Obtaining a mortality baseline free of influenza epidemic effects using models with no covariates

Baltazar Nunes[*], Department of Epidemiology,
National Health Institute Dr. Ricardo Jorge, Portugal

Isabel Natário, CEAUL, Department of Mathematics,
Faculty of Sciences and Technology, New University of Lisbon, Portugal

M. Lucília Carvalho, CEAUL, Department of Statistics and
Operation Research, Faculty of Sciences, University of Lisbon, Portugal

December 18, 2008

Abstract

The occurrence of influenza epidemics during winters, in the northern hemisphere countries, is known to be associated with observed excess mortality for all causes. A large variety of methods has been developed in order to estimate, from weekly or monthly mortality time series, the number of influenza-associated deaths in each season. The

[*]Address: Departamento de Epidemiologia, Instituto Nacional de Saúde Dr.Ricardo Jorge, Av. Padre Cruz, 1649-016 Lisboa, Portugal. Telephone: +351217562453 Email: baltazar.nunes@insa.min-saude.pt

present work focus on the group of methods characterized by fitting statistical models to interrupted mortality time series. The study objective is to find a common ground between these methods in order to describe and compare them. They are unified in a single class, being categorized according to three main parameters: the model used to fit the interrupted time series and obtain a baseline, the chosen periods used to estimate the influenza epidemic periods and the procedure used to fit the model to the time series (iterative or non iterative). This generalization led quite naturally to the construction of a set of user friendly R-routines for implementing all these models. These routines were applied to data on about 20 years of weekly Portuguese number of deaths by pneumonia and influenza showing that, in this case, the parameter that had the highest impact on influenza-associated deaths estimates was the chosen period used.

Keywords: influenza, baseline, excess deaths, cyclical regression, ARIMA

1 INTRODUCTION

In the northern hemisphere countries, during influenza epidemic periods, a rise in mortality from all causes is usually observed, mainly in the elderly population (aged 65 years or more)[1]. This increase can be associated with influenza epidemics since the influenza infection might cause complications that can lead to the hospitalization and/or death of the infected individual

[2, 3]. In this context, and from a public health point of view, the quantification of the seasonal influenza epidemics impact on the population and its description in terms of the dominant virus strain and level of vaccine coverage is of the utmost importance.

The measurement of influenza impact in terms of deaths or hospitalizations is never accessed by the number of deaths with influenza as the main cause in national mortality registries because this value is usually very low, even during the most severe epidemics. This is mainly due to the difficulty in establishing a connection between a complication (pneumonia or other respiratory diseases, circulatory system diseases, etc) and a previous or current influenza infection, due to the lack of a laboratory confirmed diagnosis. As a consequence, the use of official death registries, with influenza as cause to measure the influenza epidemic impact would underestimate its real effect [1]. This has led researchers to look for reliable methods to estimate influenza-associated deaths, using as a starting point a mortality time series and, when available, additional information from influenza epidemiologic surveillance systems on the seasonal epidemic characteristics.

Generally the methods used to estimate the excess deaths attributable to the influenza epidemics are divided in three steps:

1. obtaining a baseline of the number of deaths, by a certain time unit, in the absence of influenza epidemics;

2. using the baseline to identify the periods where there is evidence of an

excess of deaths attributable to influenza epidemics;

3. subtracting this baseline from the observed number of deaths, during these periods.

In this sense the observed excess of deaths above the baseline, when associated with influenza epidemic periods, could, in the absence of other explainable events, be attributed to an influenza epidemic. The state-of-the-art of the methods to estimate the excess deaths attributable to influenza epidemics offers a large variety of different alternatives, all applicable to identical situations and aiming essentially the same purpose. These methods can be classified into two general methodological approaches and, within those, they vary in a considerable number of aspects. In the first group, the methods are based on statistical models that include influenza activity indicators as explanatory covariates. The methods in the second group are characterized by not considering covariates and also by excluding from the estimating process all the parts of the mortality time series where there is evidence of influenza epidemics occurrence.

The present study will be focused on the second group of methods. A close analysis of this group shows several differences among them, essentially on the type of statistical model employed (cyclical regression[5, 6, 7, 8, 9] or ARIMA models[12, 10]), on the method used to build the baseline (non iterative[8, 7, 9] or iterative[12, 5, 6, 10]) and on the choice of the periods to be excluded from the mortality time series (epidemic periods

defined using influenza-like illness surveillance systems[12, 5, 8, 10] or fixed periods, like December to April[6, 7, 9]). All these differences can lead to unequal influenza-associated deaths estimates. Differences between reported estimates have been identified leading specialists into a profound discussion without a final agreed conclusion [13, 14].

Here we were able to unify these methods in a single class, in such a way that it allows the description and comparison of their applicability and results. This proved to be an important step in the conceptualization of the statistical methods used to estimate influenza-associated deaths, clarifying all the steps performed and options taken to compute the desired estimates. This unification was also the basis to build an `R`-routine that easily estimates the influenza-associated deaths by any of the methods in the class. This platform is quite user friendly even for those less familiarized with the theoretical statistical developments that have led to the results. The application of this tool could also empower other researchers to critically analyze the differences and similarities between the estimates obtained with a variety of method choices, allowing in this way a more comprehensive analyzes of their data.

2 SOME ESSENTIAL CONCEPTS

2.1 Influenza epidemic period, influenza season and flu-year

An *influenza epidemic* is defined as the occurrence, in a specific population, of a **number of cases** of influenza above what is usually expected, during a certain period of time, referred as the *epidemic period*. Usually the epidemic periods are unknown and therefore must be estimated.

The annual fixed period of time during which the influenza epidemics might occur, starting sooner or later, with larger or smaller duration, is named *influenza season*. In Portugal as in other northern hemisphere countries, this period starts in October of each year ending in May of the next calendar year. This is the period when the influenza surveillance systems are more active as the occurrence of an influenza epidemic outside this period has an almost a null probability. Taking into account the beginning and ending of the influenza season, a *flu-year* a is defined as the 52 (or 53 when the first calendar year of the *flu-year* is bissextile) weeks that start at week 27 of any calendar year n and ends at week 26 of the calendar year $n+1$.

Let E_a denote the estimate of an influenza epidemic period occurred during flu year a. The choice of these periods is much dependent on the level of information one has on the occurrence of influenza cases and on the temporal evolution of the influenza incidence rate in the population. In fact, to obtain a correct diagnosis of influenza, a confirmation of the

influenza virus presence is necessary, procedure that is not usually carried out. In the majority of the situations only the clinical diagnosis is obtained, without the laboratory confirmation, and if this situation occurs the case can only be classified as influenza-like illness.

Therefore, the information that is usually available consists on the temporal evolution of the influenza-like illness incidence rates, complemented by information on the influenza virus circulation among the population. In the majority of the developed countries this information is collected by surveillance systems specifically design for the effect, that are based on a sample of individuals set under surveillance.

When this information does not exist, or is not available, same authors [7, 9] have set E_a as the fixed time period, enclosed in the influenza season period, that goes from December to April of the next calendar year. Other solution is to use the time series of mortality specific by influenza (ICD 9th Revision:487; ICD 10th Revision:J9-J11), and define E_a as the periods where the mortality by influenza rises above the expected. In principle, given the under registration of deaths with influenza cause, this last option should be a less sensible but more specific method, given the low lethality rate of influenza infection.

2.2 Time series of the weekly number of deaths

Consider $y_{t,a}$ to be the time series of the number of deaths observed in week t, $t = 1(27), ..., 52(26)$[1] of flu-year a, $a = 1, ..., A$, where A represents the number of flu-years in study. This time series will be the main object of analysis, since the major goal is to estimate the excess number of deaths attributable to influenza epidemics from it. Usually the most used time series is the weekly (or monthly) number (or rate) deaths by all causes or pneumonia and influenza.

2.3 Periods with excess deaths attributable to influenza epidemic

Consider D_a the period of weeks where **an excess deaths** in $y_{t,a}$ is attributed to an influenza epidemic, in flu-year a. This period is defined by an observed increase in $y_{t,a}$, during the E_a period, above the expected in the absence of the effect of an influenza epidemics. Additionally, during this period, there must be no other events that can be the cause of the observed excess deaths.

[1] The values presented in parenthesis represent year calendar week number

2.4 Mortality baseline in the absence of the influenza epidemics effect

Excluding from the time series the parts where there is evidence of an influenza epidemic occurrence, one obtain the following *interrupted time series* denoted hereafter by $y^{\star}_{t,a} = \{y_{t,a} : (t,a) \notin E_a\}$.

Let $\beta_{t,a}$ be the baseline resulting from fitting a statistical model to the time series $y^{\star}_{t,a}$ or, as some authors have considered, to $x_{t,a}$, the weekly number of deaths in the absence of the influenza epidemics, defined as follows:

$$x_{t,a} = \begin{cases} y_{t,a}, & (t,a) \notin D_a; \\ \tilde{y}_{t,a}, & (t,a) \in D_a \end{cases}$$

where $\tilde{y}_{t,a}$ represents some preliminary estimate of the weekly number of deaths in the absence of the effect of an influenza epidemics for the week t of the flu-year a.

3 DESCRIPTION OF METHODS IN STUDY

The studied methods are all characterized by obtaining a mortality baseline in the absence of influenza epidemics effects using an interrupted mortality time series. Generally these methods fit a statistical model to $y^{\star}_{t,a}$ to obtain a baseline $\beta_{t,a}$ that is used to identify the periods with excess deaths attributable to an influenza epidemic, D_a.

To be able to jointly describe all these procedures one has to identify the unifying characteristics and also their differences in order to summarize

them in a few classes. We found three sources of dissimilarity

1. **Statistical model used to fit the interrupted time series**: There are mainly two types of models used in the literature:

   (a) multiple linear regression models [5, 6, 7, 8, 9], formed by a polynomial component to explain the series trend and a sinusoidal component that captures the seasonality observed. Generally these models are given by:

   $$y_s = \alpha + \sum_{i=1}^{m} a_i s^i + \sum_{j=1}^{l} b_{1,j} \sin \frac{j2\pi s}{52} + b_{2,j} \cos \frac{j2\pi s}{52} + \varepsilon_s$$

   where $a_i,\ i = 1, ..., m$, are the parameters of the order m polynomial function used to explain the trend, $b_{1,j}$ and $b_{2,j}$ are the parameters of the sinusoidal function with periods $52/j,\ j = 1, ..., l$, used to explain the eventual seasonality and $\varepsilon_s \sim N(0, \sigma^2)$ with $s = t + (a-1)52,\ t = 1, ..., 52$ and $a = 1, ..., A$.

   (b) seasonal ARIMA[15] only applied by [10, 12] to this problem.

2. **Choice of the E_a periods**: In some of the reviewed papers this period was the epidemic period defined (estimated) by the operating Influenza Surveillance Systems (ISS) using data on clinical diagnosis of influenza-like illness and viral strains isolates [5, 8, 12]. In this case the chosen E_a periods are different from flu-year to flu-year.

   Other authors [6, 7, 9] defined E_a as a fixed set of weeks (December to April), in each flu-year, where the occurrence of an influenza epidemic

with effects on mortality is more likely. This period is always included in the influenza season.

3. **Procedure used to fit the statistical model and to identify the D_a periods**

   (a) Non-iterative: the model is fitted to all points of the interrupted time series $y^{\star}_{t,a}$ [7, 8, 9] at once. Here the baseline $\beta_{t,a}$ corresponds to the estimated values given by the model for each week t. In [8] the D_a periods are defined as the set of weeks, contained in the E_a periods, that initiate with two consecutive weeks with a number of deaths above the upper 95% confidence limit of the baseline and end with two consecutive weeks with a number of deaths bellow the same upper limit. On the other hand [7, 9] have defined the D_a periods applying the previous method only to the mortality time series specific for influenza and not to the time series in study (i.e pneumonia and influenza).

   (b) Iterative: generally, these methods consist in forecasting a baseline for each flu-year i using a statistical model fitted to $x_{t,a}$ for a training set of T previous flu-years. This training set can have a fixed dimension T (equal for all iterations, e.g. 5 years) [5, 6], or be given by all previous years of flu-year i of that iteration [12]. In the iterative methods, D_i is identified in each iteration i

as the period of weeks contained in the correspondent E_i period that initiates with two consecutive weeks with a number of deaths above the upper 95% confidence limit of the forecasted baseline and terminate with two consecutive weeks with number of deaths bellow the same upper limit. After the D_i identification the series $x_{t,a}$ is updated. Here [12] update $x_{t,a}$ by substituting its values in the D_i periods by the values of the forecasted baseline, obtaining a preliminary estimate of the mortality time series in the absence of influenza epidemic $\tilde{y}_{t,a}$ during those periods. Other authors simply use an interrupted time series were the values in the D_i periods are excluded from the original time series.

4 GENERAL FRAMEWORK

Given the above description it is possible to accommodate a large number of methods in a wide framework of methods varying according to the three points considered, model fitting procedure, type of model and period where to identify the excess mortality periods attributable to the influenza epidemics D_a - see Table 1.

Note that we were able to further identify three new methods, never considered before, from taking all possible combinations of alternatives:

1. `It_SA_F` iterative, using the the seasonal ARIMA model with E_a pe-

riods fixed;

2. `SA_F` and `SA_E` that apply the seasonal ARIMA models with a non iterative procedure;

The `It_SA_F` does not seem to present any practical application problems. For methods `SA_F` and `SA_E` we propose to adjust a multiple linear regression model to the interrupted time series, substitute then the E_a periods by the model expectations, and then apply the seasonal ARIMA models to this new series.

In order to evaluate the performance of the models presented (Table 1) two items are considered:

1. Adjustment quality of the baseline $\beta_{t,a}$ to the original time series $y_{t,a}$ for $(t,a) \notin D_a$, measured in terms of:

   - Residual Mean Square

     $$RMS = \sum_{(t,a)\notin D_a} \frac{\hat{e}_{t,a}^2}{m}$$

     where m is the number of weeks without excess deaths attributable to influenza $\hat{e}_{t,a} = y_{t,a} - \beta_{t,a}$ for $(t,a) \notin D_a$;

   - Auto-correlation function of the residuals;

2. Number of excess deaths attributable to influenza epidemics for each flu-year a.

3. Correlation of the influenza associated deaths estimates between methods, has an empirical concordance measure.

It is important to notice that for item 2 the true values are not known. In this sense we can only compare the obtained results with each other and identify the methods that estimate the higher and lower values of excess deaths.

5 APPLICATION EXAMPLE

The analyzed data consists on the weekly number of deaths by pneumonia and influenza in Portugal from 1980-81 to 2003-04 flu-years obtained from the National Mortality database of the Portuguese Statistics Institute (Figure 1).

As presented in Table 2 the E_a periods were either set as fixed periods (from week 48 (December) to week 17 (April)) or equal to the non fixed epidemic periods that were defined as follows:

- From 1980-81 to 1989-90 the influenza epidemic periods were defined using the weekly number of deaths by influenza (ICD 9th Revision 487). These periods were set as the consecutive weeks (more than two) with the number of deaths above the 95 percentile of the weekly empirical distribution of the number of deaths by influenza in the

period comprised by the flu-years 1980-81 to 1989-90.

- For the flu-years from 1990-91 to 2003-04 the epidemic periods used were the ones defined by the Portuguese Sistema de Vigilância Integrada da Gripe (ISS), of the Instituto Nacional de Saúde Dr.Ricardo Jorge [16]. The final classification is presented in Table 2.

In this time series we have also substituted the known heat-waves periods [11] by the average of the number of deaths in the last week before and the first week after the heat-wave.

The models used in the application example were chosen in following way:

1. The multiple regression models used were chosen by a regular best fit model criteria, in a preliminar analysis.

   - non iterative procedure:

   $$y_s = \alpha + \beta_1 s + \beta_2 s^2 + \beta_3 s^3 + b_1 \sin\frac{2\pi s}{52} + b_2 \cos\frac{2\pi s}{52} + \varepsilon_s$$

   - iterative procedure:

   $$y_s = \alpha + \beta_1 s + \beta_2 s^2 + b_1 \sin\frac{2\pi s}{52} + b_2 \cos\frac{2\pi s}{52} + \varepsilon_s$$

2. The seasonal ARIMA models used were:

   - non iterative procedure: chosen by an automatic model identification algorithm [17];

- iterative procedure: a model analogous to the one proposed in [12], $ARIMA(2,0,0)(1,1,0)_{52}$:

$$(1-\phi_1 B-\phi_1 B^2)(1-\Phi_1 B^{52})\nabla^{52} y_s = \varepsilon_s,$$

where $s = t+(a-1)52$; $t = 1,...,52$; $a = 1,...,24$; $\varepsilon_s \sim N(0,\sigma^2)$.

In the present application example the D_a periods were always defined as the set of weeks, contained in the E_a periods, that initiate with two weeks with a number of deaths by pneumonia and influenza above the upper 95% confidence limit of the baseline and end with two weeks with a number of deaths bellow the same upper limit.

5.1 Results

The methods that considered the E_a periods as fixed presented lower RMS than the methods using year-variable sized periods, estimated by the ISS. Within each of these two different groups (fixed period or ISS periods) it was observed that the seasonal ARIMA model always presented lower values of RMS when compared to the multiple regression model – Figure 2). Autocorrelation in the residuals outside of the D_a periods was observed for all the methods that used multiple regression models.

From Figures 3 and 4 it can be observed that when the E_a period is set fixed, the number of influenza-associated estimated deaths is clearly higher. On the other hand, when the E_a periods are set by the ISS it is

patent an higher uniformity in the number influenza-associated estimated deaths, between the methods. This observation was confirmed by the high correlation coefficients obtained between non iterative methods and those using E_a defined by the ISS. Analyzing the results produced by the iterative approach one can identify a more disagreement behavior after flu year 1993-1994 mainly when the E_a period was fixed.

6 DISCUSSION

The present work constitutes an important step in the contextualization of the methods that use interrupted mortality time series to estimate influenza epidemics associated deaths. This achievement was obtained by the identification of the principal similarities and differences between the main existing methods, leading to the definition of a methodological class and a subsequent parametrization of its members in terms of the type of statistical models, period chosen to estimate the epidemic period and procedure employed to fit the model to the interrupted mortality time series and identify the periods with excess deaths associated to influenza epidemics.

Additionally, one other important output of this project is a set of `R`-routines that were enabled by the above methods parametrization. These user friendly routines constitute an important tool for the estimation of the influenza associated deaths, as the researcher can easily compare the results obtained by varying each one of the parameters involved. All the produced

R-codes will soon be available on the internet.

Pelat et al [18] have also developed a tool that is able to automatically obtain a baseline and quantify the excess of deaths attributable to influenza allowing the user to vary some parameters. Nevertheless they have only included cyclical regression models and have not made comparisons between the results obtained with the different methodological approaches.

We have to state that we have not parameterized all the differences between methods. For example, one that was not considered was the possible substitution of the values in the interrupted periods of the time series by corresponding forecasts obtained from a training set period. Note that here we have not substitute the values in the interrupted periods when the model chosen was the cyclical regression one but, when the model was the ARIMA one, we have substitute them by previous years forecasts, in accordance with the original papers where they have been proposed. This option was taken to simplify the model comparisons and because it was consider a sub parameter issue of the iterative procedure.

Regarding the comparison results between methods made before, it is important to note that it is only supported by the application to one single time series, the weekly number of deaths by pneumonia and influenza in Portugal. In our opinion, this fact is not a problem as the inconsistencies observed in the results between methods are enough to question the ideia that all of them produce equivalent results.

In the method comparison the main question is not which one will produce the best results, because we do not know what is the true number of influenza associated deaths. So we have to rely on a relative comparisons in results between methods and their associated class parameters.

In this sense, we have observed that the most sensitive parameter, i.e the one which changes have produced higher differences in the estimates, was the chosen E_a period. As should be expected the higher estimates were obtained when this period was set fixed from December to April. Intuitively one can say that this E_a period is be more sensitive and that the E_a period defined by the ISS gives a more specific estimate but also more conservative. In our opinion this parameter is the cornerstone of the epidemiological association between the occurred influenza epidemics and the time related excess of deaths observed in the mortality time series. We have further obtained a lower RMS when the E_a period was fixed, which can be explained by the fact that in this situation a great part of the remaining time series belongs to Spring and Summer where the observations variance is lower.

Considering the type of model fitted we have seen that the seasonal ARIMA model presents always lower RMS than the cyclical regression model and also non auto-correlated residuals out off the D_a periods. The main consequence of these results is a lower upper 95% confidence limit for the methods using seasonal ARIMA models. This efficiency as been also reported by other authors [12].

The baseline building process parameter also seems to have an important impact on the estimates. In this application we have seen that the estimates from the iterative procedure differ from the estimates of the non-iterative procedure essentially after the 1993-94 flu-year. From Figure 1 we can see that this is the point where the time series presents an important change in trend. Generally speaking, the iterative processes are more likely to produce estimates further apart from the observed ones than the estimates obtained by fitting the model directly to the entire time series. This is more evident where, as happens in here, the iterations are made with weekly observations for an year ahead. In fact, the first proposals of the iterative approach in the literature were not to be applied to this problem but to build baselines for surveillance purpose [12, 19], i.e. to forecast the future baseline for a to be observed period that would be the target of an surveillance.

7 CONCLUSION

In conclusion, the work described in this paper made possible to define a class of models that contains all the models, without covariates, that use interrupted time series to estimate influenza-associated deaths. The models in this class are categorized according to three main parameters: the model used to fit the interrupted time series and obtain a baseline, the chosen periods used to estimate the influenza epidemic periods and the type of procedure used to fit the model to the time series (iterative or non iterative).

This generalization led quite naturally to the construction of a set of user friendly R-routines for estimating these influenza-associated deaths with any of the methods in the class, that will be very soon available on the internet for download.

References

[1] N. J. Cox and K. Subbarao Global Epidemiology of Influenza: Past and Present *Annual Reviews of Medicine* 2000; **51**:407–421.

[2] Circular Informativa Nº35/DSCS/DPCD. Vacinação contra a gripe sazonal em 2007/2006
http://www.dgs.pt [3 June 2008].

[3] W. H. Barker, H. Borisute and C Cox A Study of the Impact of Influenza on the Functional Status of Frail Older People *Archives of Internal Medicine* 1998; **158**:645–650.

[4] Farr W. In *Vital Statistics* Office of Sanitary Institute, 1885; p 330.

[5] Lui K-J and Kendal A.P. Impact pf in°uenza epidemics on mortality in the united states from october 1972 to may 1985. *American Journal of Public Health* 1987; **77**(6):712–716.

[6] L. Simonsen, M.J. Clarke, D- Williamson, D.F. Stroup, N.H. Arden and L.B. Schonberger The impact of influenza vaccination on seasonal mor-

tality in the US elderly population. *American Journal of Public Health* 1997; **87**(12):1994–1950.

[7] L Simonsen, T.A. Reichert, C. Viboud, W.C. Blackwelder, R. J. Taylor, M.A. Miller The Impact of Influenza Epidemics on Mortality: Introducing a Severity Index *Archives of Internal Medicine* 2005; **165**:265–272.

[8] W.H.P. Zucs, U. Buchholz and H. Uphoff. Influenza associated excess mortality in germany, 1985 - 2001 *Emerging Themes in Epidemiology* 2005; **2**(6):1–9.

[9] C. Rizzo, C. Viboud, E. Montomoli, L. Simonsen, M.A. Miller Influenza-related mortality in the Italina elderly: No decline associated with the increasing vaccination coverage *Vaccine* 2006; **24**:6468–6475.

[10] J.L. Antunes, E.A. Waldman, C. Borrell, T.M. Paiva. Effectiveness of influenza vaccination and its impact on health inequalities. *International Journal of Epidemiology* 2006; **36**(6):1319–1326

[11] P. Nogueira and E. Paixão Models for mortality associated with heat-waves: update of the Portuguese heat health warning system. *International Journal of Climatology* 2007; **28** 4: 542-562.

[12] K. Choi and S.B. Thacker An evaluation of influenza mortality surveillance 1962-1979. *American Journal of Epidemiology* 1981; **113** 3: 215-216.

[13] Doshi P Are US flu death figures more PR than science? *Brithish Medical Journal* 2005; **331**:1412.

[14] L. Simonsen, R. Taylor, C. Viboud, J. Dushoff, and M. Miller US flu mortality estimates are based on solid science *Brithish Medical Journal* 2006; **332**:177-178.

[15] G. Box and G. Jenkins. *Time series analysis: Forecasting and control* (1st edn) San Francisco: Holden-Day, 1970.

[16] *Gripe Pandémica e Sazonal: Programa de Intervençao do INSA* Instituto Nacional de Saúde Dr. Ricardo Jorge, 2006.

[17] Hyndman R.J., Khandakar Y Automatic Time Series Forecasting: The forecast Package for R *Journal of Statistical Software* Julho 2008; **27**:3.

[18] C. Pelat, P-Y Boëlle, B. J. Cowling, F. Carrat, A. Flahault, S. Ansart, and A-J Valleron Online detection and quantification of epidemics *BMC Med Inform Decis Mak* 2007; **7**:29.

[19] Serfling RE Methods for Current Statistical Analysis of Excess Mortality Pneumonia-Influenza Deaths *Public Heath Reports* 1963; **78** 6: 494-506.

Name	Alias	Model	Period	Fitting procedure	T
`It_RM_F`	Sim97	Regression	Fixed period	Iterative	5
`It_RM_E`	Lui87	Regression	ISS	Iterative	5
`It_SA_F`	none	SARIMA	Fixed period	Iterative	all
`It_SA_E`	Choi81	SARIMA	ISS	Iterative	all
`RM_F`	Sim05,Riz06	Regression	Fixed period	No iterative	NA
`RM_E`	Zucs05	Regression	ISS	No iterative	NA
`SA_F`	none	SARIMA	Fixed period	No iterative	NA
`SA_E`	none	SARIMA	ISS	No iterative	NA

Table 1: Classification of the proposed methods for comparison, according to the fitting procedure (iterative or not), the model (seasonal ARIMA or multiple regression) and the E_a periods (fixed period or a period estimated by the national Influenza Surveillance Systems, ISS).

	Epidemic periods (weeks)				
flu-year	**start**	**peak**	**end**	**nº weeks**	**max incidence**
1980-81	49	3	12	16	NA
1981-82	-	-	-	-	NA
1982-83	1	2	7	7	NA
1983-84	10	11	12	3	NA
1984-85	3	3	4	2	NA
1985-86	52	3	3	4	NA
1986-87	-	-	-	-	NA
1987-88	-	-	-	-	NA
1988-89	-	-	-	-	NA
1989-90	1	3	6	6	NA
1990-91	7	9	11	5	148.4
1991-92	45	52	5	13	92.4
1992-93	6	11	14	9	117.7
1993-94	46	49	1	8	168.8
1994-95	3	5	8	6	84.1
1995-96	42	44	51	10	86.8
1996-97	47	50	8	15	111.3
1997-98	-	-	-	-	-
1998-99	51	3	8	10	252.9
1999-00	2	5	8	7	156.5
2000-01	-	-	-	-	-
2001-02	1	5	10	10	239
2002-03	48	50	50	3	76.1
2003-04	44	47	52	9	166.7

Table 2: Definition of the epidemic periods E_a for the *flu-years* 1980-81 to 2003-04 (NA: not available). Incidence values are presented by 10^5 inhabitants. From 1980-81 to 1989-90 epidemic periods were defined by the influenza death cause criterium, from 1990-91 to 2003-04 the epidemic periods were defined by Influenza Surveillance System.

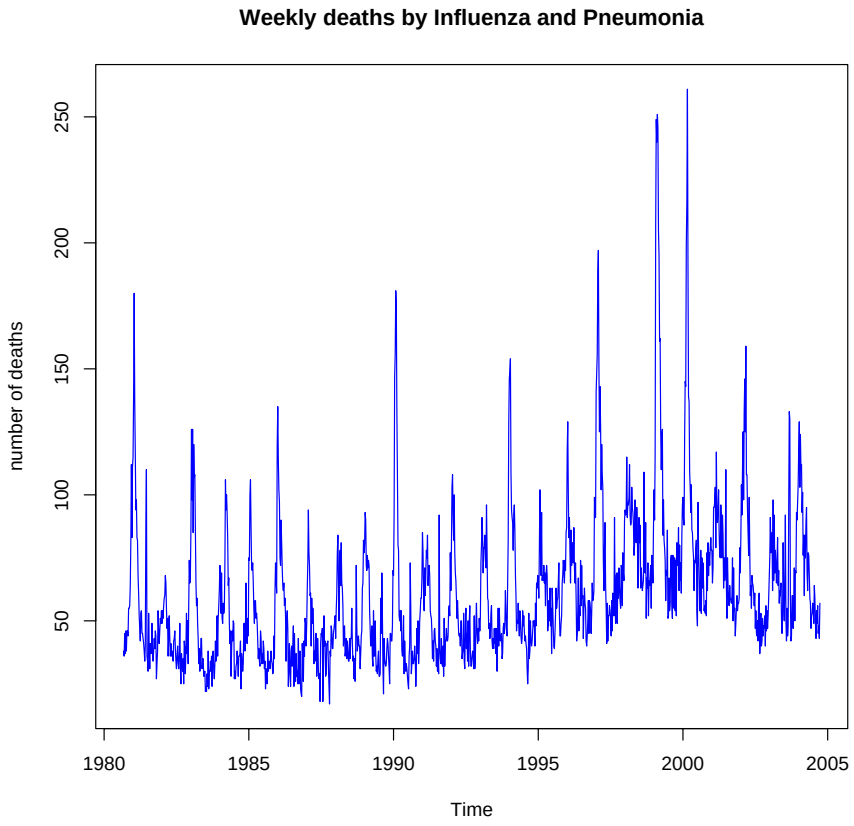

Figure 1: Weekly distribution of the number of deaths by influenza and pneumonia in Portugal from 1980-81 to 2003-04.

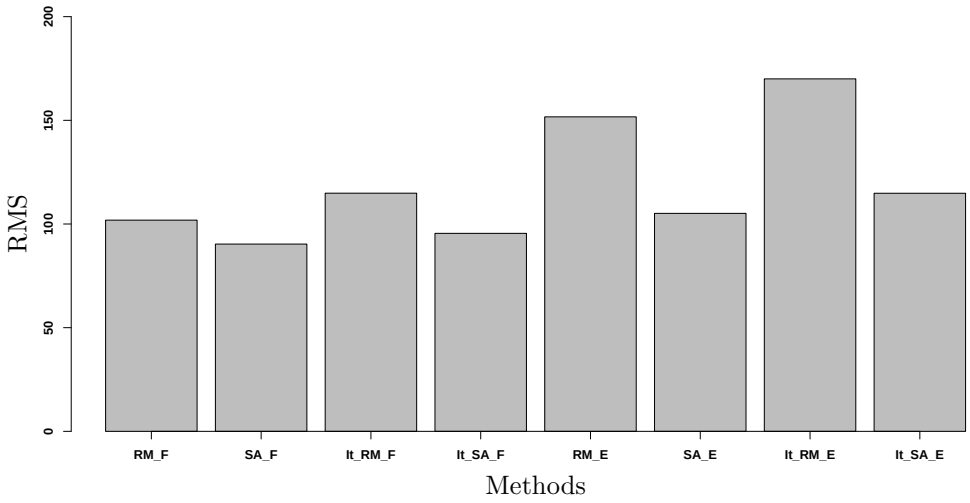

Figure 2: Residual Mean Square Errors of the studied models.

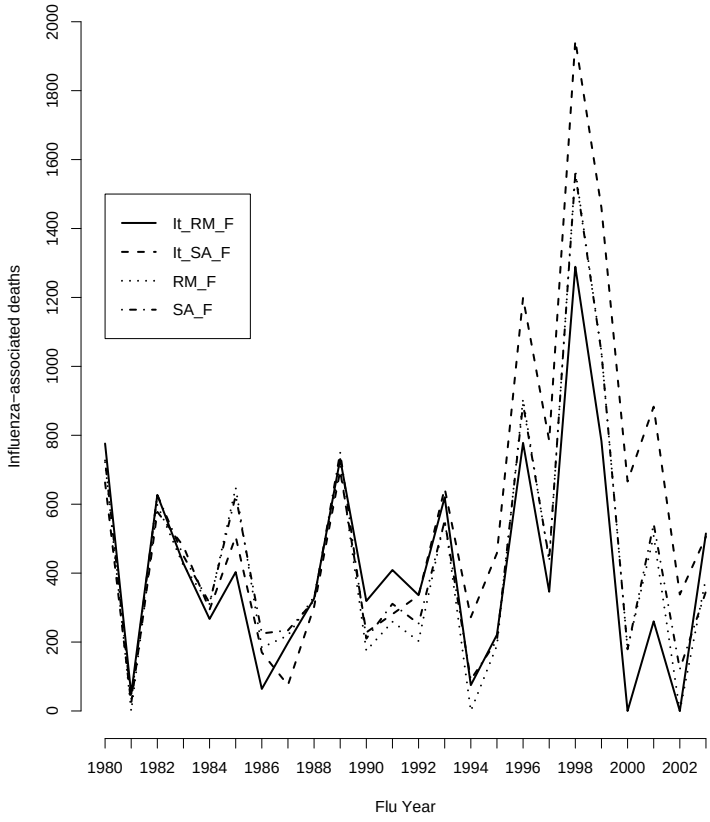

Figure 3: Estimated influenza-associated deaths from 1980-81 to 2003-2004 according to the type of method, considering E_a as fixed period from week 48 (December) to week 17 (April).

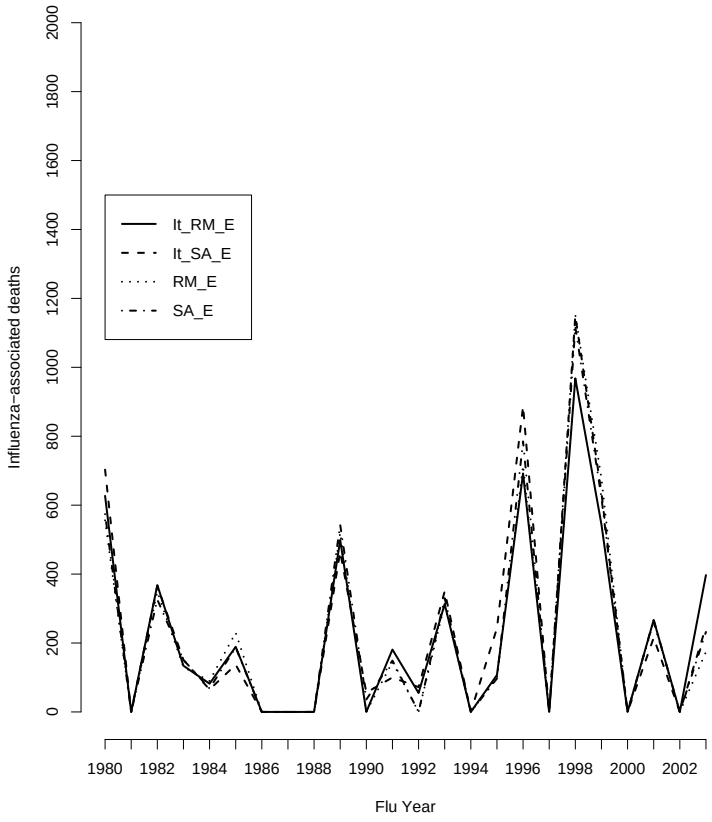

Figure 4: Estimated influenza-associated deaths from 1980-81 to 2003-2004 according to the type of method, considering E_a period as the epidemic periods previously defined.