ON A METHOD OF EVALUATION OF EPIDEMIC INDICES RELATING TO CHOLERA

By

SUDDHENDU BISWAS M.Sc.,

Demography Section, Indian Statistical Institute

1. INTRODUCTION

Cholera, more often known as Asiatic cholera, is a grim disease ascribed to a specific infection with *Vibrio cholerae*. It has been present in India from the remotest antiquity and persists as a perennial scourge along the Ganges river. Since some cases of severe diarrhoeal disorders or baciliary resembling this disease are often reported as Cholera, Statistics relating to the incidence of this disease should be accepted only with a certain amount of reservation. However, the mortality statistics of cholera incidence in cities and towns could be more easily relied upon, and these clearly indicate persistence of cholera in certain densely populated deltaic regions. Nevertheless, cholera has not only confined itself within the borders of Asia, but had been responsible for wide-spread epidemics in almost all countries of the temperate zone in the nineteenth century. With improvement in sanitation and advancement of the health consciousness of the people in the European countries and with the strict enforcement of international quarantine regulations, this disease has virtually disappeared from the European countries. It is, however, typical of tropical countries that this disease still persists in an endemic form in certain densely populated areas, especially in the Gangetic delta, which renders these areas huge reservoirs of infection. Poor sanitation and the low level of health consciousness alone, could not have rendered this disease endemic, in these areas for there are other areas in India densely populated and handicapped equally in respect of sanitation and level of health consciousness, where the disease was never endemic. This fact led to a series of investigations about the characteristic features of this area, such as its general topography, climatic conditions and their association with high endemicity.

Bryden (1869) was the pioneer in this investigation, and his study revealed certain interesting features about the geographical distribution of the incidence of cholera. He showed with the help of a map that the tract traversed by the disease over different regions in India had always their origin from one or the other of the highly endemic areas of the Gangetic delta. These findings were later confirmed by Rogers (1928) by a more rigorous statistical analysis of the cholera mortality data. Incidentally, Rogers also ascribed a vitally important physical factor, viz., a minimum pressure of 0.4 mm. of absolute humidity to the spread of the disease. He further stated as this meteorological factor is a constant, the course of the transmission of the disease remained virtually unaltered inspite of the extension of the inter-district and inter-state communication which took place during the 60 years following Bryden's investigations. Subsequently, Russel and Sunderrajan investigated into the periodity of cholera incidence (1928) in different provinces of India and their correlation with the meteriological condition. More recently, Raja, Swaroop and Lal (1941) carried out an intensive study in Bengal with monthly death rates due to cholera obtained for a period of 30 years for each district (1901-'31) wherein they attempted to classify the districts into types characterised by their endemic and epidemic patterns. For the investigation of the degree of endemicity a frequency distribution was made of varying intervals of absence of cholera. The yearly variances within the cholera districts provided with an index for the classification of the districts by the degree of epidemicity.

The study of endemicity was continued by Swaroop (1951) with the coverage extended over all the districts of the Indian Union and the data included within the scope of the study were the weekly incidence figures along with the monthly and yearly mortality figures. A statistical evaluation of the three types of endemicity indices based on the yearly, monthly and weekly data was made for the classification of the districts into varying levels in respect of endemicality. As a result of this classification, certain important features showing the association of endemic prevalence of cholera with certain topographic and climatic conditions of these areas concerned, were observed. Most of the studies so far carried out related to the endemic prevalence of cholera and their seasonal pattern. The only study carried out so far relating to the epidemicity of cholera was the one made by Raja, Swaroop and Lal (1941). As this study was based on annual mortality figures, the indices worked out were rather crude in as far as the variations within the years were pooled thereby reducing considerably the efficacy of the index as a measure of epidemicity. The indices worked out in the above study neither give any idea about the frequency of the epidemic nor does it help us to study the important characteristics of the epidemic pattern which differentiates the so called non-endemic areas with the endemic ones. Further, the endemic areas could always act as vast reservoirs of infection and in these days of rapid communication the infection could be spread even to the remote areas which have been so long enjoying a fair amount of security against the ravages of the disease. The so-called non-endemic areas experience occasional outbreaks, the frequency of such outbreaks depending on the degree of contact with the endemic areas. In these days of rapid inter-district or inter-state communication, the non-endemic areas so long considered as geographically insulated from endemic areas, present an equally important problem in the field of public health as those areas in which the disease persist as a perpetual menace. An attempt is, therefore, made in this paper to study the epidemic pattern of cholera in non-endemic areas and contrast those patterns with those obtained in endemic areas from three points of view:

(1) The frequency of epidemic outbreaks.
(2) The average duration of epidemics.
(3) The average intensity of epidemics.

2. DATA

At the very outset of this study, it was considered advisable to restrict the choice of endemic and non-endemic areas to those inhabited by compact communities since clear patterns of epidemic prevalence are likely to emerge only when perfect homogeneity of epidemiological factors within the communities concerned, are assured. On the other hand, due consideration had to be paid to retaining a fairly good size of the population of the unit selected, so as to reduce the erratic results arising from the smallness of the sample. Taking account of these facts, a subdivision (with an average population of 700,000) was considered to be reasonably satisfactory choice for the study of epidemic patterns. Nineteen subdivisions were selected for this study among 45 subdivisions of West Bengal so as to obtain an adequate representation of the endemic as well as non-endemic areas of West Bengal. The population in these subdivisions as enumerated in the 1941 and 1951 censuses and the average annual cholera mortality rates for the period, viz., January 1949 to June 1956 are shown in Table I.

3. ANALYSIS

The weekly mortality figures of each selected subdivisions were arranged, in a two-way table, the rows and the columns denoting the year and the chronological order of the weeks within the year respectively. By this arrangement variation of the weekly rates in any one column could be rendered independent of the seasonal effect of cholera mortality.

If one can assume the complete absence of secular trend and the absence of any epidemic, then the weekly death rates entered in any particular column can be considered as a statistically independent set of observations from a definite Poisson distribution because the seasonal factors affecting cholera mortality have been held constant within any specified column. Any weekly rate included in a column which is large enough to be considered as statistically an outlying observation must probably indicate the distortion of epidemic balance. Therefore, a statistical method to detect such outlying observations

TABLE I

Population and cholera mortality rate of 19 Sub-divisions of West Bengal.

	Name of the Sub-division		Population as enumerated in 1941 census	1951 census	Population density per square mile 1951	Average annual cholera mortality rate from 1949-1953
1.	Calcutta (City) ...	...	2,108,891	2,520,921	88,933	85.88
2.	24-Parganas Sadar	...	1,225,485	1,313,948	1,368	57.29
3.	Barasat Sdn. ...	...	317,261	393,980	1,026	22.84
4.	Basirhat Sdn. ...	...	594,077	713,619	873	82.68
5.	Diamond Harbour Sdn.	...	818,309	901,120	714	97.10
6.	Barrackpore Sdn.	...	579,995	877,900	7,371	24.95
7.	Bongaon Sdn. ...	...	133,104	280,742	653	23.15
8.	Howrah Sadar ...	...	830,345	928,456	5,333	64.84
9.	Uluberia Sdn. ...	...	659,959	682,917	1,769	92.98
10.	Hooghly Sadar ...	...	398,969	454,573	1,019	22.22
11.	Serampore Sdn. ...	...	634,275	729,331	2,085	36.47
12.	Midnapore Jhargram Sdn.	...	426,245	461,703	389	11.48
13.	Burdwan Sadar ...	...	737,651	802,507	623	21.80
14.	Murshidabad Sadar	...	499,749	544,228	826	12.31
15.	Malda Sadar ...	...	844,315	937,380	674	20.58
16.	Bankura Bishnupur Sdn.	...	352,959	353,896	496	18.08
17.	Jalpaiguri Sadar ...	...	524,884	546,142	421	6.22
18.	Darjeeling Sadar ...	...	147,327	169,631	470	0.00
19.	Darjeeling Siliguri Sdn.	...	90,014	116,475	437	6.87

** Rates per 100,000 population.

The weekly mortality rates of each of the 19 sub-divisions per 100,000 of its mid-year population were obtained from the registration data for a span of 7½ years (January, 1949 to June, 1956).

is a prerequisite to the study of epidemics.

The column observations are at the outset of our analysis transformed into a set of normal variables ($X_1, X_2, \ldots\ldots X_n$) where n=7 or 8, by the usual square root transformation process. Subsequently, the normal variates are transformed to a new set of variates ($\tau_1 \tau_2 \ldots\ldots \tau_n$) denoted hereafter τ_i (i=1, 2, 3, ... 7 or 8) where $\tau_i = \frac{x_i - \bar{x}}{s}$ (i=1, 2, 3 .... 7 or 8), $\bar{x}$, s being column means and standard deviations respectively. The theoretical distribution of τ as given by Thompson is

$$p(\tau) = \frac{\sqrt{\frac{N-1}{2}}}{\sqrt{(N-1)}\,{}_1\tau\sqrt{\frac{N-2}{2}}} \cdot \left(1 - \frac{\tau^2}{N-1}\right)^{\frac{N-4}{2}}$$

or in a more concise way $t = \sqrt{\frac{N-1}{N-2+t^2}}$

where t follows a students' t distribution with (N-2) degrees of freedom. The critical values of τ exceeding the 95% level of probability (right side) for different degrees of freedom has been computed and shown in Table II.

TABLE II

Upper limits of τ for different degrees of freedom (95% probability)

Degrees of freedom (1)	Upper limit of τ (2)	Degrees of freedom (1)	Upper limit of τ (2)
1	1.39	9	1.69
2	1.56	10	1.74
3	1.61	11	1.80
4	1.63	12	1.92
5	1.64	13	2.04
6	1.64	14	2.38
7	1.65	15	3.41
8	1.65		

Any value of τ in a given column which exceeded the 95% (right side) level of probability were rejected straight away and the mean and the standard deviations of the residual observations of the column were computed afresh and the same procedure of rejecting the observations (τ_i) was employed. This process was repeated till the residual values of (τ_i) in the column admitted no further rejection. (Pearson & Chandrasekhar. 1936).

If now, we consider that during the weeks relating to the rejected observations the subdivisions had experienced epidemics we shall be marking a number of epidemic patches in the sequence of 390 weeks that have been included in our study. This method, however, is not very helpful in precisely determining the onset and end of an epidemic. Cases are not very infrequent

TABLE III

TABLE 3—*Estimates of epidemic indices for 19 Sub-divisions (period of study January 1949 to June 1956)*

	Name of the Sub-divisions		Frequency of epidemic patches and rank		Average duration of epidemic patches and rank		Average intensity of epidemic patches and rank	
	(1)		(2)	(3)	(4)	(5)	(6)	(7)
1.	Calcutta	...	9	VI	9.89	I	6.72	I
2.	24-Parganas,	Baraset ...	13	II	5.77	XIV	2.64	XI
3.	„	Basirhat ...	12	III	9.73	II	5.76	III
4.	„	Diamond Harbour	12	III	8.83	VI	6.24	II
5.	„	Barrackpore ...	16	I	5.13	XV	2.11	XIV
6.	„	Sadar ...	11	IV	6.82	XI	4.10	VI
7.	„	Bongaon ...	8	VII	7.75	X	5.42	IV
8.	Howrah,	Uluberia ...	11	IV	9.55	IV	3.93	VIII
9.	„	Sadar	10	V	8.60	VII	4.67	V
10.	Hooghly,	Serampore ...	12	III	6.75	XII	2.53	XII
11.	„	Sadar	9	VI	8.89	V	2.07	XV
12.	Midnapore,	Jhargram ...	13	II	4.62	XVI	1.61	XVII
13.	Burdwan,	Sadar	9	VI	9.56	III	4.06	VII
14.	Murshidabad,	Sadar ...	12	III	6.50	XIII	3.40	X
15.	Malda,	Sadar	9	VI	8.44	VIII	3.92	IX
16.	Bankura,	Bishnupur ...	6	VIII	8.15	IX	2.37	XIII
17.	Jalpaiguri,	Sadar ...	6	VIII	3.00	XVIII	1.70	XVI
18.	Darjeeling,	Siliguri ...	5	IX	3.33	XVII	1.42	XVIII
19.	Darjeeling,	Sadar ...	0	X	0.00	XXI	0.00	XIX

where the week immediately after or before an epidemic week (detected by Thompson's criterion) for the same year is not discarded as epidemic, though the rate of the former is sometimes larger than that of the latter. This anomaly arises owing to the fact that the variables in the columns either preceding or succeeding the column of the aforesaid week being generally high may have lower values of τ adjacent to the one rejected. Also two adjacent epidemic patches may sometimes be separated by too short an interval, say, one week which on epidemiological considerations seem to be highly improbable. Such anomalies, however, have been eliminated by the method described below.

If the observation relating to the ith week of the jth year, X_i, j, has been rejected as an outlying observation of the ith column whereas X_{i+1}, j, X_{i+2}, j relating to the succeding two weeks have not been rejected by the application of the same method on (i+1)th, (i+2)th, columns respectively, then by the method of extending patches we will consider X_{i+1}, j also as an outlying observation, if X_{i+1}, j is $> X_i, j$ or is greatest among the observations of (i+1)th column. If X_{i+1}, j does not satisfy the above conditions, then X_{i+2}, j may be probed in a similar manner, and if X_{i+2}, j also does not satisfy the above conditions, then X_i, j must be considered as the end of a patch. If, however, X_{i+2}, j satisfies the conditions stated above, but not X_{i+1}, j then the latter should also be considered as an out lying observation. In general, if two adjacent patches are separated by a single nonepidemic week, the two patches are joined together into a single patch. If X_{i+1}, j or X_{i+2}, j are rejected thus, using these observations as starting points one could continue extending the patch by drawing successively until no further extension is possible by this rule. The same rule of extensions of the patch can be made on the left side to obtain the onset of epidemic.

Formation of epidemic patches leads us to the crux of the problem since our study of the epidemic patterns of all the subdivisions is based on their characteristic features, viz., (1) number of epidemic patches within the given period of study for different subdivisions (2) average length of epidemic patches for different subdivisions (3) average intensity of epidemic curves, (average maximum weekly rates over all the epidemic patches).

4. RESULT

The number of epidemic patches in the sequence of 390 weekly mortality figures (January 1949 to June 1956) with their average lengths and intensities are shown in Table III above.

It is true that these patches do not strictly indicate periods during which the particular communities referred to had experienced epidemic situations. What can at most be inferred about the periods pertaining to those patches, is that the mortality figures experienced by those particular communities wer somewhat unusual for those communities in that part of the year. The only advantage in employing statistical method for picking out the so called unusual weeks and extending them on either side to form a continuous patch was to introduce a certain amount of objectivity in the marking of such patches, so that the indices of the 19 subdivisions given in Table III can be considered as more or less statistically sound.

5. CONCLUSION

Any inference drawn from these indices is strictly limited to those that follow a post mortem examination, for one does not know anything about the epidemiological conditions prevailing in the subdivisions during the weeks included in the patches. Since an epidemic is the result of a change in the balance between the host population, environment, and the nature of invading parasites and moreover, as a multitude of factors could effect such changes, the assessment of the factors from their results are merely a matter of conjecture.

An overall glance at the figure in Table III reveals that in the subdivisions, viz., Jalpaiguri sadar, Siliguri and Darjeeling sadar, the epidemics are not only very rare but also have exceedingly short duration and low intensity. These three form a class of non-epidemic and nonendemic regions. By virtue of their typical topographic features, low population density and relatively high degree of insulation from the rest of the country, the infection rate is low and the import of new cases is also low

Among the rest of the subdivisions, varying epidemiological features are revealed by these three indices. For instance, Calcutta had relatively fewer outbreaks of epidemics, but the epidemics lasted for considerably longer periods and also attained high peaks; and in respect of their average duration as well as intensity the highest recorded indices go to the credit of Calcutta. This is not somewhat surprising if we bear in mind that Calcutta has a uniformly high density of population, bulk of it being bustee dwelllers, such that the disease prevails in a highly endemic form.

Peculiarly enough, subdivisions like Bongaon and Malda, etc. had a similar epidemiological, pattern, viz., relatively few outbreaks of epidemics but in respect of average duration as well as intensity of epidemic curves, they were significantly high. These two subdivisions are characterised by continuous additions to their population due to the inflow of refugees from Eastern Pakistan as a consequence of which the epidemiological patterns similar to those of highly endemic areas has gradually developed in these subdivisions.

Subdivisions like Howrah Sadar, Uluberia, Basirhat and Diamond Harbour not only have frequent outbreaks of epidemics but were also ranked high in respect of average duration and intensity of epidemics. These places by virtue of their topography and high density of population show the characteristic features of high endemic areas and in addition, by virtue of their close proximity to the city of Calcutta experience frequent outbreaks of cholera epidemic due to import of fresh cases from Calcutta, probably by vegetable and fish vendors.

Another class of four subdivisions, viz., Midnapore, Jhargram, Murshidabad, Barrackpore, Baraset exhibit very typical epidemiological features. They have experienced during a span of 7½ years 13 to 16 outbreaks of epidemics but the duration and intensity of such epidemics are remarkably low. This might be typical of comparatively nonendemic areas having frequent import of new cases from endemic areas

Finally, for subdivisions like 24-Parganas Sadar, Serampore, Hooghly Sadar, Burdwan and Bankura, though no definite epidemiological pattern is revealed those prevailing in these areas could however, be ascribed to semi-endemic conditions.

These three indices, can, therefore, be consiered as summarising the epidemiological patterns available in various subdivisions of West Bengal. But in this paper, the computation of these three indices were based on data extending over a period of 7½ years, and as such the values obtained should be considered as very approximate and consequently not adequate for drawing precise inferences. But this study answers the epidemiological problem only partially in as the epidemiological factors operating on the populations at the time of epidemics are unknown. A more extensive investigation is needed to probe into the factors that cause the variation from one community to another. Vaguely it may be stated from the results obtained in this study, that there exists areas in the close proximity of Calcutta and other densely

(Continued on page 196)