MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-447

Attempting to predict the fate of an ongoing epidemic. Lessons from A(H1N1) influenza in USA

José Luis Hernández Cáceres

Center for Cybernetics Applications to Medicine, Havana Higher Institute of Medical Sciences, Cubanacan, Playa, Ciudad de La Habana, Cuba

Abstract

A simple method is proposed for predicting the fate of an epidemic outburst from early data. The method is based on the Richards model, and linearizations are proposed for obtaining preliminary values. A second step with nonlinear estimation fed with preliminary values as initial guess values may be attempted if field conditions allow the computation. The method was tested on data from 2001 dengue outbursts in both Havana and Winward Islands (French Polynesia). Predictions were satisfactory and an attempt of true prediction based on daily data for the 2009 H1N1 influenza outburst in the USA was undertaken. Comparison of early predictions with actual values obtained 3 months later suggests that some of the discrepancies are not due to method's inacuracy, but to real improvement of infection rate as the H1N1 outburst proceeded. The method can be applied in any setting where cumulative number of cases is properly recorded.

Keywords:

Mathematical Model, Epidemic prediction, Dengue, Influenza A(H1N1)

Introduction

In 2001 both the Cuban capital and French Polynesia were affected by dengue epidemics. Both places belong to the high dengue risk tropical belt, and dengue was known to both populations. In Cuba half a million persons (more than 5% of the country's population) suffered from Dengue-1 in 1977, whereas 300 000 were affected by dengue-2 in 1982 [1]. In French Polynesia, 17% of the population caught the disease during the 1989 dengue-1 epidemic [2]. An epidemic of Dengue-2 affected Santiago de Cuba in 1997, but was succesfully controled [3].

The 2001 Havana dengue-3 epidemic lasted for almost 9 months[1]. Dengue-1 was flagellating French Polynesia for 10 months. In spite of similar time course, one epidemic affected less than 0.6% of the population of Havana, whereas the other involved 16% of French Polynesians [2].

The paid prizes, however, were similarly high: In Cuba , huge human and financial resources were diverted to curb the epidemic. In French Polynesia, pediatric health infrastructure almost collapsed [2].

These two examples may illustrate the spoil caused by epidemics worldwide. Fighting them is a major public health challenge, and no country can proclaim to be completely safe [4]. Public opinion apparently sees the solution to most of epidemics in the development of vaccines as well as in their availability [4-5].

It seems reasonable to assume that the success of public health response to an epidemic also depends on the possibility to early predict its time evolution [6].

Here, we maintain the viewpoint that early prediction of an ongoing epidemic is a task approachable in a medical informatics framework. We recognize that only the first steps are being made in that direction. Thus, deeper theoretical work is needed together with closer collaboration with public health authorities. At the same time the task is not very appealing for theoreticians. For data miners relevant information appears when huge data sets are analyzed and sophisticated methods are applied. In the case of early prediction of an ongoing epidemic, computations are based on rather coarse approximations to small sets containing between 2 and 10 data points. The soundness of such a task can be negligible mathematically. However, when figures have deal not with abstract numbers, but with human beings under threat the motivation may find a proper space.

As an illustration, the method described in this paper predicted, based on early data available on weeks from 3rd to 10th of the outburst that the cumulative incidence of dengue hemorrhagic fever (DHF) for Windward Islands in French Polynesia would be between 6 and 11 cases/1000 inhabitants. Real cumulative incidence, determined after week 40, totaled, 3.07. Closer to this figure were estimates obtained with standard nonlinear approximation methods, but only available after the 10th week. Standard methods failed for the early stage, and our early forecasts, though relatively rough, could be very valuable as a guidance for preparing public health response during the most critical stage of epidemic fighting. For Havana, on the basis of data available from weeks 3rd till 5th, the total number of cases was predicted between 2 700 and 7 000, 45% of the actual number of cases: 12 889.

During the preparation of this manuscript, an attempt of "true" prediction was undertaken for H1N1 epidemic in the USA. A very important step is model selection [7-9].

A very useful expression for incidence rate was derived theoretically from the classical SIR model developed by Kendrick and McCormick 70 years ago[10]:

$$Incidence=A*4/(\exp(b*t-c)+\exp(c-b*t))^2 \quad (1),$$

where A corresponds to the maximum incidence and Tau=c/b is the time from the beginning until the peak of the outburst . For realistic conditions, this approximation is satisfactory, in particular, it fits nicely to a more complicated model for dengue fever conveyed by the mosquito[11].

As an illustration, in figure 1 incidence data for Winward Islands DHF are fitted to model 1.

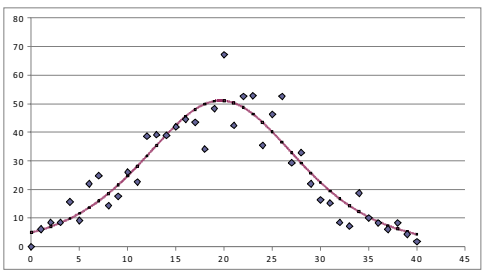

Figure 1-Theoretical bed occupancy with DHF patients in Leeward Islands[2]. Abscissa: Week since the starting of the epidemic; Ordinates: Percentage of theoretical bed capacity. Solid line was estimated from equation (1) using the Hooke Jeeves method.

Incidence data tend to be noisy, but cumulative incidence combine the advantages of a smooth filtering with faithfulness to original data; when little is known about the mechanisms of the ongoing disease, it is advisable to select the simplest models being capable of adapting to different possible variants. For cumulative data S(t), the Richards model can be valid [12-14]:

$$S(t)=K/(1+\exp(R(Tau-t))) \quad (2)$$

K corresponds to the total number of cases and equals

$$K= (S(0)*(1+\exp(R*Tau)) \quad (3)$$

Tau has the meaning of the peak time for incidence; the basic reproductive number R_0 (defined as the average number of secondary cases generated by one primary case) can be estimated as

$$R_0=\exp(T_g*r) \quad (4)$$

Where T_g is the transmission time, or the mean time between the appearance of symptoms in the primary case and the appearance of symptoms in a secondary case [12]. For dengue fever a value of T_g=2.71 weeks has been accepted [14].

There are several approximation methods to fit data into nonlinear models, such as Simplex, Hooke-Jeeves, Gauss-Newton, that have been implemented in different commercial statistical packages. These allow, in principle to simultaneously estimate several parameters from a data set. However, straightforward estimation beyond the domain of observed values with a highly nonlinear function, is not always reliable. As an example, Hooke Jeeves approximation for DHF in Windward Islands yielded cumulative incidence estimates surpassing 40 000, obviously a senseless value. In other words, data are behaving as those typical for ill-posed inverse problems. A practical way to try to deal with this kind of drawbacks is via limiting the space of possible solutions, and imposing to the them certain plausible requirements. In this case, the use of linearizations and manual stepwise estimation of values seems to be recommended. The rationale of our method is based on this philosophy.

Materials and Methods

General Description of the Prediction Method

The first step to linearize the Richards model is as follows. If R(Tau-t)>>1, the expression for S(t) (2) can be seen as S(t)≈ exp (R(t-Tau)), thus the value of R can be obtained from the slope of the curve log(S) vs time.

The next attempt to linearize is via a Taylor expansion of the exponential function. Expanding it, and having deal with realistic numerical values, we found that a good approximation for inverse values of expression can be:

$$(1/S(t))^{(1/4)} \approx -(R(t-Tau)) \quad (5)$$

The right side becomes equal to zero when t=Tau; thus from the relationship between the 4th root of the inverse of the cumulative data and elapsed time, it is possible to obtain a good guess for the time to the peak of the outburst (or "turning point" in the terminology of Richards model [14]) .

From the determination of Tau and R, the value of parameter K can be found from (2) for any value of time t. These estimates obtained "by hand" are subsequently entered as initial guess values for a Hooke Jeeves estimation of data into function (2). Two parameters are fixed an the third one is estimated. After several iterations a refined set of parameters is obtained. In some practical situations, the computer demanding nonlinear estimation step can be omitted.

Results

Dengue hemorrhagic fever in Leeward Islands, French Polynesia

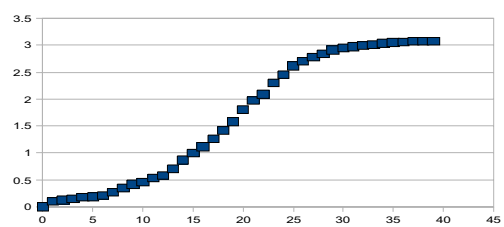

Figure 2-- Cumulative incidence of DHF in Leeward Islands [2]. Axes: Week since the star vs Cumulative incidence.

Hooke Jeeves estimation from the whole data set yielded: K=3.16, R=0.22, and T(weeks)=18.59.

The basic reproduction number (R_0) can be calculated as R_0=1.82 based on a 2.71-weeks transmission time (T_g) [14].

The application of the our method for early prediction based on data from weeks 3rd to 6th, yielded the following values (Table 1):

Table 1-- Early predictions for the Richards model parameters. DHF. Leeward Islands

Week	T	R	K
3	19.6	0.22	6.22
4	21.7	0.20	6.45
5	24.1	0.19	7.17
6	28.1	0.16	7.37
Geometric Mean	23.2	0.19	6.79

Assuming a transmission time of 2.71 weeks, a value of R_0=1.67 is obtained. As appreciable, predictions for T and R are satisfactory whereas for parameter K early estimates roughly doubled the final outcome. We regard that, in practical terms this information is valuable, especially considering that they were given 8 months before the end of the outburst. This value of R_0 is in agreement with other reports using the Richards model for its estimation [14].

Havana dengue-3 epidemic, 2001

The model estimated focumulative numbers for dengue fever cases reported for Havana [1] yielded:
R=0.31
Tau=18
K=12900
(R_0=2.31)
Early estimates for weeks from 3rd to 5th are shown in Table 2.

Table 2 -- Early predictions for the Richards model parameters. Dengue Fever, Havana.

Week	T	R	K
3	11.5	0.42	2661
4	15.33	0.38	6904
5	14.38	0.36	3786
Geometric Mean	13.64	0.39	4113

(R_0=2.88)

In this case the prediction for cumulative number of cases are below 50% of the actual value.

This discrepancy respect to parameter K might reflect our previous finding that Havana Dengue epidemic behaved as several independent outbursts. Curiously, this value is concordant with the predicted size of the early outbreak [11].

An attempt of true prediction: H1N1 in the US.

The possibility for real a priori predictions about an ongoing epidemic came with the advent of H1N1 flu epidemic in the US. Since the first notification of two H1N1 cases in April 2009, the Centers for Disease Control (CDC) were informing, on a daily basis, all the confirmed cases of H1N1 [15]. The information was freely available for any Internet user.

From the beginning of the outburst, predictions were based on a combination of estimates based on the linearizations described above, and then using these estimates as initial guess values for a Hooke Jeeves estimation.

Figure 4 is showing the cumulative cases numbers for the US H1N1 Flu outbreak.

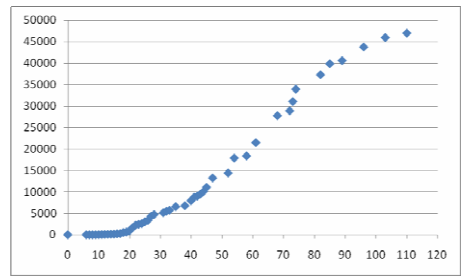

Figure 3-- Cumulative cases of H1N1 in USA(2009). Legend: Abscissa: Day since the starting of the epidemic; Ordinates: Cumulative cases. Redrawn from daily CDC reports [15]

A Hooke Jeeves estimation yielded the following "final" estimates, K=48234.3, R=0.070, and T(days)=63,67.

A value for T_g=2.3 is plausible for influenza A(H1N1) [15-16], corresponding to R_0=1.17, which is in agreement for the first early report in the range of R_0=1.4-1.6 for H1N1 in Mexico [17].

For days from 7th to 20th the following predictions were obtained (Table 3)

Table 3-- Early predictions for the Richards model parameters. H1N1 flu outbreak, US (2009).

Days	Tau	R	K
7	43.98	0.19	9197
8	37.68	0,228	9477
9	22.26	0.577	43541
10	20.9	0.575	20584
11	19.68	0.566	8108
12	28.07	0.30	9990
13	30.07	0.34	38919
14	28.76	0.336	21442
15	31.94	0.331	47121
16	31.97	0.327	44800
17	32.20	0.317	38013
18	29.35	0.328	20252
19	30.89	0.328	32863
20	30.78	0.328	31685

Geometrical mean values were:

Tau	*R*	*K*
29.3	***0.345***	***22576***

(R_0=2.11)

Discussion

Predictions in all the three examples differed from the "true" values obtained from complete data sets. In order to assess the reliability of the method, it is necessary to explore possible sources of disagreement.

Total number of cases (K). At early stages of the epidemic the parameter K is estimated as a nonlinear function of the estimates of Tau, R, and the number of cases at time zero (expression 2). Small errors in either Tau or R can lead to large differences in the estimated value of K. Moreover, the incidence reported at time zero can easily vary in a large relative quantity as it can be the difference between 1 and 2 cases, a situation very likely in everyday practice of data reporting. Thus early estimates for K between 50% and 150% of the "actual" value can be regarded as "acceptable". We expect that this range expected cases, predicted several months before the end of the outburst may be a useful figure for decision makers.

In data for Havana and French Polynesia, early predictions for Tau were between 70% and 130% of the final values. This was not the case for H1N1 in the US where the early estimate for Tau was 45% of the "final" value.

Similarly, estimates for parameter R were similar to "final" values for dengue outbursts in both French Polynesia and Havana (at least 75% of final values), but very divergent for H1N1 in US (almost 500%).

These differences between early predictions and final values can be due either to inaccuracies of the method or to specific features of the data.

We tried to explore method's predictions for parameters Tau and R using simulated data with two different parameters sets. One set corresponds to early estimations for H1N1 and the other set containing the final values. If the accuracy of early prediction is poor, it is to expect that no differences will be found between predictions for each data set.

These results are summarized in Table 4. Predictions reflect the geometrical mean from 19 estimations (from day 2 till day 20). As obtained, estimations in this range of values were clearly different between data sets. The largest difference was for parameter Tau and was 65% of the actual value (19 vs 29.6). For parameter R early predictions were very accurate, better than 97% of actual values.

Table 4-- Early Predictions for model parameters for two simulated data sets. Geometrical means for the first 20 days are shown.

Week	T	R
real	29.26	0.345
estimate	19	0.344
%	65%	100%
real	60	0.07
estimate	65.6	0.068
%	109%	97%

Thus simulation results suggest that differences in the estimates for parameter r are not due to inaccuracies of early prediction. An additional reason for this conclusion might be the dependence of estimates for r as the epidemic proceeded.

Data fitted to an exponential decay curve(r^2 =0.84), and might reflect a real reduction in parameter r, and, subsequently of the basic reproductive number as the epidemic proceeded. Given the nature of this epidemic with an unknown virus and the subsequent strong response by health authorities it is to expect a sharp reduction of infectivity after the initial days of the outburst.

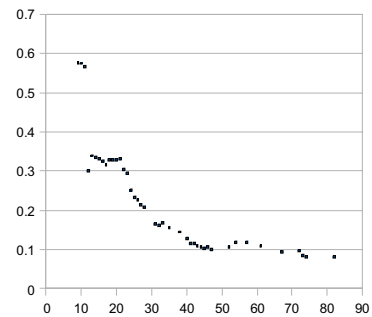

Figure 4-- Time evolution of estimates for parameter R estimated for reported H1N1 cases in USA(2009). Legend: Abscissa: Week since the starting of the epidemic; Ordinates: estimate for parameter R until the given day.

For parameter Tau , a trend towards the increase was also appreciated (r^2=0.91).

Taken together, the reduction in parameter R with increases in parameter Tau can be interpreted as a positive impact of public health decisions to fight the epidemic in the US. As it is known, reducing transmission rate without a reduction of a number of susceptibles leads to an elongation of Tau [8].

Limitations of the present study.

As any research with modeling of real data, the present study is limited by reporting accuracy and under-reporting. In the case of H1N1 in the US, only CDC confirmed cases were studied, these numbers reflect only a small fraction of real H1N1 cases. However, our results might be regarded as a good reflection of the real dynamics of the epidemic.

Limitations related to the method include model selection, limitations of theoretical approximations assumed, as well as those related to assumption abiding under real circumstances. Richards model has been applied for the description of SARS as well as dengue fever and seems to be a good first choice model for different outbreaks. To approximate nonlinear functions with coarse linearizations is always accuracy-costly, and randomness, can additionally hinder theoretical predictions. Simulations suggest that for realistic sets of parameters, excelent accuracy is chieved for R; acceptable predictions can be obtained for Tau and 50-200% or real K-values, based on early estimates from the first 3-5 weeks of an outburst lasting for almost 20 weeks. Assumptions are necessary for getting nice models, but they are not always congruent to reality [17].

Richards model is good for a spatially homogeneous population with parameters unchanged with time. We showed earlier that the 2001 dengue epidemic in Havana corresponds to several independent isolated foci (a nice result from spread prevention measures [11]). It is to expect that for the US, the outbreak can be represented as well as several relatively independent clusters. By actively fighting the spread of the disease, infection parameters change. This is not assumed in the model. The proposed method apparently detects changes associated to intervention measures. The main advantage of the method, is that it can be applied in any setting where an epidemic outbreak appears. A table of logarithms a pencil and a sheet of paper are the only requirements.

Thus a simple method has been proposed for early prediction of the fate of an ongoing epidemic. Its application to real data suggest that some of the discrepancies between early predictions and values obtained at the end of the outburst are not necessarily due to the poor accuracy, but to spatial heterogeneity of the outbreak, or to the impact of prevention measures taken during the time of the outbreak.

Acknowledgments

This research is part of a Cuban National Project on Medical Informatics on "Mathematical models for the Study of Epidemics". My deep gratitude to an anonymous referee for positive comments and encouragement, as well as to Riccardo Bellazzi and Johanna Westbrook from Medinfo 2010 organizing committee for excellent feedback

References

[1] Peláez O, Guzmán MG, Kourí G, Pérez R, San Martín JL, Vázquez S, Rosario D, Mora R, Quintana I, Bisset J, Cancio R, Masa AM, Castro O, González D, Avila LC, Rodríguez R, Alvarez M, Pelegrino JL, Bernardo L, and Prado I. Dengue 3 Epidemic, Havana, 2001.Emerging Infectious Diseases Vol. 10, No. 4, April 2004, 719-722.

[2] Hubert B. Type 1 Dengue fever epidemic in French Polynesia - 2001.Secretariat of the Pacific Community, 2002

[3] Valdés L, Guzmán MG, Kourí G, Delgado J, Carbonell I, Cabrera VM. Epidemiology of dengue and hemorrhagic dengue in Santiago, Cuba 1997. Rev Panam Salud Publica 1999;6(1):16-25.

[4] Laver G and Garman E. The Origin and Control of Pandemic Influenza. Science 7 September 2001: Vol. 293. no. 5536, pp. 1776 – 1777

[5] Coker R. Swine flu. Fragile health systems will make surveillance and mitigation a challenge. BMJ 2009;338:b1791. Published 30 April 2009, doi:10.1136/bmj.b1791

[6] Cohen J and Enserink M. As Swine Flu Circles Globe, Scientists Grapple With Basic Questions. Science 1 May 2009. Vol. 324. no. 5927, pp. 572 – 573

[7] Casagrandi R, Bolzoni L, Levin SA, Andreasen V. The SIRC model and influenza A. Mathematical Biosciences 200 (2006) 152–169

[8] Derouich M, Boutayeb A, Twizell EH. A model of Dengue fever. BioMed. Eng. Online 2 (2003).

[9] Dumont Y, Chiroleu F, and Domerg c. On a temporal model for the Chikungunya disease: Modeling, theory and numerics. Mathematical Biosciences 213 (2008) 80–91

[10] Kermack W O, McKendrick AG. Contributions to the mathematical theory of epidemics. Proc. Royal Soc. A, 115, 700-721 (1927).

[11] Hernández Cáceres JL. Extracting useful information from dengue incidence data. Revista Cubana de Informatica Medica. Vol 7, N2 (2007). Available at: http://www.cecam.sld.cu/pages/rcim/revista_13/articulos_htm/caceres.htm

[12] Zhou G and Yan G. Severe Acute Respiratory Syndrome Epidemic in Asia. Emerging Infectious Diseases Vol. 9, No. 12, December 2003

[13] Richards FJ. A flexible growth function for empirical use. J Exp Botany 1959;10:290–300.

[14] Hsieh YH and Ma S. Intervention Measures, Turning Point, and Reproduction Number for Dengue, Singapore, 2005. Am. J. Trop. Med. Hyg., 80(1), 2009, pp. 66-71]

[15] Center for Disease Control. Interim Guidance for Clinicians on Identifying and Caring for Patients with Swine-origin Influenza A (H1N1) Virus Infection May 4, 2009 4:45 PM ET. Available at http://www.cdc.gov/h1n1flu/identifyingpatients.htm

[16] Carrat F, Vergu E, Ferguson NM, Lemaitre M, Cauchemez S, Leach S, and Valleron AJ. Time Lines of Infection and Disease in Human Infuenza:A Review of Volunteer Challenge Studies. Am J Epidemiol 2008;167:775 –785

[17] Bennett M. 1998 Population and transmission dynamics of cowpox in bank voles: testing fundamental assumptions. Ecol. Lett. 1, 82–86.

Address for correspondence

José Luis Hernández Cáceres. Center for Cybernetics Applications to Medicine, Havana Medical University, Calle 146 esquina a 25, Cubanacan, Playa, Ciudad de La Habana, Cuba.
Email: cacerjlh@infomed.sld.cu