A Data Mining Perspective on the Prevalence of Polio in India

Neera Singh
Indian Institute of Information Technology, Allahabad, India

Sonali Agarwal and Ramesh C. Tripathi
Indian Institute of Information Technology, Allahabad, India

Abstract— **Polio, although has been eradicated from many parts of the world still continues to be prevalent in countries like India, Nigeria and Pakistan. This is a source of concern because a single human carrier in any part of the world is a threat and might lead to resurgence in polio all over the world. There are several reasons for this e.g. lack of information, misinformation, lack of delivery of health services, social stigma attached to it by various communities etc. The purpose of the current work is to analyze the data for the state of Uttar Pradesh and study the age, place, gender, number of doses of polio vaccine administered and various other attributes in the prevalent polio cases .Also to study the data and observe the months when the maximum number of new cases are occurring (season wise variation in incidence of Polio).**

.

Index Terms— **Polio, Prevalence, Eradication, Data Mining**

I. INTRODUCTION

Polio remains one of the major causes of lameness in tropical countries especially in India (see Figure 1), Nigeria and Pakistan. Although somewhat eradicated in temperate countries but the presence of a single human carrier remains a threat leading to the resurgence of disease around the world. Oral Polio vaccine remains the major option to control the disease. The rate of residual paralytic poliomyelitis of 5.4 per 1000 children born in 1969–1970 was 2.3 times higher than the rate of 2.3 per 1000 born in 1973–1974, which may be related to increasing vaccination of children less than one year of age [1]. Oral Polio Vaccine was introduced in India in 1961, when imported vaccines became commercially available [2], introduction of which has seen a drastic fall in the number of polio incidence annually. However polio still seems to be rampant in the states of Uttar Pradesh and Bihar owing to illiteracy, poor health delivery system and misinformation among certain communities regarding its usage. Various studies have also been undertaken to study the prevalence of the post-polio

syndrome was found to be 28.5% of all paralytic cases (95% confidence interval 24.4–32.6). The risk of post-polio syndrome was significantly higher among patients who sustained substantial permanent impairment after polio and among females [3].

Polio data is mainly obtained from school studies, but this continues to be flawed since not all children are attending school. More precise and comparable data about the risk of poliomyelitis could be obtained by incorporating a standard case definition, by using house to house case finding methods in representative community based samples, by analyzing and presenting rates in more clearly defined ways and by selecting a stable population for study [4].Prevalence data can be obtained by studying the number of residual lameness cases. Prevalence data for lameness due to poliomyelitis can be translated into rates of annual incidence of paralytic disease [5].

The purpose of the lameness survey is to chalk out the areas in terms of low, moderate and high incidence and to evaluate the success of the programme after 70-80% coverage of the eligible children was done with at least three doses of vaccine in areas with high incidence [6].The purpose of the study is to predict the most prone districts of Uttar Pradesh(Data of various districts of Uttar Pradesh was analyzed) and to form clusters of areas with somewhat similar risks so that a common programme can be undertaken to deal with the disease in all

these districts. Also a 15 year data of the area of Greater Bombay was studied to analyze the seasonal variations in the incidence of Polio cases.
.

II. METHODOLOGY

A data of recorded cases of Paralytic Poliomyelitis in Greater Bombay (January 1961 to December 1974) was taken. Years 1968 and 1972 had highest incidence of fresh polio cases (see Figure 2).Months of July, August and September recorded having the highest number of cases in these years (see Figures 3&4).This points to a season wise propensity of these months (maybe owing to their high temperature and humidity and water contamination in the rainy season) for occurrence of polio cases.

Also a data of various districts of Uttar Pradesh(602 cases from 27 Districts) was taken and WEKA tool was used for data mining to predict which age group, communities, and districts were prone to which particular strain of polio so that they can be clustered and managed accordingly.
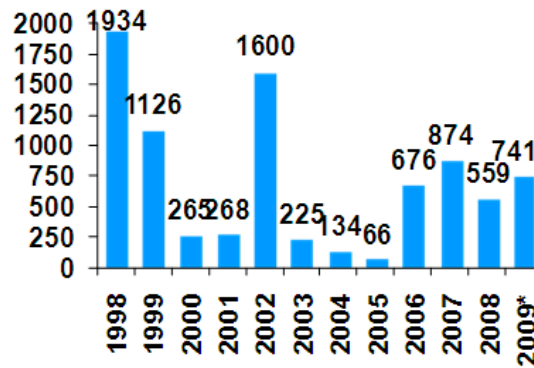

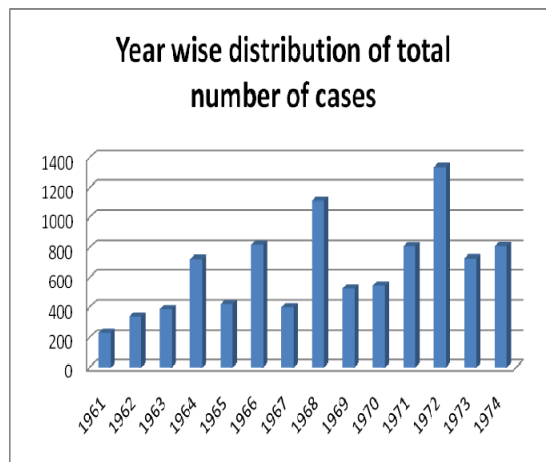Figure1. Polio cases in India from 1998-2009

Figure2. Showing the years with highest number of cases in   Greater Bombay
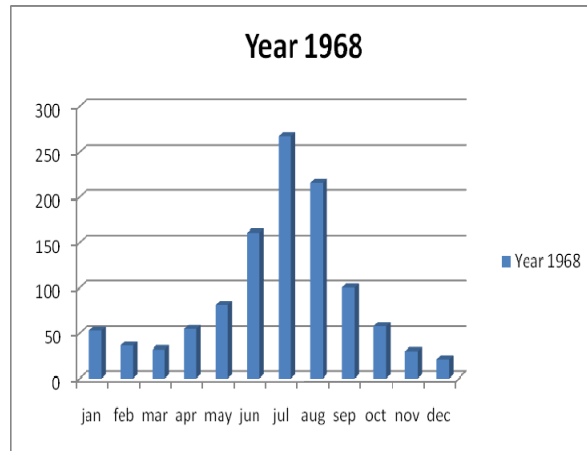
Figure3. Showing the months with highest number of polio incidence in the year 1968
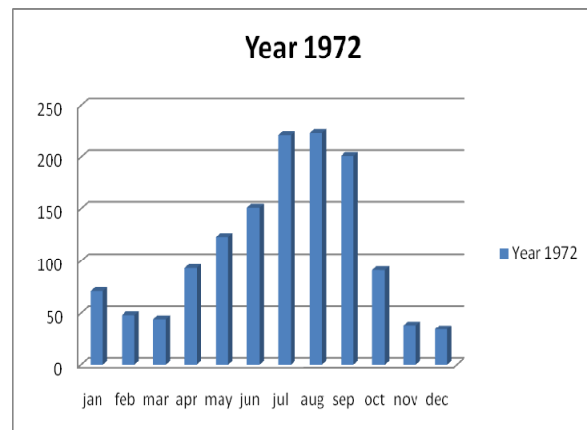

Figure4. Showing the months with highest number of polio incidence in the year 1972

III. RESULTS

Data of 602 cases from 27 districts was taken. The following districts recorded the highest number of cases (see Figure 5):-Ghaziabad (75), Moradabad (74), JyotibaphuleNagar (51), Bulandshahar (43), Badaun (52), Bareilly (36) and Muzaffarnagar (33).

Clustering was performed on this data and 4 clusters were obtained based on age, sex, community, strain of polio whether the case is rural or urban (See Table 1).

Table1. Showing the number and percentage of total cases in a cluster

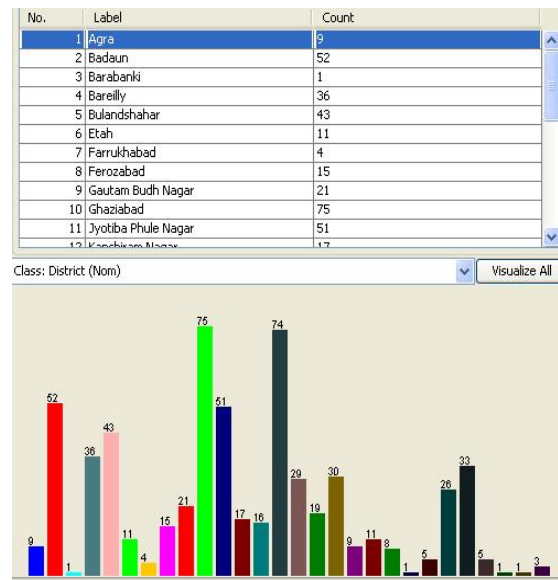| Cluster Number	Number of cases belonging to that cluster	Percentage of cases
1	208	34%
2	224	375
3	157	26%
4	17	3%

Figure5. Showing the districts and their number of polio cases

A decision tree (see Figure 6) was obtained signifying if the age is more than 35months, chances of polio is almost none. If the individual is above 35 month old and is a Muslim chances of having polio is very high. If the age of the individual is below 35 months and is a Hindu and age is greater than 15 months probability of having polio is low and if he is below 15 months of age probability of having p_3 and p_1+p_3 strain of polio is very high. Probability of having p1 strain polio is high if the child is less than or equal to 11 months of age.

   Thus if an individual is more than 3 years old and has polio, chances are that he/she is a Muslim. In Hindus polio is more common among the infants below15 months age. Probability of having p_1 strain is high specially in age groups less than or equal to 11 months old children while p_3 and combination of p_1 and p_3 is more probable in age groups older than 11 months and younger than 15 months.

Also certain association rules (see below) were generated i.e. if an individual is a Muslim in rural setting and the case is hot (fresh) chance is he/she is having p_3 strain polio. If the setting is rural and the patient is a Hindu and the case is not hot (not a fresh case) chances are he is having p_1 strain.

 If the setting is urban and the patient is a Muslim chance is that the strain of polio is p_1.If the religion is Hindu, sex is Female, case is not hot, chances are the strain of polio is p_1.If the religion is Muslim, sex is Male, case is hot then the polio strain is likely to be p3.

 1. Setting Rural / Urban =R Hot=Y 261 ==> Type=P3 252   conf:(0.97)
 2. Religion=M Setting Rural / Urban =R 309 ==> Type=P3 298   conf:(0.96)
 3. Religion=M 367 ==> Type=P3 353   conf:(0.96)

4. Hot=Y 304 ==> Type=P3 290   conf:(0.95)
5. Setting Rural / Urban =R 527 ==> Type=P3 502   conf:(0.95)
6. Sex=M Setting Rural / Urban =R 308 ==> Type=P3 292   conf:(0.95)
7. Sex=F 250 ==> Type=P3 237   conf:(0.95)
8. Sex=M 356 ==> Type=P3 335   conf:(0.94)
9. Setting Rural / Urban =R Hot=N 266 ==> Type=P3 250   conf:(0.94)
10. Hot=N 302 ==> Type=P3 282   conf:(0.93)

IV. DISCUSSION

Polio incidence is subjected to seasonal variation with the number of cases more in the months of July, august and September probably owing to high temperature and humidity and also due to water contamination during the rainy season.

Based on region wise clustering in the state of Uttar Pradesh, districts of Moradabad, Ghaziabad, Jyotiphulenagar were the most prone. Districts with similar occurring cases in terms of age, polio strain, age group, community and region (rural/urban) can be targeted accordingly and polio control regime can be applied similarly in groups with somewhat same attributes.

A decision tree indicating that age group of >35 months and having polio, chances are that the patient is a Muslim. In Hindus it is common in ages less than 15 months.p_3 strain and combination of p_1 and p_3 was more common in >11 months and <15 months age group. While p_1 strain was more common in age group less than 11 months.

Also certain association rules were generated indicating that a rural setting, hot case, male child and a Muslim chances of p_3 strain is high while rural setting, case is not hot, female child and a Hindu religion chances are that strain is p_1. Hindu in an urban setting chances of polio is very low while Muslim in an urban setting chances are that the polio strain is p_1.
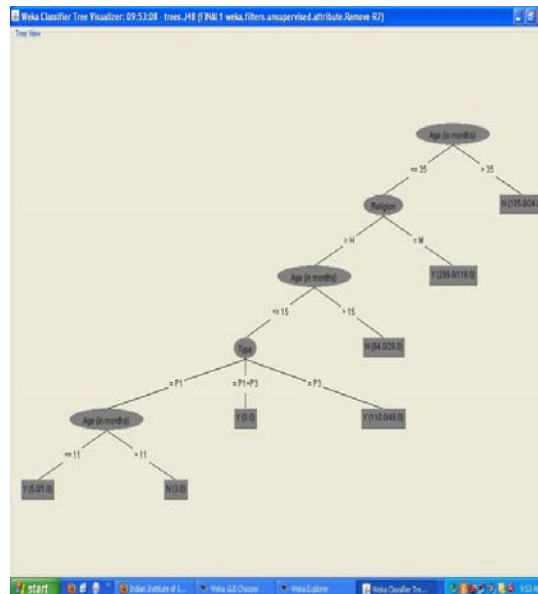

Figure6. Showing the decision tree for the polio case with respect to age group, community and polio strain

IV. CONCLUSION

Certain seasons (July-September) and certain regions (state of Uttar Pradesh in India) are prone to having polio cases. The Muslim community is more prone due to misinformation regarding the polio vaccine. The rural population is more prone due to lack of proper infrastructure and poor delivery of health services in these areas. Certain strains affect certain population and age groups. The idea is that these sections of the population should be targeted with the relevant vaccine for proper control of polio. If only the high risk groups, regions and time of polio spread can be identified, effective health programmes can be targeted at these sections for adequate control and prevention of spread of this disease.

REFERENCES

[1] Albert B. Sabin and Ernesto Silva, Residual paralytic poliomyelitis in a tropical region of Brazil, 1969–1977: prevalence surveys in different age groups as indicators of changing incidence. American Journal of Epidemiology Vol. 117, No. 2: 193-200.

[2] T. Jacob John, *Reviews of Infectious Diseases*, Poliomyelitis in India: Prospects and Problems of control, Vol. 6, Supplement 2. International Symposium on Poliomyelitis Control (May - Jun., 1984), pp. S438-S441.

[3] Jonathan Ramlow, Michael Alexander, Ronald LaPorte, Caroline Kaufmann and Lewis Kuller**.** Epidemiology of the Post-Polio Syndrome.American Journal of Epidemiology Vol. 136, No. 7: 769-786.

[4] Roger H. Bernier, Some Observations on Poliomyelitis Lameness Surveys. *Reviews of Infectious Diseases*, Vol. 6, Supplement 2. International Symposium on Poliomyelitis Control (May - Jun., 1984), pp. S371-S375.

[5] F. M. LaForce, M. S. Lichnevski, J. Keja, & R. H. Henderson, Clinical survey techniques to estimate prevalence and annual incidence of poliomyelitis in developing countries. Bulletin of World Health Organization ,58{4} ,609-620 {1980}.

[6] R. F. Basu and J. Sokhey Prevalence of Poliomyelitis in India, Indian Journal of Pediatrics,Volume 51, Number 5 515-519, DOI: 10.1007/BF02776613