Effect of temperature, relative humidity and rainfall on rotavirus infections in Kolkata, India

A. SUMI*, K. RAJENDRAN², T. RAMAMURTHY², T. KRISHNAN², G. B. NAIR², K. HARIGANE³ AND N. KOBAYASHI¹

¹ Department of Hygiene, Sapporo Medical University School of Medicine, Sapporo, Hokkaido, Japan
² National Institute of Cholera and Enteric Diseases (NICED), Kolkata, India
³ Department of Nursing, Tenshi College, Sapporo, Hokkaido, Japan

Received 29 February 2012; Final revision 31 July 2012; Accepted 3 September 2012; first published online 8 October 2012

SUMMARY
Rotavirus is a common viral cause of severe diarrhoea. For the underlying cause of rotavirus seasonality, the meteorological factor has been suspected, whereas quantitative correlation between seasonality and meteorological factor has not been fully investigated. In this study, we investigated the correlation of temporal patterns of the isolation rate of rotavirus with meteorological condition (temperature, relative humidity, rainfall) in Kolkata, India. We used time-series analysis combined with spectral analysis and least squares method. A 1-year cycle explained underlying variations of rotavirus and meteorological data. The 1-year cycle for rotavirus data was correlated with an opposite phase to that for meteorological data. Relatively high temperature could be associated with a low value of isolation rate of rotavirus in the monsoon season. Quantifying a correlation of rotavirus infections with meteorological conditions might prove useful in predicting rotavirus epidemics and health services could plan accordingly.

Key words: Analysis of data, climate – impact of, rotavirus, statistics, surveillance.

INTRODUCTION
Rotavirus is the most common aetiologic agent of severe diarrhoea in children worldwide [1]. Globally, rotavirus accounted for 527,000 deaths in children aged <5 years in 2004, representing 5% of all deaths in children in this age group. Twenty-three percent of deaths due to rotavirus disease were estimated to occur in India alone [2]. Annually, India paid US$41–72 million in medical expenses for the treatment of rotavirus diarrhoea [3]. For prevention of this disease and alleviation its economic burden on India, it is necessary to understand the epidemiological factors of rotavirus epidemics, for this the mechanism of seasonality of such epidemics is of great interest [4–10].

An observational investigation of the seasonality of rotavirus disease visually confirmed that higher numbers of rotavirus infections in the tropics were found at colder and drier times of the year [6, 10]. This temporal behaviour of rotavirus epidemics was confirmed in the case of Kolkata, India, with surveillance data collected during the period 1979–1981 [11]. Furthermore, it was seen that seasonality of rotavirus infections in Kolkata are inversely correlated with three kinds of meteorological conditions: temperature, relative humidity and rainfall.
More recently, in Kolkata, systematic surveillance for diarrhoeal aetiologies in hospitalized patients was conducted from November 2007 to December 2009. Based on this surveillance, it was reported that rotavirus epidemics in Kolkata indicated a regular, yearly cycle with a peak in the winter months [12]. There is considerable interest in discovering whether or not temporal patterns of rotavirus epidemics in Kolkata from 2007 to 2009 had an inverse correlation with temperature, relative humidity and rainfall as has been reported for the period 1979–1981 [11]. In addition, it is possible that the occurrence of rotavirus infection in the monsoon season in Kolkata is associated with high temperature, as in the case of Bangladesh [4].

In order to investigate the relationship between rotavirus epidemics and meteorological factors, some studies have been conducted with a linear Poisson regression model which has been widely used as one method of time-series analysis [4, 5, 8, 9]. Conversely, other studies interpreted the temporal variations of rotavirus infections by the Susceptible/Exposed/Infected/Recovered (SEIR) model, which is a well-known nonlinear dynamical system for epidemics of infectious diseases [13]. However, the linear Poisson regression model using random noise has a weakness in interpreting multiple periodicities with characteristic fluctuations caused by nonlinear dynamics. Our group has already proposed a newly devised method for time-series analysis, which allowed us to make the extremely precise determination of periodic structures of nonlinear time-series including a short data sequence [14–17].

The aim of the present study was to investigate the following two points for Kolkata during the period 2007–2009: (i) whether temporal patterns of rotavirus infections were inversely correlated with those of temperature, relative humidity and rainfall, as reported previously [11], and (ii) whether the occurrence of rotavirus infection in the monsoon season is associated with high temperature, as is the case of Bangladesh [4]. For aims (i) and (ii), we applied our method of analysis to the time-series data of rotavirus infections and meteorological conditions in Kolkata.

DATA

Rotavirus data

Surveillance data used in our study were collected at the Infectious Diseases and Beliaghata General Hospital (ID&BGH) in Kolkata between November 2007 and December 2009 [12]. For data collection, every fifth hospitalized patient with diarrhoea or dysentery without other associated illness on two randomly selected days in a week was enrolled, resulting in 45,004 patients. Of these 45,004 patients, stool specimens were collected from 2519 cases for aetiological study. In Figure 1, monthly data of cases enrolled via surveillance, positive cases of rotavirus from stool specimens, and isolation rate of rotavirus are illustrated. In the present study, the time-series analysis was conducted for monthly data of isolation rate of rotavirus gathered over 26 months from November 2007 to December 2009 (26 data points).

Meteorological data

Data on daily maximum and minimum temperature, relative humidity and rainfall were collected in the study region by the Meteorological Department, Kolkata [18], and the time span and number of data were identical to those for the rotavirus data. Daily average temperatures were computed as the mean of the daily maximum and minimum values. The monthly means for average temperature, relative humidity and total rainfall were calculated from daily records.

Season was categorized as pre-monsoon (March–May), monsoon (June–September), post-monsoon (October–November) and winter (December–February).

METHODS

We used the method of time-series analysis combined with spectral analysis based on the maximum entropy method (MEM) in frequency domain and least squares method (LSM) in time domain, as proposed in our previous work [14–17]. The time-series analysis revealed an association between rotavirus epidemics and meteorological factors, which are not visually apparent. For explanation of the present method of time-series analysis, we used the time-series data of the isolation rate of rotavirus shown in Figure 1.

Setting up the original data for analysis

The pre-processing of the original data of the isolation rate of rotavirus (Fig. 1) was conducted by the choice of an equal sampling interval, and the modified time-series data thus obtained are illustrated in Figure 2a. As a result, 26 data points of the original
data (Fig. 1) were translated into 25 data points in the modified data (Fig. 2a).

**Spectral analysis**

We assumed that the modified time-series data $x(t)$ (where $t =$ time) in Figure 2a are composed of systematic and fluctuating parts [19]:

$$x(t) = \text{systematic part} + \text{fluctuating part}.$$  \hspace{1cm} (1)

To investigate temporal patterns of $x(t)$ (Fig. 2a), we performed MEM spectral analysis, which is useful for investigating periodicities of short time-series, such as the time-series data used in the present study [14–17]. The MEM spectral analysis produces a power spectral density (PSD), from which we can obtain power representing the amount of amplitude of $x(t)$ at each frequency (note the reciprocal relationship between the scales of frequency and period). The formulation of MEM-PSD has been described previously [15]. In Figure 3a, PSD for the modified data (Fig. 2a) is shown.

**LSM**

The validity of the results of MEM spectral analysis was confirmed by calculation of the least squares fitting (LSF) curve to the modified data (Fig. 2a) with MEM estimated periods. In Figure 2a, the LSF curve calculated with six MEM estimated periodic modes is shown. The level of the reproducibility of the modified time-series data by the optimum LSF curve is evaluated by Pearson correlation coefficient ($\rho$) with SPSS version 17.0J (SPSS, Japan). Formulation of the LSF curve is described in the Appendix.

To investigate seasonality of the modified data (Fig. 2a), the LSF curve of the modified data was calculated with a seasonal cycle corresponding to a 1-year period ($T_1$), and is illustrated in Figure 4. By subtracting the LSF curve (Fig. 4) from the modified data (Fig. 2a), we obtained the residual time-series (Fig. 5a).

**Cross-correlation**

To investigate correlation of rotavirus infections with meteorological conditions (temperature, relative humidity, rainfall) in detail, we calculated the cross-correlation, $r$, between a series of rotavirus data and meteorological data. The calculation of a cross-correlation was conducted by using the residual time-series (Fig. 5a). Cross-correlation is a measure of similarity of two series as a function of a time-lag, $d$, applied to one of them. The formulation of cross-correlation is given in the Appendix.

**RESULTS**

**Temporal variations of rotavirus data and meteorological data**

Under the same procedure as used for analysis of the original data of the isolation rate of rotavirus (Fig. 1),
we obtained the modified time-series data for the meteorological conditions (temperature, relative humidity, rainfall). The modified time-series data thus obtained are illustrated in Figure 2(a–d).

Rotavirus infections (Fig. 2a) were observed throughout the year, and the frequency of detection was higher during winter and pre-monsoon (December–May) with peaks in February 2008 and March 2009, when low temperature, low relative humidity and low level of rainfall were observed (Fig. 2b–d). On the other hand, low values of the isolation rate of rotavirus (Fig. 2a) were observed in the monsoon and post-monsoon seasons (June–November), i.e. months of high temperature, high relative humidity and high level of rainfall (Fig. 2b–d). The seasonal pattern of temperature (Fig. 2b) indicates high values during pre-monsoon (March–May), monsoon (June–September) and post-monsoon (October–November) periods. Regarding relative humidity (Fig. 2c), the seasonal pattern indicates high values during monsoon and post-monsoon periods; a small peak was also observed in winter (December–February) in 2009. The seasonal pattern of rainfall (Fig. 2d) indicates two peaks with high values during the post-monsoon and monsoon periods.

Fig. 2. Comparison of the optimum modified time-series data and least squares fitting curve calculated with the six dominant periodic modes (—) with the modified data (- - -) for (a) isolation rate of rotavirus, (b) temperature, (c) relative humidity and (d) rainfall.
Spectral analysis and LSF analysis

PSDs for the modified time-series data (Fig. 2) are shown in Figure 3. Therein, in each PSD, the most prominent spectral line was observed at \( f = 1.0 \) [1/year]; frequency] corresponding to the 1-year cycle. In Figure 3a, a spectral peak in the low-frequency range of PSD \((f < 1.0)\) reflects longer-term oscillations that the 1-year cycle (2·38 years). Some distinct spectral peaks in the high-frequency range of PSD \((f > 1.0)\) reflect the shorter-term oscillations than the 1-year cycle. The reason for spectral peaks in the high-frequency range, \( f > 1.0 \), can be explained as follows: (i) shorter-term oscillations rather than the 1-year cycle superposing on the 1-year cycle, (ii) the harmonics corresponding to integer multiples (e.g. \( f = 2·0 \)) of the dominant spectral line at \( f = 1·0 \), and (iii) the superposition of (i) and (ii).

Six dominant spectral-peak frequency modes are shown in Table 1 with the corresponding periods and intensities (powers) of the spectral peaks. In Figure 2, each LSF curve calculated with six dominant periodic modes reproduces the modified data well. Thus, the periodic modes detected by MEM spectral analysis for each modified data (Fig. 3, Table 1) were confirmed to be appropriate. The good fitness of each LSF curve to the modified data was supported by the result that the values of \( r \) between the modified data and the LSF curve cover the high level: 0·96, 0·98, 0·96 and 0·97 of the rotavirus, temperature, relative humidity and rainfall data, respectively.

**Seasonal cycle of rotavirus data and metrological data**

The LSF curves calculated with the 1-year cycle \((T = 1·0)\) were normalized in amplitude, and were overlapped as shown in Figure 4. In the figure, the oscillation is shown in opposite phase between the LSF curve for rotavirus infection and the curve for meteorological condition (temperature, relative humidity, rainfall). The peaks of the LSF curves for temperature, rainfall and relative humidity data precede those of the rotavirus data by 5, 6 and 7 months, respectively. The value of \( r \) between the LSF curve of rotavirus data and that of each metrological data is strongly negative: \(-0·944\), \(-0·982\) and \(-0·751\) for temperature, rainfall and relative humidity, respectively.

**Shorter-term cycle of rotavirus data and metrological data**

In Figure 5, the residual data are shown for isolation rate of rotavirus, temperature, relative humidity and rainfall. The value of \( r \) is shown in Figure 6 for temperature, relative humidity, and rainfall. It is notable
Rotavirus infections and meteorological conditions

Fig. 4. Normalized least squares fitting (LSF) curves calculated with the 1-year cycle: isolation rate of rotavirus (%), temperature (°C), rainfall (mm) and relative humidity (%).

Fig. 5. Residual data obtained by subtracting the least squares fitting curves calculated with the 1-year cycle from the modified data for: (a) isolation rate of rotavirus, (b) temperature, (c) relative humidity and (d) rainfall.
that, in the case of temperature (Fig. 6a), the positive values of $r$ were observed at $d = 3$ and 4. This result indicates that relative high values of rotavirus infections also occur after 3 and 4 months of high temperature, although the opposite phase was observed between rotavirus and temperature data (Fig. 4). With regard to relative humidity and rainfall (Fig. 6b and 6c, respectively), this sort of tendency for $r$ values at $d = 3$ and 4 was not observed.

**DISCUSSION**

The significant result of our study was to demonstrate that the 1-year periodic mode for rotavirus data was correlated in an opposite phase to that for the meteorological data recorded from November 2007 to December 2009 in Kolkata, India (Fig. 4): i.e. rotavirus infections increase as the values of temperature, relative humidity and rainfall data decrease, and vice versa. It should be noted that our present result is comparable with that of Saha et al.’s study using data recorded from July 1979 to June 1981 in Kolkata [11], irrespective of today’s increasing availability to detect rotavirus (the ELISA test for Saha et al.’s study [11] and the polyacrylamide gel electrophoresis (PAGE) and silver staining tests in the present study [12]). For detection of rotavirus, specificity of PAGE was reported as 100%, higher than that of ELISA, although ELISA showed slightly higher sensitivity than PAGE [20]. Therefore, in our study, incidence of rotavirus may be more accurate than the earlier study, providing more appropriate conclusions in the analysis of the effects of meteorological conditions on rotavirus infections.

In our study, temporal patterns of rotavirus infections and meteorological conditions were elucidated by conducting a time-series analysis combined with MEM spectral analysis and LSM. As a result, we obtained valuable knowledge on the temporal structures of the time-series data which had remained unresolved until now. In previous studies, the correlation between rotavirus infections and meteorological conditions were investigated with the linear Poisson regression model [4, 5, 8, 9]. However, this model, which assumed the fluctuating part in equation (1) as

| Table 1. Characteristics of the six dominant spectral peaks shown in Figure 3 |
|---------------------------------|----------------|----------------|----------------|----------------|----------------|----------------|
|                                | f              | Period (year) | Power          | f              | Period (year) | Power          |
| Rotavirus                      | 0.43           | 2.38          | 17.01          | 0.31           | 3.36          | 0.45           |
| Temperature                    | 1.05           | 0.95          | 94.67          | 1.01           | 0.99          | 11.44          |
| Relative humidity              | 1.68           | 0.60          | 12.38          | 2.06           | 0.49          | 2.32           |
| Rainfall                       | 2.12           | 0.47          | 16.14          | 2.82           | 0.35          | 0.04           |
|                                | 2.83           | 0.35          | 18.79          | 4.98           | 0.20          | 0.18           |
|                                | 3.57           | 0.28          | 14.14          | 5.54           | 0.18          | 0.16           |

Fig. 6. Cross-correlations $(r)$ between the residual data of isolation rate for rotavirus and that for: (a) temperature, (b) relative humidity and (c) rainfall.
random noise, has a weakness for interpreting multiple periodic structures with characteristic fluctuations caused by nonlinear dynamics. On the other hand, the method of time-series analysis conducted in our study is applicable to any time-series without any restriction, and enables us to elucidate multiple periodicities of seasonal variations in detail (Table 1).

For the 1-year period listed in Table 1, the acrophase of LSF curves calculated with the 1-year cycle for temperature is observed in June and that for rainfall and relative humidity lag behind (July and August, respectively). This result might reflect a meteorological phenomenon of the monsoon season; as the temperature rises, there are increases in rainfall and relative humidity. Our result that the cross-correlation, $r$, between rotavirus infections and temperature (Fig. 6a) indicates the positive values at $d = 3$ and 4 might relate to the fact that, in Kolkata, the temperature exceeded 29 °C in April in 2008 and 2009 (Fig. 2b), and after 3–4 months, low values of isolation rates of rotavirus were observed (Fig. 2a). Thus, it can be considered that, in Kolkata, relatively high temperature in the monsoon season (June–September) is associated with the occurrence of rotavirus infection, as in the case of Bangladesh [4].

The reason for the occurrences of rotavirus infections in the monsoon season in Kolkata may be exerted through complex temperature-dependent pathways, as considered in the case of Bangladesh [4]. To understand the correlation of rotavirus infections with high temperature in Kolkata in detail, further accumulation of surveillance data of rotavirus infection will be necessary.

In the present study, asymptomatic infection of rotavirus was not investigated. Philips et al. [21] reported that the highest prevalence of asymptomatic rotavirus infection in England was in children aged <2 years (20–30%), with almost one-third of children in this age group being infected. There was no marked seasonality in the prevalence of asymptomatic rotavirus infection in children aged <5 years. In addition, Cox & Medley [22] reported that, in UK, anti-rotavirus IgM was seen in all age groups throughout the year with little obvious seasonal variation in the distribution of antibody levels, and this phenomenon is in contrast to the distinct seasonality of symptomatic rotavirus infections. With respect to the case of Kolkata, the epidemiological relationship between symptomatic and asymptomatic rotavirus infection remains unclear, and more work is needed to understand the role of seasonality in asymptomatic rotavirus infections.

Maximum peaks of rotavirus infections in the cold season (February 2008 and March 2009) in our study (Fig. 2) were similar to studies in other geographical areas of India [23–28]. Based on previous studies, two routes for rotavirus infection can be considered for the cold season, i.e. outdoor and indoor transmission. If an infection is transmitted outdoors, based on laboratory evidence [29–31], an important effect of low temperature might be the prolonged survival of the virus in the environment, increasing the possibility of exposure to infectious virus. On the other hand, if an infection is transmitted indoors, based on a case-control study conducted in England and Wales [32], one possible interpretation for the cold weather effect on rotavirus infections could be low ambient temperatures encouraging individuals to stay indoors, thereby increasing their exposure to contaminated air and surfaces.

For the route of transmission of rotavirus infections in the cold season in Kolkata, one possibility is that, based on strong evidence that rotavirus is a waterborne pathogen, the increase of rotavirus infections in winter results from a lack of clear water because of low rainfall (Fig. 2a, d). However, the following three reasons suggest that water alone may not be responsible for all rotavirus transmission [10]: (i) the high rates of infection in the first 3 years of life regardless of sanitary conditions, (ii) the failure to document faecal–oral transmission in several outbreaks of rotavirus diarrhoea, and (iii) the marked spread of rotavirus over large geographical areas during winter in temperature zones. On the other hand, airborne spread of aerosolized particles may be responsible for the seasonal pattern of rotavirus disease in Kolkata [10], and wind might play a role in dispersal of the aerosolized particles [28].

With a mechanical force such as rotavirus transmission in India, it is possible that wind plays an important role in epidemic patterns which can affect the temporal patterns of rotavirus epidemics, in accord with a study of rotavirus in Pune, India [28]. In order to understand underlying cause of rotavirus in India and to use the results obtained effectively for health service, it is necessary to conduct a systematic study to quantify the impact of wind on rotavirus epidemic pattern for the whole of India. It is expected that the present method of time-series analysis will contribute to the investigation of the temporal patterns of wind and the estimation of its correlation with rotavirus infections.

https://doi.org/10.1017/S0950268812002208 Published online by Cambridge University Press
epidemics, as well as for temperature, relative humidity and rainfall, as conducted in the present study.

APPENDIX

LSF calculation

LSF calculations are performed using an appropriate function:

\[ X(t) = a_0 + \sum_{n=1}^{N_p} \left\{ a_n \sin(2\pi f_n t) + b_n \cos(2\pi f_n t) \right\}, \quad (A1) \]

where \( f_n \) (\( = 1/T_n \); where \( T_n \) = the period) is the frequency of the \( n \)th component, \( a_0 \) a constant, \( a_n \) and \( b_n \) \((n = 1, 2, 3, \ldots, N_p)\) the amplitude, and \( N_p \) the total number of components. The optimum values of parameters \( a_0, a_n \) and \( b_n \) \((n = 1, 2, 3, \ldots, N_p)\) in equation (A1), with the exception of \( N_p \), are determined exactly from \( T_n \) estimated by MEM spectral analysis.

Cross-correlation

For a series of residual data for rotavirus \( x(i) \) and for the meteorological condition \( y(i) \), where \( i = 0, 1, 2, \ldots, N - 1 \) \((N \) is the total number of the series), the cross-correlation \( r \) at delay \( d \) is defined as

\[ r(d) = \frac{\sum_{i=0}^{N-1} (x(i) - m_x)(y(i-d) - m_y)}{\sqrt{\sum_{i=0}^{N-1} (x(i) - m_x)^2 \sum_{i=0}^{N-1} (y(i-d) - m_y)^2}}, \quad (A2) \]

where \( m_x \) and \( m_y \) are the means of the corresponding series. In the present study, equation (A2) is computed for \( d = 0, 1, 2, 11 \). We set the initial data point of the first series \((x(0) \) and \( y(0))\) as November 2008 \((N = 15)\). When the index in the series is \( < 0 \) \((i - d < 0)\), the data point of \( y(i-d) \) corresponds to \( d \) month previous points from \( y(0) \) (November 2008).

ACKNOWLEDGEMENTS

This study was supported in part by a Grant-in-Aid of Scientific Research (Grant no. 22406017) and the Japan Initiative for Global Research Network on Infectious Diseases (Okayama University, Japan – National Institute of Cholera and Enteric Diseases, India) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

DECLARATION OF INTEREST

None.

REFERENCES

15. Ohtomo K, et al. Relationship of cholera incidence to El Niño and solar activity elucidated by


