortality in the European Union: 2009, edition. Luxembourg.

Herţa Gongola, L. (2004). Western European "identity" versus Eastern European "identity"? Studia Universitatis Babes-Bolyai-Studia Europaea, (1-2), 135-142.

Hinde, A. (2014). Demographic methods. Routledge.

Hyndman R.J. (2015). forecast: Forecasting functions for time series and linear models. http://github.com/robjhyndman/forecast. R package version 6.1.

Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available atwww.mortality.org or www.humanmortality.de (data downloaded on [29/07/2015]).

Knoth S, Schmid W (2004). Control Charts for Time Series: A Review. En: Lenz H, Wilrich P (eds.), Frontiers in Statistical Quality Control 7. 210-236. Springer-Verlag, Berlin.

Lancaster H.O. (2012). Expectations of life: a study in the demography, statistics, and history of world mortality. Springer Science & Business Media, New York, NY.

Montgomery DC (2012). Introduction to Statistical Quality Control. 7.aed. John Wiley & Sons, Hoboken, NJ.

Peña, D., 2005. Análisis de Series Temporales. Alianza Editorial, Madrid.

Psarakis S, Papaleonida G (2007). SPC procedures for monitoring autocorrelated processes. Quality Technology and Quantitative Management, 4(4):501-540.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Scrucca L (2004). qcc: an R package for quality control charting and statistical process control. R News, 4(1):11-17. http://CRAN.R-project.org/doc/Rnews/

Trapletti A, Hornik K (2015). tseries: Time Series Analysis and Computational Finance. http://CRAN.R-project.org/package=tseries. R package version 0.10-34.