ISSN:0976-4933
Journal of Progressive Science
Vol. 01, No.02, pp 124-136 (2010)

Comparative prediction for the Japanese encephalitis casea by using linear and non-linear modelings

Lal Sahab Singh* **and Anil Kumar Sharma**
Department of Mathematics & Statistics,
Dr. RML Avadh University, Faizabad (U.P.)

Abstract

In this paper, prediction for registered and death cases of Japanese Encephalitis has been carried out by making use of linear and non-linear models. Further, a trend of occurrence of J.E. has also been discussed. Prediction of trend of occurrence has been subject to validation by applying the statistical technique. The present study is taken up for most affected areas of eastern Uttar Pradesh; Gorakhpur and adjoining districts of Bihar and Nepal enveloped by Japanese Encephalitis.

Introduction

Attempt to control Japanese encephalitis (JE) is neglected in part because it's virtually unheard of in the industrialized world. Hosted by pigs and wading birds and transmitted by mosquitoes, the virus mainly strikes poor rural communities in Southeast Asia and the Western Pacific. The World Health Organization (WHO) has estimated that JE claims 10,000 to 15,000 lives a year [8]. But because of low awareness among us and the disease is tough to diagnose, these figures may be gross underestimates. Japanese encephalitis (JE)-epidemics have been reported in many parts of the country. The incidence has been reported to be high among pediatric group with high mortality. The incidence of JE in recent times is showing an increasing trend. It appears that JE may become one of the major public health problems in India, considering the quantum of the vulnerable pediatric population, the proportion of JEV infections among the encephalitic children and wide scattering of JE-prone areas. JE burden can be estimated satisfactorily to some extend by strengthening diagnostic facilities for JE confirmation in hospitals and by maintenance of contact with the nearby referral hospitals to collect the particulars on JE cases. Vaccination proves to be the best to protect the individual against any disease (2004). In the case of JE, it is essential to immunize the pigs (amplifying host) also to interrupt the transmission of the disease. Recognition of JE, based on serological surveys, was first made in 1955 in Tamil Nadu. Subsequent surveys carried out by National Institute of Virology, Pune indicated that about half the population in South India has neutralizing antibodies to the virus. In the last decade, there has been a major upsurge of JE in Assam, Andhra Pradesh, Bihar, Goa, Karnataka, Manipur, Maharashtra, Madhya Pradesh, Tamil

* E-mail: singhdrlalsahab7@gmail.com

Nadu, UP, Pondichery and West Bengal (Singh and Agarwal, 2005). Almost every year, so-called "undiagnosed illnesses" invade India and unfailingly claim thousands of lives. Viral encephalitis, a major global emerging public health problem, is among them. Although viruses are the most common cause of encephalitis, bacteria, fungus, and parasites may also be responsible for infection. In India, although many encephalitis outbreaks have been reported since 1955, several have remained undiagnosed. In the absence of a defined cause, these outbreaks were tentatively attributed to Reye's syndrome, dengue, chikungunya, Japanese encephalitis (JE) and measles. Between July and December 2005, a large and severe epidemic of viral encephalitis was seen in northern India. The disease gripped Uttar Pradesh, border areas of Bihar and Nepal.

The majority of the cases during this epidemic came from eastern Uttar Pradesh (Gorakhpur and adjoining areas), which is the paddy growing "Terai area." Uttar Pradesh (a northern state of India) lies between latitudes 24º and 31º north and longitudes 77º and 84º east and is surrounded by Uttaranchal in the northeast, Haryana and Himachal Pradesh in the north, Delhi and Rajasthan in the west, Madhya Pradesh in the southwest, Chhattisgarh in the south, Bihar in the southeast, and Nepal along the East. After the 2001 census, Uttar Pradesh is the most populous state in India, accounting for 16.4% of the total population of the country. Population density of this state is 689 persons per square kilometer, while it is 324 persons per square kilometer for the country. Children aged 0 to 6 years make up 18.35% of the population, of which 9.58% are male and 8.77% are females, while approximately 40% of the total population belongs to the 0-12 year age group. The rural population is 79.22% (Baily, 1979). There are three distinct seasons: Summer (March to June, with temperatures ranging from 27.5º-32.5 ºC, with a max 45 ºC); Monsoon (July to October, with rainfall of 1,000-2,000 mm in the east, and 600-1,000 mm in the west); and winter (November to February, with temperatures ranging from 12.5º-17.5 ºC). The entire state has a tropical monsoon climate. JE is an acute viral zoonotic infection of the central nervous system (CNS), which produces meningomyelo-encephalitis. It poses a serious public health problem with an increasing frequency of epidemics and outbreaks in many parts of the Indian subcontinent and South East Asian countries over the last four decades. Now approaching newer areas such as Papua New Guinea and Australia, it has been classified as new emerging disease (Baily, 1997). JE virus (JEV) is an arthropod-borne flavivirus and is transmitted to human beings by *Culex tritaeniorrhynchus* and other related rice field-breeding mosquitoes of genus *Culex*. The disease commonly affects children and is a major cause of acute childhood encephalopathy. JEV or antigenically related virus has been identified serologically in different parts of India since the mid-fifties. Although initial cases of JE were reported in the 1950s in India, Uttar Pradesh saw its first epidemic in 1978 (Mathur *et. al*.,1982 and Umanai *et. al*., 1985). Since then, this encephalitis has taken more than 10,000 lives in the state (6, Webb and Perriera, 1956).

Last ten years, most of the cases of acute encephalitis came from Gorakhpur region of Uttar Pradesh The affected area in North India eastern Uttar Pradesh is known as the "Terai area." Floods are an annual feature in the region giving rise to water logging. The warm, humid climate of the region provides an excellent breeding ground for *Culex tritaeniorrhynchus* and *Culex vishnui* mosquitoes, which are vectors of JE. Therefore, during the rainy season, an increase in the population density of the mosquito in this division is observed. The area is densely populated so mosquito-human contact is very frequent. Villages in this rice-growing region abound in stray and reared pigs. Studies from peninsular and eastern parts of India indicate that pigs are the main vertebrate host of the virus and the major reservoir of the infection. Pigs, besides other animals, are widely prevalent in both rural and urban areas of Uttar Pradesh. However, epidemiological and ecological aspects of the illness are yet to be studied in this part of the country

(Sarfling 1973). Due to the evolution of new viral strains and/or reemergence of older viral strains, children lack herd immunity. Although health management facilities are improved in the area, there is still a lack of adequate resources and proper facilities for health care and hygiene (Rathman, 1969). These factors may be the responsible for the intensifying trend of JE in North India. A model using a second harmonic Fourier series for the expected weekly influenza was proposed by Serfling (Saxena *et.al.* 2006) followed by Mukhopadhyay *et.al*. (1993) that provided the mathematical model and showed its fitness for Thailand data. In this paper, a fresh attempt has been made to conduct the study for worst affected Gorakhpur districts of Uttar Pradesh by developing the linear and non linear regression models to predict the occurrences in the current year so that necessary public health measures may be taken in advance to combat the dreadful disease. Although the data relate to the Gorakhpur district of eastern U.P. but the model may also be tried for other parts of Uttar Pradesh as well.

Model development

Linear regression analyzes the relationship between two variables, X and Y. For each subject (or experimental unit), you know both X and Y and you want to find the best straight line through the data. In some situations, the slopes and/or intercepts have a scientific meaning. In other cases, you use the linear regression line as a standard curve to find new values of X from Y, or Y from X. There are two primary reasons for fitting a regression equation to a set of data, d First, to describe the data; second, to predict the response from the carrier. The rationale behind the way the regression line is calculated is best seen from the point-of-view of prediction. A line gives a good fit to a set of data if the points are close to it. Where the points are not tightly grouped about any line, a line gives a good fit if the points are closer to it than to any other line. For predictive purposes, this means that the predicted values obtained by using the line should be close to the values that were actually observed, that is, that the residuals should be small. The basic idea of nonlinear regression is the same as that of linear regression. Nonlinear regression is characterized by the fact that the prediction equation depends nonlinearly on one or more unknown parameters. Whereas linear regression is often used for building a purely empirical model, nonlinear regression usually arises when there are physical reasons for believing that the relationship between the response and the predictors follows a particular functional form. Nonlinear regression is an extended linear regression technique in which a nonlinear mathematical model is used to describe the relationship between the response variable and the predictor variables. A nonlinear regression model is a model that contains at least one of the parameters in a nonlinear form.

Field data is often accompanied by noise. Even though all control parameters (independent variables) remain constant, the resultant outcomes (dependent variables) vary. A process of quantitatively estimating the trend of the outcomes, also known as regression or curve fitting, therefore becomes necessary. The curve fitting process fits equations of approximating curves to the raw field data. Nevertheless, for a given set of data, the fitting curves of a given type are generally not unique. Thus, a curve with a minimal deviation from all data points is desired. This best – fitting curve can be obtained by the method of least squares.

The method of least squares assumes that the best-fit curve of a given type is the curve that has the minimal sum of the deviations squared (***least square error***) from a given set of data. Suppose that the data points are (x_1, y_1), (x_2, y_2), (x_3, y_3),.....(x_n, y_n) where x the independent variable is and

y is the dependent variable. The fitting curve $f(x)$ has the deviation (error) d from each data point, i.e., $d_1 = y_1 - f(x_1), d_2 = y_2 - f(x_2), ..., d_n = y_n - f(x_n)$. According to the method of least squares, the best fitting curve has the property that:

$$\prod = d_1^2 + d_2^2 + \ldots + d_n^2 = \sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} \left[y_i - f(x_i) \right]^2 = a \text{ minimum}$$

Polynomials are one of the most commonly used types of curves in regression. The applications of the method of least squares curve fitting using polynomials are briefly discussed as follows. To obtain further information on a particular curve fitting, please click on the link at the end of each item. Or try the calculator on the right.

The least-squares line method uses a **straight line** $y = a + bx$ to approximate the given set of data, $(x_1, y_1), (x_2, y_2), (x_3, y_3), ...(x_n, y_n)$, where $n \geq 2$ and the least-squares parabola method uses a **second degree curve** $y = a + bx + cx^2$ to approximate the given set of data, $(x_1, y_1), (x_2, y_2), (x_3, y_3), ...(x_n, y_n)$, where $n \geq 3$.

Validity for the prediction is quintessence for the model and therefore it is seriously programmed to conduct with each model which has prediction value. This function enables us to compare the distribution of classes of observations with an expected distribution. Our data must consist of a random sample of independent observations, the expected distribution of which is specified.

Pearson's chi-square goodness of fit test statistic is:

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

where *Oi* are observed counts, *Ei* are corresponding predicted count and n is the number of classes for which counts/frequencies are being analyzed.

The test statistic is distributed approximately as a chi-square random variable with α-1 degrees of freedom. The test has relatively low power (chance of detecting a real effect) with all but large numbers or big deviations from the null hypothesis (all classes contain observations that could have been in those classes by chance).

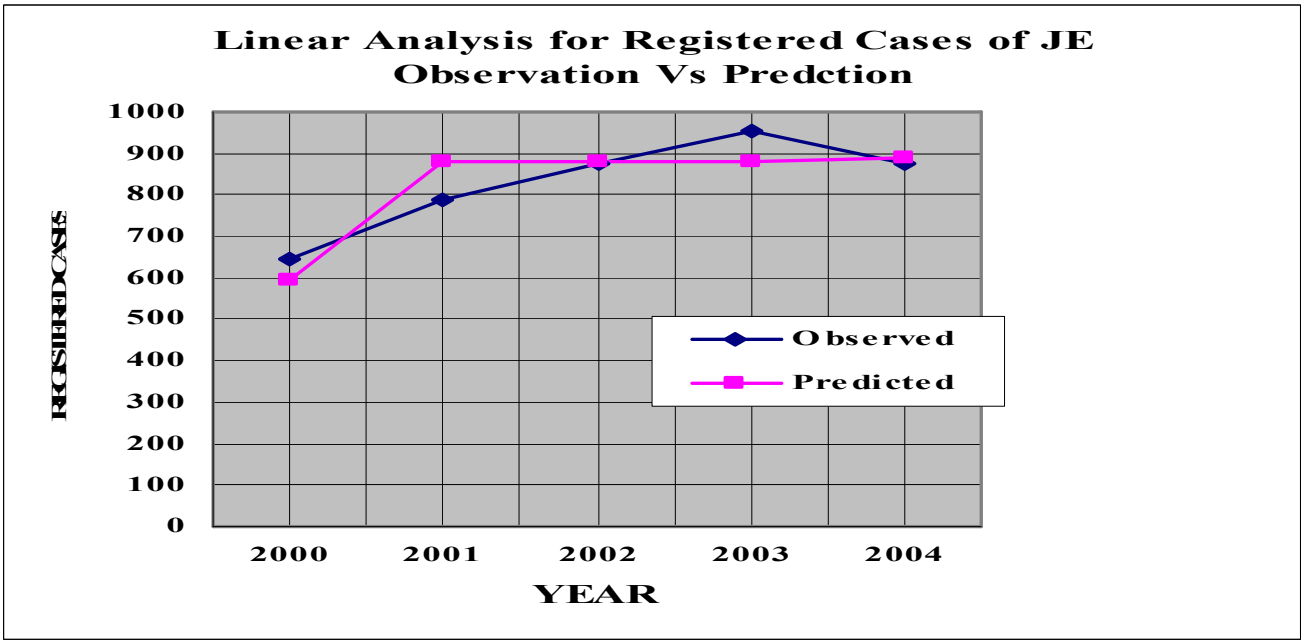
A statistical hypothesis is an assumption about a population parameter. This assumption may or may not be true. The best way to determine whether a statistical hypothesis is true would be to examine the entire population. Since that is often impractical, researchers typically examine a random sample from the population. If sample data are not consistent with the statistical hypothesis, the hypothesis is rejected out of the process of testing the null hypothesis against an alternative hypothesis.

Analysis of model and result prediction

For this study, we collected the secondary data on the suspected JE patients from B.R.D. Medical College, Gorakhpur. Information on various aspects was available in the original data about patients but the factors related to our study were" Yearly" cases of JE patients. The data of registered and death cases of pediatrics JE patients were available to us from 2000 to 2004 and from 2005 to 2009 of all age group JE patients. So, the analysis was done accordingly. After cleaning the data, we have classified and tabulated the data and further diagrammatic representation was done for better understanding with the help of software package MS Excel. In the following section, a linear and non-linear modeling and their fitting for the registered and death cases have been done as conceptualized and developed in the earlier section. The following tables in various cases are presented as outputs or findings of the model under consideration.

Table 3.1 Linear Analysis for registered cases of pediatrics J.E. patients (2000-2004)

Year	Observed	Predicted	C.F. (Observed)	C.F. (Predicted)	$\frac{(A_i - P_i)^2}{P_i}$
2000	646	595	646	595.2	4.33
2001	787	881	1433	1475.9	1.24
2002	875	881	2308	2356.6	1.00
2003	952	880	3260	3237.3	0.15
2004	876	891	4136	4128.0	0.07
N=5	4136	4128	11783	11783	6.79



Graph 3.1

Test Statistics

$\chi^2_{(observed)} = 6.79$

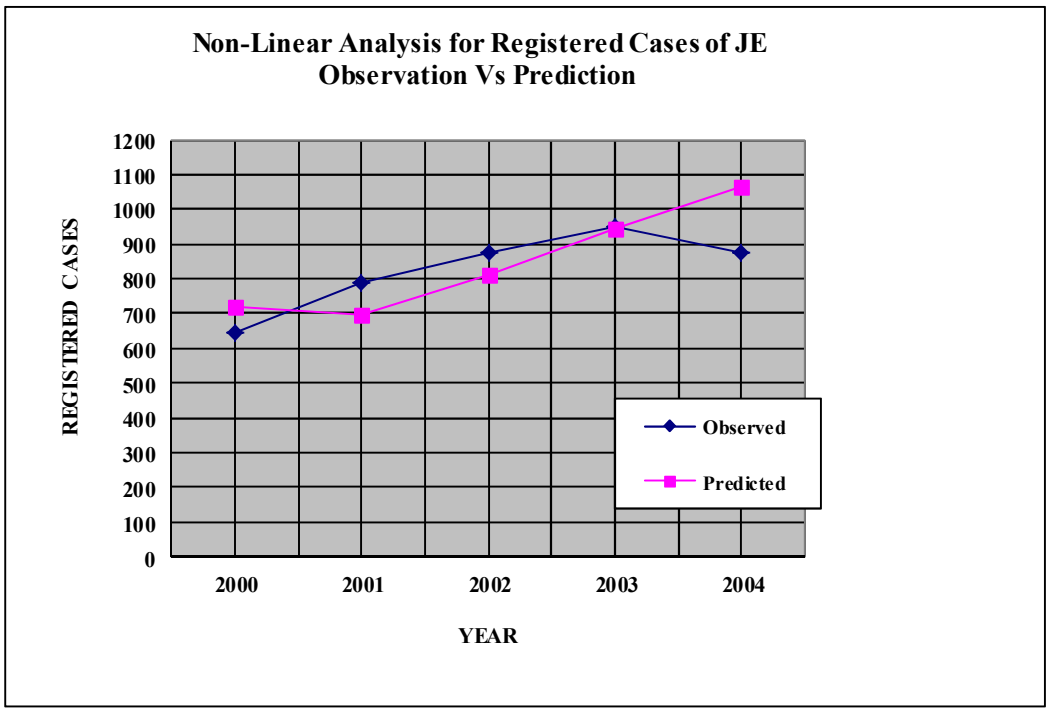
$\chi^2_{(standard)} = \chi^2_{4\ df} = 9.48$ (at 5% level of significance)

Here, $\chi^2_{(observed)} < \chi^2_{(standard)}$

If we closely look at the above computing results, we can conclude that the predicted value is highly insignificant and hence validity of prediction and the proposed hypothesis is accepted i.e. predicted values are valid and useful for the analysis made for the present model.

Table 3.2 Non-Linear Analysis for registered cases of pediatrics J.E. patients (2000-2004)

Year	Observed	Predicted	C.F. (Observed)	C.F. (Predicted)	$\frac{(A_i - P_i)^2}{P_i}$
2000	646	717	646	717.62	7.14
2001	787	697	1433	1414.69	0.23
2002	875	812	2308	2234.18	2.43
2003	952	942	3260	3176.09	2.21
2004	876	1064	4136	4240.42	1.52
N=5	4136	4232	11783	11783	13.53



Graph 3.2

Test Statistics

$\chi^2_{(observed)} = 13.53$

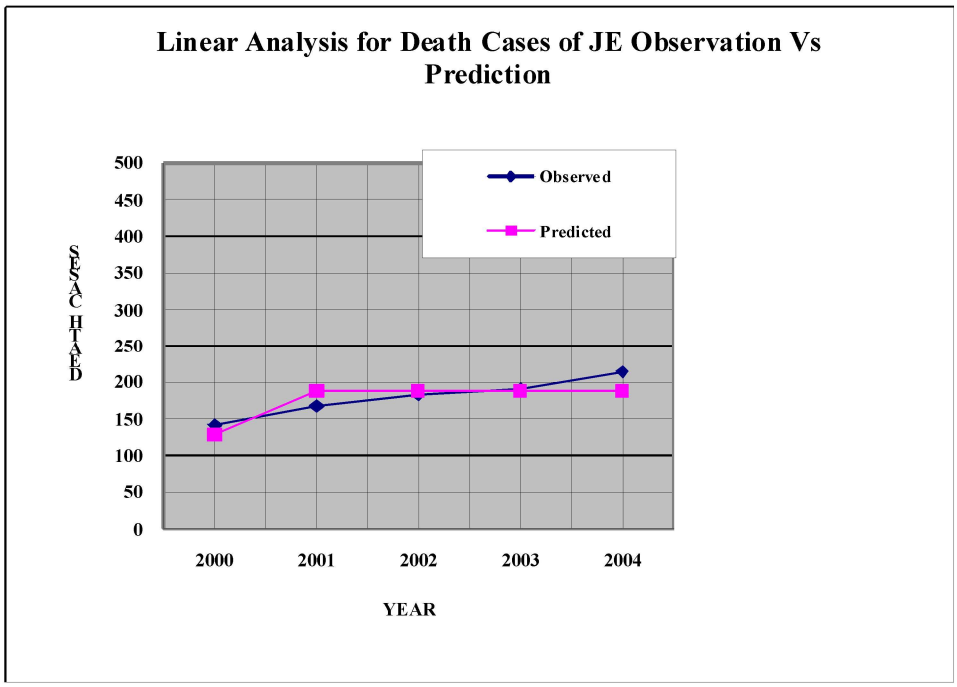
$\chi^2_{(standard)} = \chi^2_{4\ df} = 9.48$ (at 5% level of significance)

Here, $\chi^2_{(observed)} > \chi^2_{(standard)}$

From the computed and standard values of test statistic, difference is highly significant and thereby prediction is rejected for the further use.

Table 3.3 Linear Analysis for death cases of pediatrics J.E. patients (2000 -2004)

Year	Observed	Predicted	C.F. (Observed)	C.F. (Predicted)	$\frac{(A_i - P_i)^2}{P_i}$
2000	142	128	142	128.0	1.53
2001	168	188.5	310	316.5	0.13
2002	182	188.5	492	505.0	0.33
2003	191	188.5	683	693.5	0.15
2004	215	188.5	898	882.0	0.29
N=5	**898**	**882**	**2525**	**2525**	**2.43**



Graph 3.3

Test Statistics

$\chi^2_{(observed)} = 2.43$

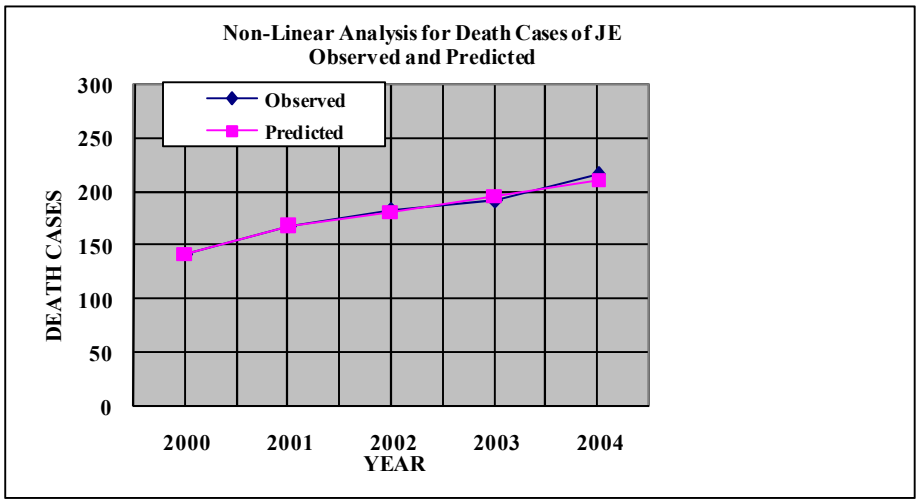
$\chi^2_{(standard)} = \chi^2_{4\ df} = 9.48$ (at 5% level of significance)

Here, $\chi^2_{(observed)} < \chi^2_{(standard)}$

In view of above results, we can conclude that the predicted value is highly insignificant and hence validity of prediction and the proposed hypothesis is accepted i.e. prediction is valid and useful for the analysis made for the present model.

Table 3.4 Non-Linear Analysis for death cases of pediatrics J.E. patients (2000-2004)

Year	Observed	Predicted	C.F. (Observed)	C.F. (Predicted)	$\frac{(A_i - P_i)^2}{P_i}$
2000	142	142	142	142.70	0.003
2001	168	167	310	309.15	0.002
2002	182	181	492	490.30	0.005
2003	191	196	683	686.15	0.014
2004	215	211	898	896.70	0.001
N=5	898	897	2525	2525	0.025



Graph 3.4

Test Statistics

$\chi^2_{(observed)} = 0.025$

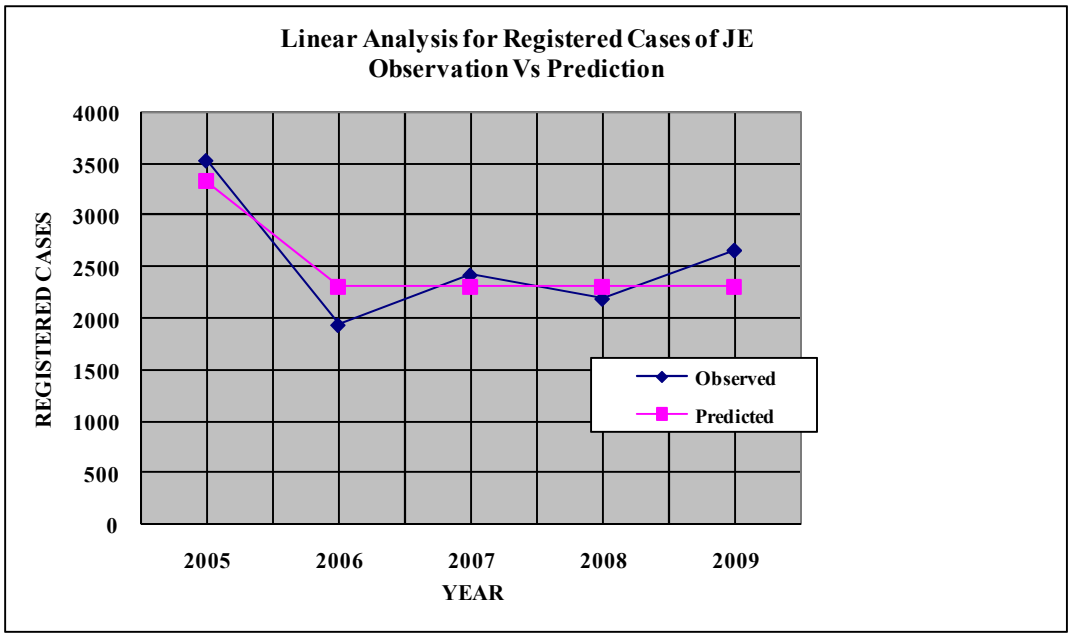
$\chi^2_{(standard)} = \chi^2_{4\ df} = 9.48$ (at 5% level of significance)

Here, $\chi^2_{(observed)} < \chi^2_{(standard)}$

This shows that difference is insignificant and prediction is valid.

Table 3.5 Linear Analysis for registered cases of all age group of J.E. patients (2005-2009)

Year	Observed	Predicted	C.F. (Observed)	C.F. (Predicted)	$\frac{(A_i - P_i)^2}{P_i}$
2005	3532	3336.6	3532	3336.6	11.44
2006	1940	2305.7	5472	5642.3	5.14
2007	2423	2305.7	7895	7948.0	0.35
2008	2194	2305.7	10089	10253.7	2.64
2009	2663	2305.7	12752	12559.4	2.95
N=5	12752	12559.4	39740	39740	22.52



Graph 3.5

Test Statistics

$\chi^2_{(observed)} = 22.52$

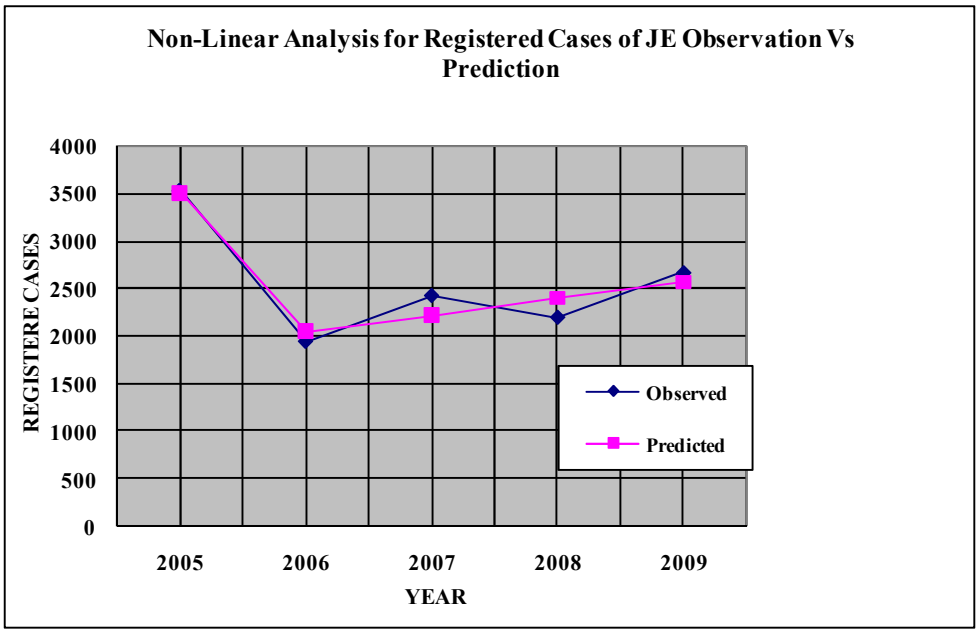
$\chi^2_{(standard)} = \chi^2_{4\ df} = 9.48$ (at 5% level of significance)

Here, $\chi^2_{(observed)} > \chi^2_{(standard)}$

This implies that prediction is rejected.

Table 3.6 Non-Linear Analysis for registered cases of all age group of J.E. patients (2005-2009)

Year	Observed	Predicted	C.F. (Observed)	C.F. (Predicted)	$\frac{(A_i - P_i)^2}{P_i}$
2005	3532	3510	3532	3510.44	0.13
2006	1940	2044	5472	5555.38	1.25
2007	2423	2219	7895	7774.16	1.87
2008	2194	2393	10089	10166.78	0.59
2009	2663	2567	12752	12733.24	0.02
N=5	12752	12733	39740	39740	3.86



Graph 3.6

Test Statistics

$\chi^2_{(observed)} = 3.86$

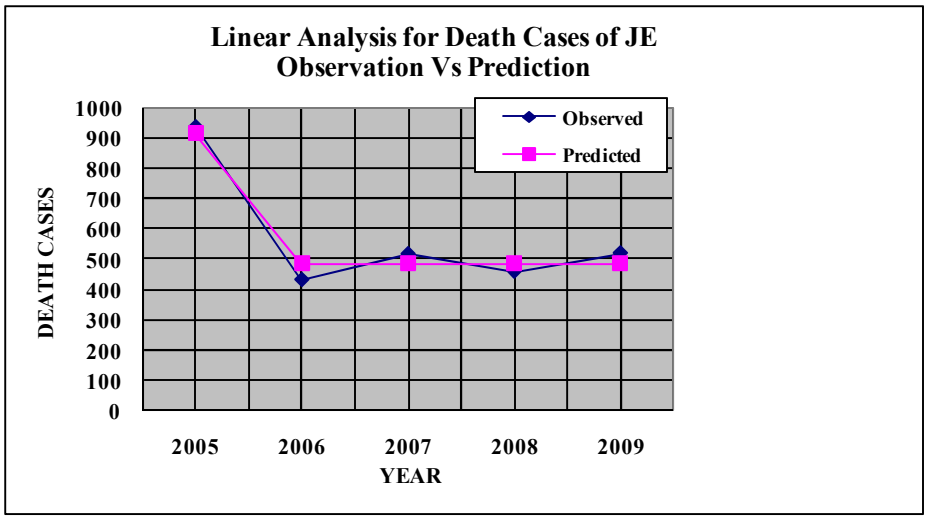
$\chi^2_{(standard)} = \chi^2_{4\ df} = 9.48$ (at 5% level of significance)

Here, $\chi^2_{(observed)} < \chi^2_{(standard)}$

This provides the sufficient evidence to reject null hypothesis.

Table 3.7 Linear Analysis for death cases of all age group of J.E. patients (2005-2009)

Year	Observed	Predicted	C.F. (Observed)	C.F. (Predicted)	$\frac{(A_i - P_i)^2}{P_i}$
2005	937	914	937	914.0	0.57
2006	431	482.3	1368	1396.3	0.57
2007	516	482.1	1884	1878.4	0.01
2008	458	482.2	2342	2360.6	0.14
2009	519	482.1	2861	2842.7	0.11
N=5	2861	2842.8	9392	9392	1.4



Graph 3.7

Test Statistics

χ^2 (observed) = 1.4

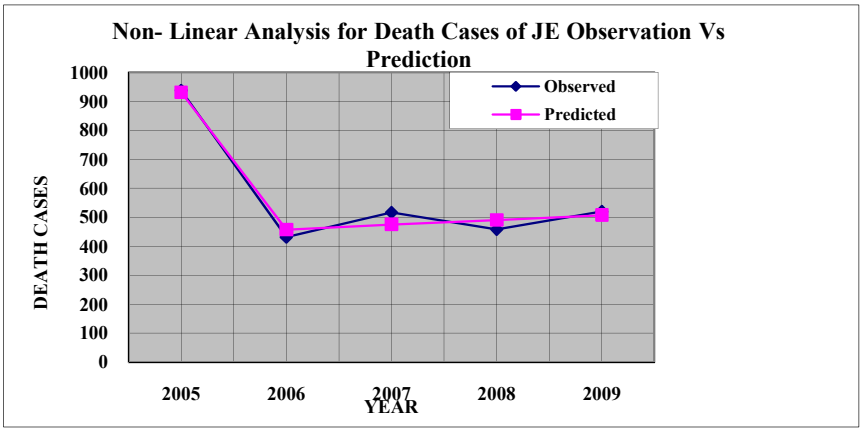
χ^2 (standard) = χ^2 4 df = 9.48 (at 5% level of significance)

Here, χ^2 (observed) < χ^2 (standard)

This situation forms the basis for the rejection of null hypothesis.

Table 3.8 Non-Linear Analysis for death cases of all age group of J.E. Patients (2005-2009)

Year	Observed	Predicted	C.F. (Observed)	C.F. (Predicted)	$\frac{(A_i - P_i)^2}{P_i}$
2005	937	930	937	930.28	0.03
2006	431	457	1368	1388.06	0.28
2007	516	475	1884	1862.12	0.25
2008	458	490	2342	2352.46	0.04
2009	519	507	2861	2859.08	0.00
N=5	**2861**	**2859**	**9392**	**9392**	**0.6**



Graph 3.8

Test Statistics

$\chi^2_{(observed)} = 0.6$

$\chi^2_{(standard)} = \chi^2_{4\ df} = 9.48$ (at 5% level of significance)

Here, $\chi^2_{(observed)} < \chi^2_{(standard)}$

This rejects the prediction

Discussion

Graphs 3.1 – 3.8 shows the observation and prediction of incidences of registered and death cases of pediatrics JE patients from 2000 to 2004 and from 2005 to 2009 for all age group JE patients. The linear and non–linear models with actual data was generally found in agreement for year 2000 to 2004 and year 2005 to 2009 and the curve of predicted, registered and death cases showed almost similar trend as that of the observed incidences of registered and death cases. The results were validated by the appropriate goodness of fit test. The validity test provides sufficient ground for linear and non – linear models for year 2000 to 2004 and year 2005 to 2009 to accept the prediction for JE.

Conclusion

A mathematical model for the prediction of incidences of JE cases has been proposed and dwelt upon in this chapter. This model has been tested with available data of registered and death cases of pediatrics JE patients which were made available to us from 2000 to 2004 and from 2005 to 2009 of all age group of JE patients from the Gorakhpur district of eastern U.P. The result shows that a proposed model simulates the true nature of the validity of prediction of registered and death cases of Japanese Encephalitis. Hence, predictions for the incidences of this dreaded disease in various years can be made satisfactorily, provided the exogenous factors influencing the general pattern maintains average values for the predicted year and nothing unusual happens in the year of prediction. This analysis provides us a scientific footing for the study of this kind of models used in JE problems in the eastern Uttar Pradesh.

References

1. Zile Singh and VK Agarwal (2005) Japanese Encephalitis: Is Routine Immunization Required? MJAFI, Vol. 61, No. 4
2. Baily, N.T.J.(1957) The mathematical theory of infectious diseases and its application, Griffin, London, 2nd edition
3. Baily, N.T.J. (1975) Introduction to the modeling of venereal disease, *J. Math. Bio*. 8, 301-322
4. Directorate General of Health & Family Welfare: Prevention and Control of Japanese Encephalitis Guidelines, Government of India, Directorate of National Vector Borne Diseases control programme, 1-10, 2004
5. Encephalitis (26-27th Sep. 2003) Current Trends & Challenges, Proceedings of Intramural CME, SGPGI, Lucknow, India
6. Government of Uttar Pradesh, Official website. Available: http://www.upgov.nic.in/.
7. Gupta S.C., Kapoor V. K. Fundamental of Applied Statistics, 9.1-9.9.
8. Immunization and Vaccine Development, Japanese Encephalitis. World Health Organization: Regional Office for South East Asia
9. Mathur A, Chaturvedi U.C. and Tandon H.O. (1982) Japanese encephalitis epidemic in Uttar Pradesh, India during 1978. *Indian J Med Res* 75: 161-169
10. Mukhopadhyay, B. B.; Tapaswi, P. K.; Chatterjee A. and Mukherjee, B.(1993) A

mathematical model for the occurrences of Japanese Encephalitis, *Mathl. Comput. Modeling*, 17, 99-103,
11. National Institute of Communicable Diseases(1998) Investigation & Control of Outbreaks Japanese Encephalitis, Directorate General of Health Services, 1-10
12. Rothman, K.J (1997) Modern Epidemiology
13. Serfling, R.E.(1963) Methods for current statistical analysis of excess pneumonia-influenza deaths, *Public Health Report*, 6,78
14. Saxena S.K.; Singh M.; Pathak A.K. and Mathur A.(2006) Reply to 'Encephalitis outbreak finds Indian officials unprepared'. *Nat Med* 12: 269-270
15. Umenai T.; Krzysko R.; Bektimirov T.A. and Assaad F.A. (1985) Japanese encephalitis: current worldwide status. *Bull World Health Organ* 63: 625-631
16. Webb J.K.G.; Perriera S.(1956) Clinical diagnosis of arthropod borne type virus encephalitis in children in north Arcot district, Madras state, India. *Indian J Med Sci* 10: 573-580

Received on 07.09.2010 and Accepted on 30.10.2010