Mem. Inst. Oswaldo Cruz, Rio de Janeiro, Vol. *87* (3): 433-436, jul./sep. 1992

FORECAST OF ACUTE RESPIRATORY INFECTIONS: EXPECTED NONEPIDEMIC MORBIDITY IN CUBA

ARMANDO AGUIRRE & EDILBERTO GONZALEZ

Division de Epidemiologia, Instituto Pedro Kouri, Apartado Postal 11500, La Habana, Cuba

A forecast of nonepidemic morbidity due to acute respiratory infections were carry out by using time series analysis. The data consisted of the weekly reports of medical patient consultation from ambulatory facilities from the whole country. A version of regression model was fitted to the data. Using this approach, we were able to detect the starting data of the epidemic under routine surveillance conditions for various age groups. It will be necessary to improve the data reporting system in order to introduce these procedures at the local health center level, as well as on the provincial level.

Key words: ARI time series analysis – ARI endemicity forecasting

The concept of excess mortality to detect underlying morbidity was initially explored by William Farr in 1842. He defined excess mortality as *"the number of deaths that exceed the usual expected for this cause, for a given period, for a determined community"* (Bailey, 1975). Farr was unable to specify the method for determining the number of deaths which should be expected as *usual* or *normal*, i. e. nonepidemic. In 1932, S. D. Collins defined the number of expected deaths, as the average number of weekly deaths during the nonepidemic years for one specific week (Collins, 1932). Then in 1962, T. C. Eickoff introduced a regression equation for the estimate of the usual expected number of deaths, upon weekly mortality rates, in which the data corresponding to epidemic weeks were omitted (Eickoff et al., 1962). In 1963, R. E. Serfling applied the regression method to the weekly mortality rates obtaining the value around which these mortality figures could fluctuate for each week. He defined an upper threshold limit of 1.64 standard deviations (σ) above the expected, which is the maximum permissible rate that the reported mortality figure could take and still be considered *not to be in excess* (Serfling, 1963). In 1974, J. Housworth and A. Langmuir used the percentage of excess, or the ratio of the excess and expected rate, as a measure of the intensity of an epidemic (Houstworth & Langmuir, 1974). In 1979, R. E. Clidford included in the calculation of normal mortality the time intervals between significant influenza viral drift and isolation, the secular trends and the temperature fluctuations associated with seasonal variations (Clifford et al., 1977). This resulted in a very complex method which would be difficult to implement for practical routine applications.

In 1981, K. Choi and S. B. Thacker proposed a method for the calculation of the weekly expected nonepidemic mortality (Choi & Thacker, 1981), which is based on a modification of the Auto Regressive Integrated Moving Average (ARIMA) model developed by Box and Jenkins (Box & Jenkins, 1976). This method was adopted by the Centers for Disease Control in the United States, and also has been recommended by the Expert Commission for Influenza Surveillance of the Council for Mutual Economic Assistance (Epid. Surv., 1987). In their paper Choi and Thacker (Choi & Thacker, 1981) established a comparison between the ARIMA and Serfling models with the error criterion *Relative error of forecast = 100x (observed deaths – expected deaths)/(observed deaths)* for nonepidemic weeks between 1964 and 1978. The forecasting error of greather than the range percentage error ±10 for Serfling model is 42.5%, but for ARIMA is 33.3% and not 31% described in the paper.

In the surveillance unit at the Pedro Kouri Institute of Tropical Medicine, a modification

Received 22 July 1991.
Accepted 22 April 1992.

of Choi and Thacker model was adopted for the forecast of the expected nonepidemic morbidity (number of outpatient medical consultations) from acute respiratory infections (Aguirre et al., 1988).

MATERIALS AND METHODS

The weekly time series of the number of medical consultations of patients with influenza-like illness attending outpatient facilities for the whole country were used as an expression of the reflection of morbidity for this syndrome (Gonzalez et al., 1988). This information is reported for the following age groups: less than 1 year, from 1 to 4, 5 to 14, 15 to 64 and 65 years and over from the 27th week of 1984, to the 26th week of 1988.

In order to determine the epidemic values in the time series via computer, the age group series were ordered nondecreasing for each week $\{x_{[1]}, x_{[2]}, \ldots x_{[N]}\}$, where $x_{[i]}$ is the rate of consultation and $i = 1, 2, \ldots, N$ is the new ordinal number for the rate $\hat{x}_t$ in ordered serie. The $x_{[i]}$ values upper the 95 per centil were considered abnormal (Fisher, 1963; Aguirre et al., 1986). They were replaced in the original serie by the expected rates of nonepidemic outpatient consultation $\hat{x}_t$ which were calculated from the Fourier filter serie

$$\hat{x}_t = a + bt + \sum_{i=1}^{3} (c_i \sin \frac{2i\pi t}{T} + d_i \cos \frac{2i\pi t}{T}) + e_t$$

where, $T = 52$ is the series periodicity, e_t is the random error and $(a, b, c_i, d_i) \in R$ are parameters estimated by the least square method for each serie (Bloomfield, 1984; Moolgabkar & Venzon, 1987).

An ARIMA (3, 0, 0) x $(1, 1, 0)_{52}$ model was fitted to these transforming nonepidemic series using a similar model proposed by Choi and Thacker (Choi & Thacker, 1981): $(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - \Phi_1 B^{52})(1 - B^{52}) x_t = \alpha_t$, where, x_t is the value of the time series at time t, α_t is white noise series, B is a backward operator, defined as $Bx_t = x_{t-1}$, ϕ_i and Φ_i are parameters, where $(\phi_1, \phi_2, \phi_3) \in R$, and $\Phi_1 \in R$. Over the forecasted endemic values, an upper threshold of maximum permissible fluctuation for the rate of consultation has been established for 95% significant level (Anderson, 1971). An elevation above the epidemic threshold for two or more weeks indicated the evidence of high rise in epidemic outbreak using the Serfling's criterion (Serfling, 1963).

RESULTS

The numerical model obtained (Faddeev & Faddeeva, 1985) was implemented on an IBM-XT personal computer. The time consumption for typical calculation was about four minutes. The curves of expected nonepidemic numbers of medical consultations due to acute respiratory infections were obtained for the epidemiologic season between the 27th week of 1988 and the 26th week of 1989, by age groups.

These curves provide the weekly values around which the actual number of medical outpatient consultations tend to fluctuate during the season. They include the upper threshold for a nonepidemic week. Values exceeding this threshold are considered to be excess morbidity and indicate an elevation of epidemic outbreak rise.

We can get the predicted values of the overall number of medical consultations in the absence of an epidemic situation by using the weekly predicted fluctuation. Figures 1-3 provide the curves on the expected and observed values for the whole population and for the age groups from 1 to 4 years and from 5 to 14 years.

During the begining of the 1988-1989 epidemiologic season for acute respiratory infections, a moderate epidemic rise was observed between the weeks 29th, ending on July 20th and the week 39th ending on September 30th for whole population (Fig. 1). However, this excess morbidity is not apparente in the age group from 1 to 4 year old as revealed the analysis (Fig. 2). Similar situation also was revealed in the age groups below 1 year, from 15 to 64 years, and 65 and over years. By the other hand, in the age group from 5 to 14 year old it was noticed an epidemic wave from the week 36th ending on September 7th to the week 41th ending on October 14th (Fig. 3). Just after the begining of the school year it was produced an epidemic increase in the school children population, mostly caused by influenza $A(H_1N_1)$ viruses as revealed by the laboratory evidences. It is clear that differences between the behavoir of the distint age group can be pointed out by the monitoring of the population by the time serie analysis combined with the laboratory results, and the corresponding ethiological hypotesis can be sugested and confirmed or not in terms of the laboratory evidences under current conditions.

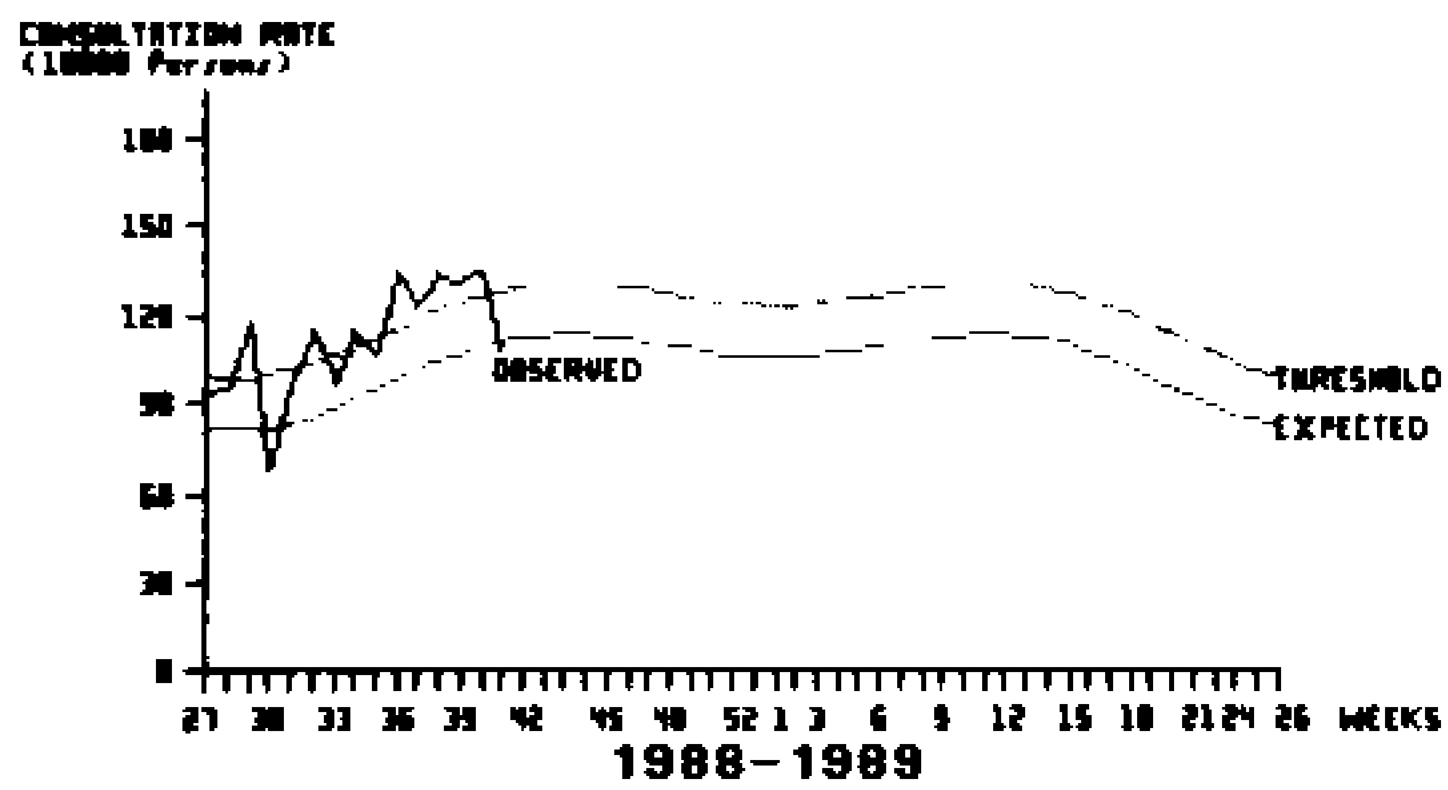

Fig. 1: Cuba nonepidemic expected and observed outpatient medical consultations rates from acute respiratory infections of whole populations.

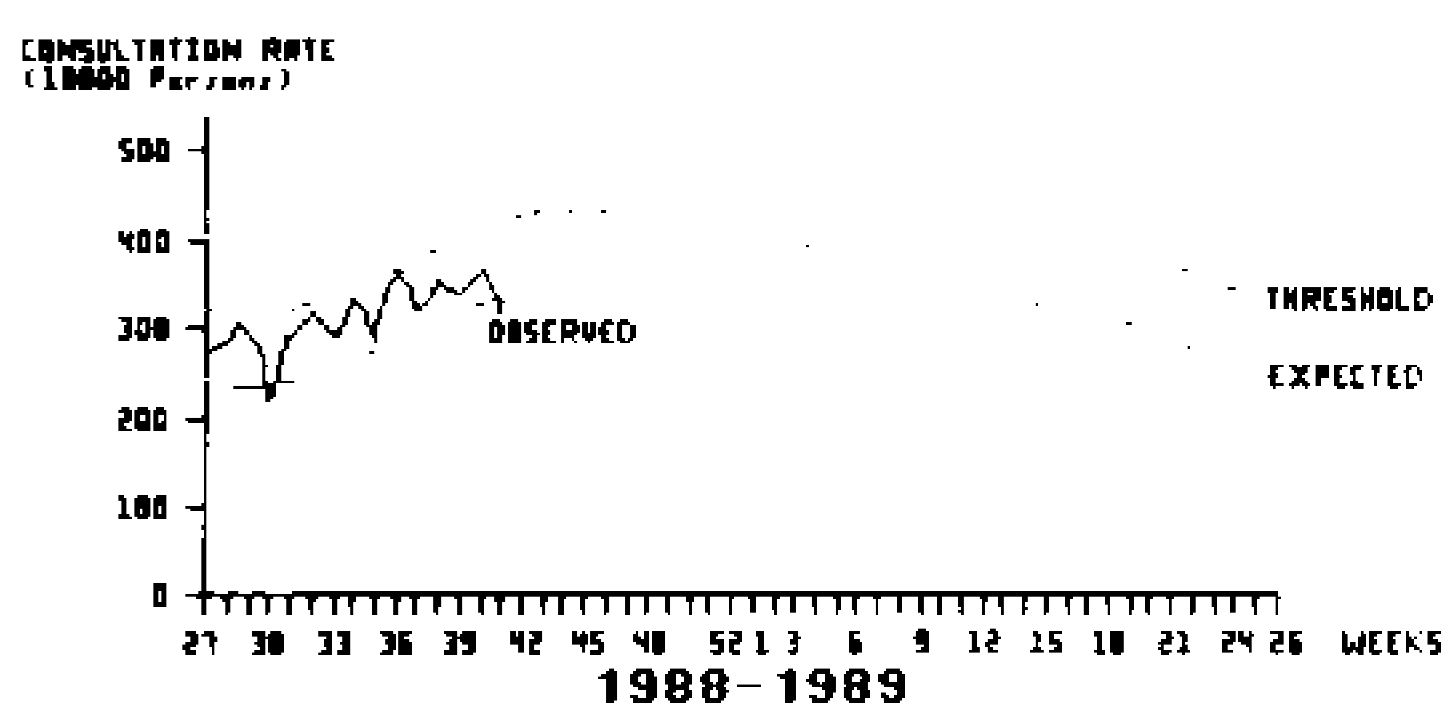

Fig. 2: Cuba nonepidemic expected and observed outpatient medical consultations rates from acute respiratory infections of children from 1 to 4 years old.

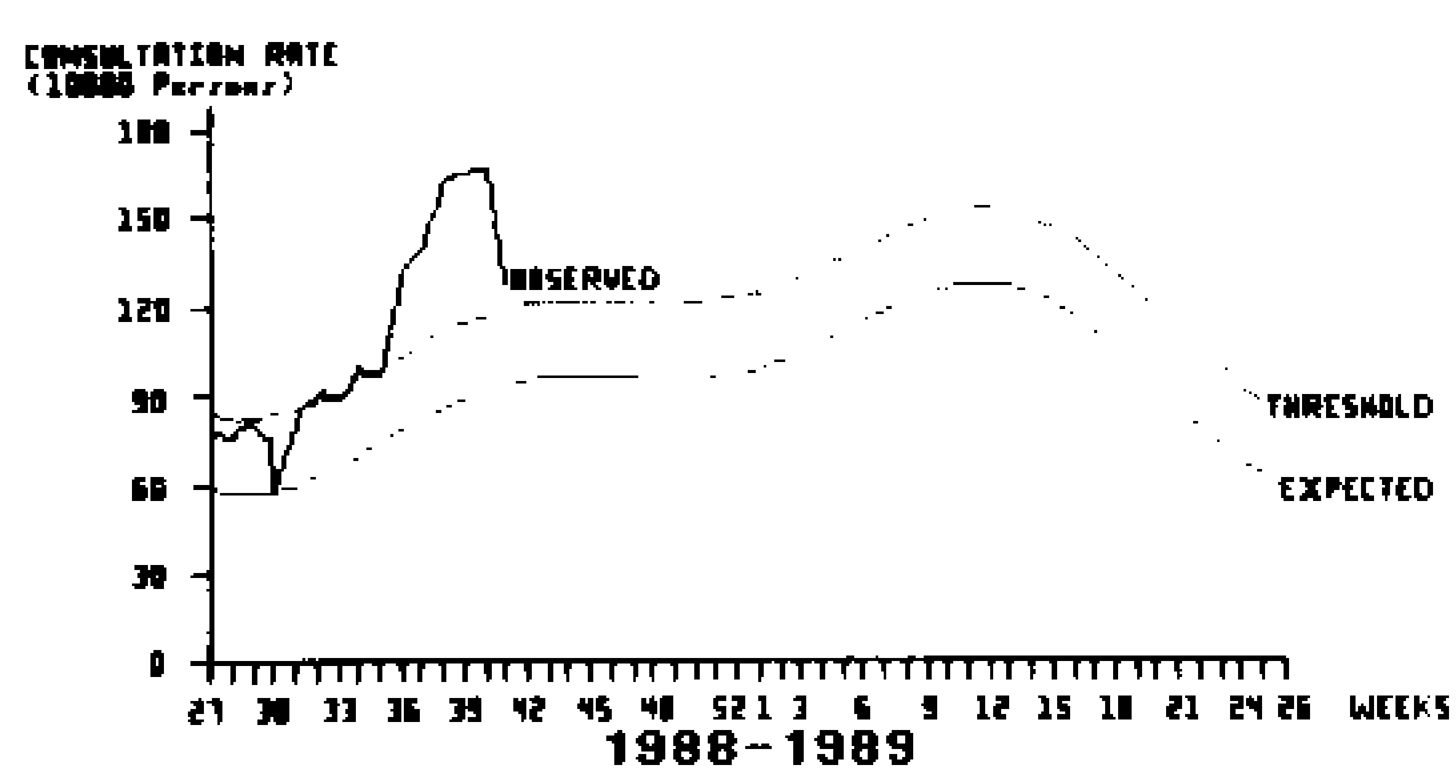

Fig. 3: Cuba nonepidemic expected and observed outpatient medical consultations rates from acute respiratory infections of children from 5 to 14 years old.

DISCUSSION AND CONCLUSIONS

The main contribution of our approach consists of the objetive discrimination of the abherrant values of the series and its replacement for the expected endemic values, work previous to any and all treatment of time series carried out by the researchers, and in our case, this is achieved with the help of the personal computers available in the local systems, as well as in the central and provincial centers of epidemiological surveillance. Such an aproach avoids subjetive interpretations and the evaluation opinions in the consideration of anomalous values of notification and its treatment as well.

Public health services delivered on a very wide range of national coverage and depending on an unique service with close cooperation between different centers of the local, provincial and central level, enable the application of these techniques to routine surveillance. However, a number of requirements are needed for the best operation of such a system. It is important that weekly data be reported in a timely fashion. On the other hand, there is an actual trend to decentralize the responsibility for preventive interventions in the primary health services in close cooperation with the epidemiologic antiepidemic units of the secondary health care services. The availability of personal computers makes it possible to introduce this tool at the local level for analyzing epidemiological data in order to carry out more accurate surveillance tasks, not only in the provincial centers, but also in the municipal and polyclinic facilities.

Presently, provincial centers of epidemiology are equiped with personal computers and they are creating data bases for different health care projects. The training of technical staff and the supply of suitable software will increase the efficiency of surveillance for acute respiratory infection epidemics, as well as for other communicable diseases where a rapid administrative response is often needed.

The application of this time series model to data of medical outpatient consultations for acute respiratory infections has provided estimates of the expected nonepidemic (endemic) weekly morbidity. In addition, the model has been used to detect the possible rise of an epidemic situation concerning the evaluation of excess morbidity figures and laboratory results. Under routine surveillance conditions, the current data transmission system must be improved to assure the timely assessment of the beginning of an epidemic.

The follow-up of fluctuation in the ARI consultation rates by aged groups, allows us to discover, with the help of the software package designed*, the specific affected groups, as shown in the example, apart from detecting the propagation dynamics of the process among those groups. A second step, already achieved in this work is the use of this approach, up to the provincial level, which allows in addition to identifying the affected age groups to point out the places where this occurs.

Another aspect of interest is that the amount of data available, when considering the unit of screening as the week is 208 data for the forecast test, with a yearly increase of 52 data by season for each group. The residual distributions obtained are satisfactory adjusted to teoretical configurations of white noise for coeficients of simple and partial autocorrelations, and the residuals behavoir as incorrelated zero-mean and constant-variance random distributions, but regularity do not follow a gaussian distribution. A sample to show the increase of available observations to oustand the forecasts, with the passing of time, we think the residual distributions could be near normal curves. On the other hand the forecats obtained for further seasons have not brougth remarkable changes in regards to previous forecasts, considering the growing trend removal, which is in favor of a clear stability in the models adjusted to the data.

ACKNOWLEDGEMENT

To Ira M. Longini Jr., for his help in the preparation of this paper.

REFERENCES

AGUIRRE, A.; GONZALEZ, E. & BURGASOV, Y. I., 1986. *Los niveles de control de las enfermedades infecciosas*. IPK, La Habana.

AGUIRRE, A.; GONZALEZ, E. & BURGASOV, Y. I., 1988. El sistema automatizado de vigilancia epidemiologica SAVE. *Rev. Cub. Med. Trop., 40:* 11-23.

ANDERSON, T. W., 1971. *The statistical analysis of time series*. John Wiley, New York.

BAILEY, N. T. J., 1975. *The mathematical theory of infectious diseases and its applications*. Griffin, New York.

BLOOMFIELD, P., 1984. *Fourier analysis of time series: an introduction*. John Wiley, New York.

BOX, G. E. P. & JENKINS, G. M., 1976. *Time series analysis: forecasting and control*. Holden Day, San Francisco.

CHOI, K. & THACKER, S. B., 1981. Time series forecast on expected pneumonia and influenza deaths. *Amer. J. Epidem., 113:* 215-226.

CLIFFORD, R. E.; SMITH, J. W. G. & TILLETT, H. E., 1977. Excess of mortality associated with influenza in England and Wales. *Inter. J. Epid., 6:* 115-118.

COLLINS, S. D., 1932. Excess mortality from causes other than influenza and pneumonia during influenza epidemics. *Pub. Health Rep., 47:* 2159-2180.

EICKOFF, T. C.; SHERMAN, I. L. & SERFLING, I. L. & SERFLING, R. E., 1962. Observation on excess mortality associated with epidemic influenza. *JAMA, 176:* 776-782.

Epidemiological Surveillance of Influenza in the Countries. Members of the Council for Mutual Economical Assistance, 1987. Permanent Comission for Cooperation in Public Health, Moscow (in Russian).

FADDEEV, D. K. & FADDEEVA, V. N., 1985. *Computational methods of Linear Algebra*. Mir, Moscow.

FISHER, R. A., 1963. *Statistical methods for research workers*. 6th ed. Oliver & Boyd, Edimburg.

GONZALEZ, E. & GOYENECHEA, A., 1988. La vigilancia epidemiologica de la influenza y otras infecciones respiratorias agudas en Cuba. *Rev. Cub. Med. Trop., 40:* 35-41.

HOUSTWORTH, J. & LANGMUIR, A. D., 1974. Excess mortality from epidemic influenza 1957-1966. *Am. J. Epid., 100:* 40-56.

MOOLGABKAR, S. & VENZON, D., 1987. General relative risk regression models for epidemiologic studies. *Am. J. Epid., 126:* 949-961.

SERFLING, R. E., 1963. Methods for current statistical analysis of excess of pneumonia and influenza deaths. *Pub. Health R., 78:* 494-506.

*This software package will be available, please send the request to main author adress.