Mem. Inst. Oswaldo Cruz, Rio de Janeiro, Vol. *87* (3): 429-432, jul./sep. 1992

THE FEASIBILITY OF FORECASTING INFLUENZA EPIDEMICS IN CUBA

ARMANDO AGUIRRE & EDILBERTO GONZALEZ

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

A large influenza epidemic took place in Havana during the winter of 1988. The epidemiologic surveillance unit of the Pedro Kouri Institute of Tropical Medicine detected the begining of the epidemic wave. The Rvachev-Baroyan mathematical model of the geographic spread of an epidemic was used to forecast this epidemic under routine conditions of the public health system. The expected number of individuals who would attend outpatient services, because of influenza-like illness, was calculated and communicated to the health authorities within enough time to permit the introduction of available control measures. The approximate date of the epidemic peak, the daily expected number of individuals attending medical services, and the approximate time of the end of the epidemic wave were estimated. The prediction error was 12%. The model was sufficienty accurate to warrant its use as a practical forecasting tool in the Cuban public health system.

Key words: influenza epidemic forecasting model – ARI morbidity forecasting

In 1971, the Rvachev-Baroyan simulation and forecasting mathematical model for influenza epidemics was tested in the National Institute of Influenza Research of the Soviet Union. The experiment was carried out under routine conditions during a real influenza epidemic. The model performed quiet well, and, therefore, it was officially adopted in the routine work of the Epidemiological Surveillance Network of the Soviet Public Health System (Baroyan & Rvachev, 1967, 1968, 1969, 1970, 1971, 1972, 1973; Rvachev, 1967, 1968a, b, 1971, 1972; Rvachev & Longini, 1985).

This model represents one of the first large-scale applications of mathematical epidemic theory, a process initiated by D. G. Bernoulli on April 30th 1760 (Sakino, 1962; Bailey, 1975).

The model has two parts: (i) the local model, for the epidemic forecast at the urban center level, and (ii) the general model, for the epidemic forecast on the level of a whole territory. Thus, the model uses a global approach for forecasting epidemics (Longini et al., 1986). In 1987 a stochastic version of the model has been developed by A. Flahault, of the Biomathematical Group of the National Institute of Medical Research at the University of Paris, with very good results (Flahault, 1986, 1987). Some variations of the original model are being integrated into the public health surveillance systems of Bulgaria (CMEA, 1987) and Mexico (Rodriguez, 1985).

We have carried out a preliminary study in Cuba to determinate the baseline parameters for the model. These include the following: (i) the cofficients of attendance for visits to the ambulatory health care centers due to influenza-like illness. These coefficients are used to distribute the weekly attendance accross days, e. g., attendance tends to be low on Friday but high on Monday, (ii) the infectious period distribution, and (iii) the relationship between the etiologic agent, e. g., Influenza A(H_3N_2), and the time of the epidemic peak for an epidemic in a typical Cuban population (Aguirre et al., 1988, 1989a, b).

METHODS AND DATA

The Rvachev-Baroyan local model is reduced to the solution of the Cauchy problem for the following integro-differential equation:

Received 22 July 1991.
Accepted 22 April 1992.

$$\frac{\partial x}{\partial t} = -\frac{\lambda}{p}\, x(t) \int_{D}^{T} y(\tau, t)\, g(\tau) dt,$$

$$\frac{\partial y}{\partial t} + \frac{\partial y}{\partial \tau} = O,$$

$$y(O, t) = \frac{\lambda}{p}\, x(t) \int_{D}^{T} y(\tau, t)\, g(\tau)\, d\tau,$$

$$x(O) = \alpha p, y(\tau, O) = a(\tau), \tau = 1, 2, 3, \ldots,$$

where, x(t) is the number of susceptibles, y(τ, t) is the number of infected persons who have been infected for τ time units, g(τ) is the infectious period distribution, p is the population size, λ is the infectious contact rate, α is the initial proportion susceptible, T is the maximum length of the infectious period, and a(τ) is the number of reported cases (individuals attending medical consultations) during the first days of the outbreak.

The discrete-time counterpart to the above model is governed by the following equation (Baroyan & Rvachev, 1973; Rvachev & Longini, 1985):

$$y(O, t+1) = \frac{\lambda}{p}\, x(t) \sum_{\tau = D}^{T} y(\tau, t)\, g(\tau)$$

This discrete-time model was implemented in an IBM-XT personal computer*. The time consumption for a typical calculation is about 20 minutes. In order to estimate the free parameters α and λ we make the assumption that the ratio of new infectives in succesive time units is a constant, K, i. e.,

$$\frac{y(t+1)}{y(t)} = K$$

for t small. It follows directly from the above equation that

$$\lambda = \frac{K^{T+1}}{\alpha \sum_{\tau = D}^{T} g(\tau) K^{T-\tau}}$$

(Baroyan & Rvachev, 1973; Rvachev & Longini, 1985). The initial values of the free parameters must be contained in $\alpha\, \varepsilon[0.60, 0.90]$ and $k\varepsilon[1.05, 1.65]$ which is based on the historical influenza data (Baroyan & Rvachev, 1973; Rvachev & Longini, 1985). We then apply a previously developed statistical procedure to find the estimates $\hat{\alpha}$ and $\hat{\lambda}$. For the forecasting of the epidemic curves we used the number of reported cases (patients attending medical consultations) during the first seven days after detection of the outbreak. Then the remainder of the epidemic is predicted.

RESULTS

The surveillance unit detected a rise in reported cases of acute respiratory infecitons during the 6th week of 1988, and an influenza epidemic was confirmed in the following week ending on February 17th. Laboratory tests provided evidence of etiologic activity of influenza $A(H_3N_2)$ virus in both school children and adults.

The application of the model based on the information obtained from the overall figures reported by the surveillance unit, provided the number of expected individuals attending the outpatient health centers or services. Then the monitoring of the daily observed cases reported by these services was carried out as usual. The values used in the model were p = 2,027,800, T = 8,g(τ) = {0.00,1.00,0.90,0.55,0.30, 0.15,0.05,0.00} which is an average of those values. The coefficient of attendance m(w) = {0.933,0.980,1.00,0.903,0.894,0.760,0.600} where, w = Monday, Tuesday, . . . , Sunday were obtained statistically by comparing the attendance of each day of the week to attendance on Wednesday for a number of influenza epidemics in Havana, e. g., attendance on Sunday is 60% of attendance on Wednesday. The estimated values for the free parameters were $\hat{\alpha} = 0.75$ and $\hat{\lambda} = 1.42$.

The results are shown in Table and Figure. The values were calculated from February 14th. The expected peak was set for March 15th and it appeared on March 1st. This diference may be explained by the bias in the empirical estimate of the expected time of the peak of the outbreak. The final decline of the epidemic wave was predicted to occur on April 17th which was also observed to be the final day. The error of predicted values was calculated by the formula:

---

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

$$\text{error} = \frac{\sigma}{a_{max}(t)} \times 100\%$$

where, a_{max} (t) is the maximum value of observed morbidity,

$$\sigma = \left(\frac{\sum_{t=1}^{L} [a(t) - y(0,t)]^2}{L} \right)^{\frac{1}{2}},$$

i. e., the mean squared deviation, and L is the number of observations. This error was estimated to be 12%.

TABLE

Expected and observed daily reported cases of acute respiratory infections during the influenza epidemic in Havana, 1988

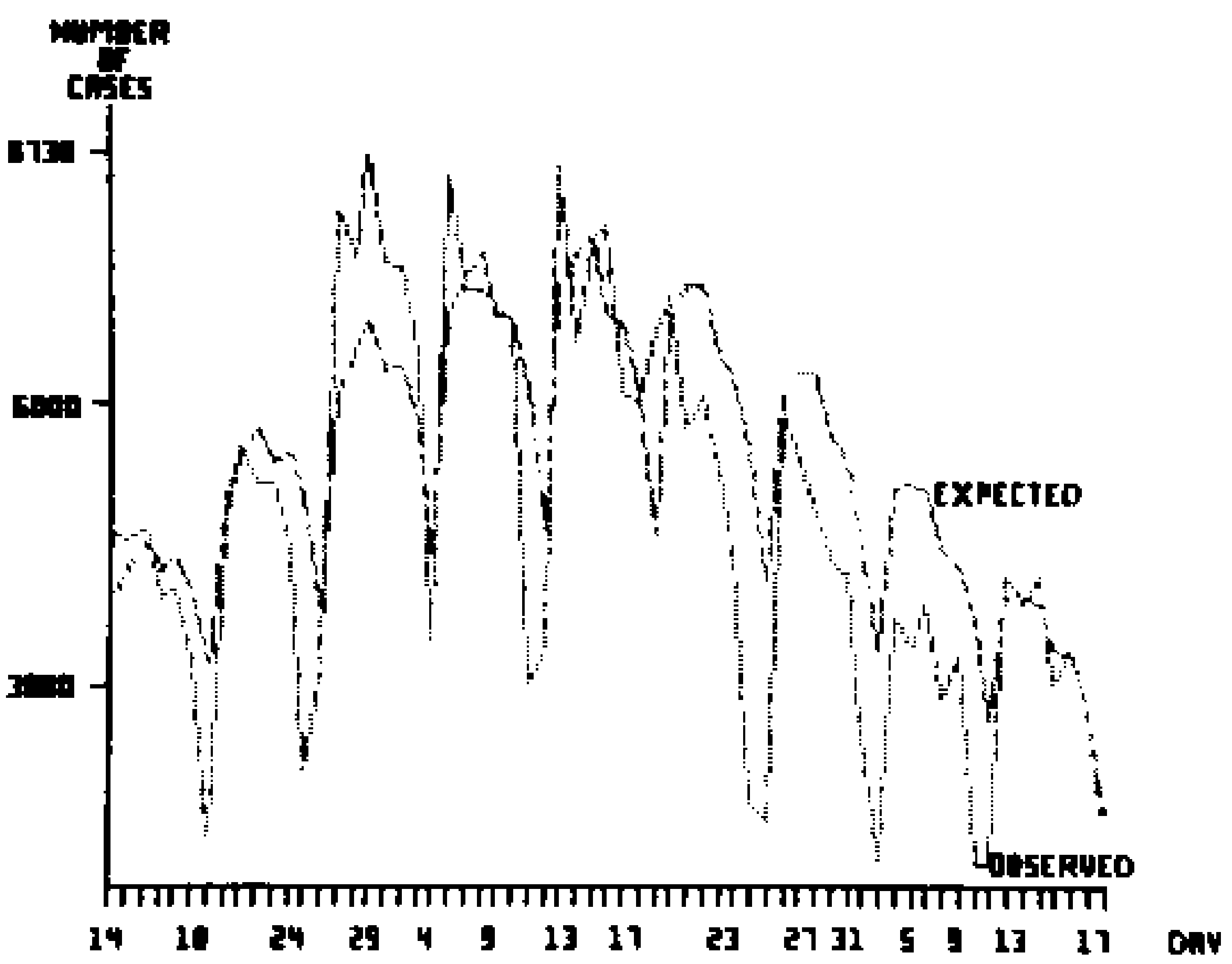
Day	Expected[a]	Observed[b]	Day	Expected[a]	Observed[b]
Feb 14	3910	4694	Mar 17	6904	6137
Feb 15	4272	4596	Mar 18	6056	6070
Feb 16	4531	4669	Mar 19	4558	4677
Feb 17	4248	3935	Mar 20	7011	7202
Feb 18	4362	4044	Mar 21	7273	5765
Feb 19	3993	3224	Mar 22	7317	6104
Feb 20	3138	1368[c]	Mar 23	6504	5277
Feb 21	5045	4844	Mar 24	6327	4462
Feb 22	5472	5539	Mar 25	5486	1768
Feb 23	5758	5188	Mar 26	4081	1560
Feb 24	5355	5166	Mar 27	6208	5912
Feb 25	5453	4480	Mar 28	6368	5513
Feb 26	4950	2118	Mar 29	6337	4944
Feb 27	3856	3189	Mar 30	5572	4831
Feb 28	6143	8088	Mar 31	5365	4250
Feb 29	6600	7591	Apr 1	4603	4514
Mar 1	6879	8738	Apr 2	3391	1181
Mar 2	6335	7560	Apr 3	5107	3756
Mar 3	6386	7503	Apr 4	5189	3400
Mar 4	5737	6516	Apr 5	5116	3886
Mar 5	4422	3498	Apr 6	4459	3887
Mar 6	6969	8469	Apr 7	4255	3367
Mar 7	7406	7260	Apr 8	3621	1098
Mar 8	7632	7244	Apr 9	2645	1114
Mar 9	6949	7029	Apr 10	3953	4173
Mar 10	6925	6960	Apr 11	3987	3896
Mar 11	6148	3056	Apr 12	3902	4207
Mar 12	4684	3508	Apr 13	3376	3041
Mar 13	7293	8660	Apr 14	3287	3386
Mar 14	7658	6711	Apr 15	2890	2763
Mar 15	7797	7800	Apr 16	1567	1327
Mar 16	7013	7930	Apr 17	850	942

[a] forecasted cases by the Rvachev-Baroyan model.

[b] reported cases

[c] only the number of cases reported in the first seven days were used to estimated the forecast (remainder).

The shape of the epidemic curve shows the daily variation in attendance with on minimum on Sunday when there are no outpatient services, however, there are still reports to the duty services.



Influenza epidemic forecast, Havana, February 15th to April 17th 1988. The forecasting error was 12%.

Samples comparison based on signs tests for expected and observed values give Z = 2.125 for large sample test statistic and p = 0.033 for two-tailed probability of equaling or exceeding this Z. Samples comparison based on the Wilcoxon's pairs and ranks test for expected and observed values give Z = –1.479 for large sample test statistic and p = 0.138 for two tailed probability of equaling or exceeding this Z. The Pearson Correlation Coefficient between expected and observed values obtained was 0.839496 for average removal and R-squared was equal to 0.70.

DISCUSSION AND CONCLUSIONS

The National Epidemiological Surveillance System links the primary care services, from which a data record and reporting subsystem is operated through the network of municipal and provincial health centers, to the top health units of the Pedro Kouri Institute of Tropical Medicine. Data collection and reporting is standardized by national guidelines, and they are stable and consistent enough to provide good quality routine surveillance. In addition the provincial centers of hygiene and epidemiology are equiped with personal computers that make it possible to manage the data in a rapid and accurate way.

Another important factor for introducing improved computerized methods of surveillance and control, with some mathematical aspects, is the increased number of professional staff in the municipal and provincial levels who are being trained in epidemiology and other related subjects, required for operating components of the surveillance models.

Since it has been possible to obtain suitable information on daily attendance to the health services of Havana, and then, to apply the local influenza forecasting model in a satisfactory way it seem feasible to incorportate the model as a routine part of the surveillance system.

Nevertheless, we will continue to carry out experiments to improve the technical aspects of model use. Personnel must be trained in specific topics concerning mathematical models. In addition, information provided by local health departments will probably have to be reorganized. It may only be necessary to make the health centers' staff at municipal and local levels aware of the possibility of using this new tool in order to get the most effective cooperation for rapid transmission of data with maximum quality and timelyness.

The number of individuals who would attend the health centers of Havana because of influenza-like illness during the epidemic period in February-April, 1988, was estimated in advance and submitted to the health authorities in time to permit the organization of general and specific control measures. This is the first time that such a model has been applied to actual influenza epidemic in Cuba. We plan to introduce the model as a routine component of the surveillance system at no additional cost.

ACKNOWLEDGEMENT

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

REFERENCES

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

AGUIRRE, A.; GONZALEZ, E. & BURGASOV, Y., 1989a. Un modelo para el pronostic de epidemias de gripe. 1a parte. *Rev. Cub. Hig. Epid., 27:* 159-166.

AGUIRRE, A.; GONZALEZ, E. & BURGASOV, Y., 1989b. Un modelo para el pronostic de epidemias de gripe. 2a parte. *Rev. Cub. Hig. Epid., 27:* 215-221.

BAILEY, N. T. J., 1975. *The mathematical theory of infectious diseases and its applications.* 2nd ed., Griffin, New York.

BAROYAN, O. V. & RVACHEV, L. A., 1967. Deterministic epidemic models for a territory with a transport network. *Kibernetica, 3:* 67-74.

BAROYAN, O. V. & RVACHEV, L. A., 1968. Some epidemiological experiments carried out with a transport network. *J. USSR Sc. Acad., 23:* 32-34.

BAROYAN, O. V. & RVACHEV, L. A., 1969. An attempt at large-scale influenza epidemic modelling by means of a computer. *Bul. Int. Epid. Assoc., 18:* 22-31.

BAROYAN, O. V. & RVACHEV, L. A., 1970. Computer modelling of influenza epidemics for large-scale systems of sities and territories. Proc. WHO Symposium on Quantitative Epidemiology, Moscow.

BAROYAN, O. V. & RVACHEV, L. A., 1971. Computer modelling of influenza epidemics for the whole country USSR. *Adv. App. Prob., 3:* 224-226.

BAROYAN, O. V. & RVACHEV, L. A., 1972. Projects of machine modelling of influenza epidemics for the territory of the USSR. *J. Virol., Microb. Immunol., 49:* 3-11.

BAROYAN, O. V. & RVACHEV, L. A., 1973. Mathematical and computer modelling of influenza epidemics in the USSR. *J. USSR Sc. Acad., 28:* 26-30.

CMEA, Council for Mutual Economic Assitance, 1987. Permanent Comission for Cooperation in Public Health. Epidemiological Surveillance and Epidemic Forecasting of Influenza. Moscow (in Russian).

FLAHAULTH, A., 1986. Modelisation de la distribution a l'attaque des maladies transmissibles: application a la grippe. DEA Biomath., octobre.

FLAHAULTH, A., 1987. Adaptation du modele de Rvachev-Baroyan a la simulation d'une epidemie de grippe en France. Revue INSERM, june.

LONGINI, I. M.; FINE, P. E. & THACKER, SB., 1986. Predicting the global spread of new infectious agents. *Am. J. Epid., 134:* 383-391.

RODRIGUES, A. E., 1985. *Un modelo de pronostico de epidemias de gripe para territorias grandes.* PhD Thesis, UNAM, Ciudad Mexico.

RVACHEV, L. A., 1967. A model for the connection between processes in the organism and the structure of epidemics. *Kibernetica, 3:* 75-78.

RVACHEV, L. A., 1968a. A modeling of the connection between epidemiological and infectious processes. *J. USSR Sc. Acad., 23:* 34-37.

RVACHEV, L. A., 1968b. Computer modelling experiment on large scale epidemics. *J. USSR Sc. Acad., 180:* 294-296.

RVACHEV, L. A., 1971. A computer experiment for predicting an influenza epidemic. *J. USSR Sc. Acad., 198:* 68-70.

RVACHEV, L. A., 1972. Modelling medico-biological processes in a community in terms of the dynamics of continous media. *J. USSR Sc. Acad., 203:* 540-542.

RVACHEV, L. A. & LONGINI, I. M., 1985. A mathematical model for the global spread of influenza. *Math. Biosc., 75:* 3-22.

SAKINO, S., 1962. Historical development of epidemic models. *Proc. Inst. Stat. Math., 9:* 127-137.