

## Prediction of epidemic cholera due to *Vibrio cholerae* O1 in children younger than 10 years using climate data in Bangladesh

F. MATSUDA<sup>1</sup>, S. ISHIMURA<sup>2</sup>, Y. WAGATSUMA<sup>3</sup>, T. HIGASHI<sup>4</sup>, T. HAYASHI<sup>5</sup>,  
A. S. G. FARUQUE<sup>6</sup>, D. A. SACK<sup>6</sup> AND M. NISHIBUCHI<sup>7\*</sup>

<sup>1</sup> Graduate School of Medicine, Kyoto University, Kyoto, Japan

<sup>2</sup> Laboratory of Mathematics, Tsurumi University School of Dental Medicine, Yokohama, Japan

<sup>3</sup> Department of Epidemiology, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Japan

<sup>4</sup> Department of Epidemiology and Healthcare Research, Kyoto University School of Public Health, Kyoto, Japan

<sup>5</sup> Disaster Prevention Research Institute, Kyoto University, Kyoto, Japan

<sup>6</sup> Clinical Science Division, International Centre for Diarrhoeal Disease Research Bangladesh, Dhaka, Bangladesh

<sup>7</sup> Center for Southeast Asian Studies, Kyoto University, Kyoto, Japan

(Accepted 26 January 2007; first published online 8 March 2007)

### SUMMARY

To determine if a prediction of epidemic cholera using climate data can be made, we performed autoregression analysis using the data recorded in Dhaka City, Bangladesh over a 20-year period (1983–2002) comparing the number of children aged <10 years who were infected with *Vibrio cholerae* O1 to the maximum and minimum temperatures and rainfall. We formulated a simple autoregression model that predicts the monthly number of patients using earlier climate variables. The monthly number of patients predicted by this model agreed well with the actual monthly number of patients where the Pearson's correlation coefficient was 0.95. Arbitrarily defined, 39.4% of the predicted numbers during the study period were within 0.8–1.2 times the observed numbers. This prediction model uses the climate data recorded 2–4 months before. Therefore, our approach may be a good basis for establishing a practical early warning system for epidemic cholera.

### INTRODUCTION

The World Health Organization (WHO) reported in 2005 that diarrhoeal diseases accounted for 18% of the deaths among children aged <5 years [1]. Although the deaths caused by diarrhoeal diseases are gradually decreasing, they are still a significant public

health problem especially for children in developing countries.

Cholera is an important diarrhoeal disease and is considered to be a disease of poverty. It is a major health problem for those living in places where the public health system is underdeveloped or collapsed. However, the case-fatality rate of cholera could become less than 1% if prompt and adequate medical measures are taken [2]. Therefore, detecting or predicting an epidemic as early as possible would be helpful for public health authorities to prepare for disease intervention and prevention. Previous studies

\* Author for correspondence: Dr M. Nishibuchi, Center for Southeast Asian Studies, Kyoto University, 46 Shimoadachi-cho, Yoshida, Sakyo-ku, Kyoto, Kyoto 606-8501, Japan.  
(Email: nisibuti@cseas.kyoto-u.ac.jp)

demonstrated a strong correlation between climate and infectious diarrhoeal diseases including epidemics of cholera [3–11]. The WHO emphasizes the importance of an early warning system for infectious disease epidemics based on climate data [12]; however, studies to establish early warning systems of practical use are still being assessed.

Cholera is endemic in Bangladesh where there are two epidemic seasons per year: a spring epidemic around April (just before the monsoons) and an autumn epidemic occurring during September–December (at the end of the monsoons) [13]. This seasonality suggests some climatic factor(s) may be associated with epidemic cholera. In addition, there is an inter-annual variation in the scale of the epidemic cholera and the pattern of the seasonal cycle in the Bengal area spanning Bangladesh and India [9]. Sea-water temperature variation in the Pacific caused by the El Niño Southern Oscillation (ENSO) plays a major role in the inter-annual variation, including the seasonal modulation of cholera in Bangladesh [6, 9, 10]. The sea surface temperature in the Bay of Bengal is a key mediating factor for the ENSO effect [8, 9] and affects a complex ecology system controlling the persistence and proliferation of *Vibrio cholerae*, virulent strains of which are responsible for cholera, in its natural habitat, brackish water and the estuarine environment [3, 14–17]. Further studies concerning the regional variables mediating the ENSO effect locally are necessary [11].

We hypothesized local climatic variables may correlate with and precede the local sea surface temperature variation and could be useful for an early warning system for cholera. Local atmospheric temperature affects local water temperature in Bangladesh and ultimately mediates the ENSO effect [6]. Rainfall is one of the mediating factors for the ENSO effect and can affect concentration of the pathogen in the aquatic environment and human exposure to the pathogen [6, 11]. Atmospheric temperature data and rainfall data are usually available from local authorities. We evaluated whether these simple local climatic variables could be useful predictors for the number of cholera patients in Bangladesh using a powerful statistic, regression analysis. We focused on the number of patients due to *V. cholerae* O1 infection (hereafter abbreviated as O1 cholera patients) in children aged <10 years. Case rates were highest among the 2–9 or 5–9 years age groups in previous epidemiological studies carried out in Bangladesh [18, 19]. We presume the <10 years age

group more precisely reflects the scale of exposure to *V. cholerae* than do persons in other age groups who are mostly asymptomatic due to acquired immunity. We studied cholera cases due to *V. cholerae* serovar O1 because this serotype was a major cause of epidemic cholera among children throughout the study period in Dhaka City.

## METHODS

### Patient data

We obtained the daily data of O1 cholera patients during the 20-year period from 1983 to 2002 from International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B). Data of the surveillance for aetiologies of diarrhoea at Dhaka Hospital of ICDDR,B have been recorded since 1979 [20]. Diarrhoeal stool from every 25th patient until 1995 and every 50th patient from 1996 onwards was examined for aetiologies and recorded at ICDDR,B. We extracted the data for O1 cholera patients aged <10 years, and estimated the monthly total number of O1 cholera patients aged <10 years. The monthly patient data were used for the statistical analysis.

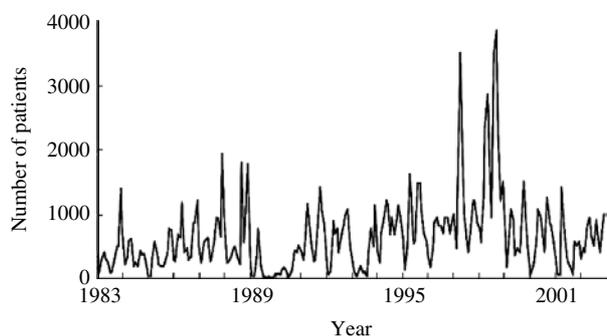
### Climate data

We obtained the daily maximum temperature, minimum temperature, and rainfall in Dhaka City that were recorded by the Bangladesh Meteorological Department during the period from 1983 to 2001. When the daily data were missing, we interpolated the data by calculating the means from the daily data before and after the missing day(s). We calculated the monthly average for the maximum temperature in Celsius (MAXT), minimum temperature in Celsius (MINT), and the monthly total rainfall in millilitres (RAIN) from the daily data; and used these monthly data for the following analyses.

### Statistical analysis

Spectral analysis was performed to examine the periodicity in the number of monthly O1 cholera patients.

Autoregression analysis was used to examine the association between lagged climate data using MAXT, MINT, and RAIN and number of O1 cholera patients aged <10 years. The time lag of the climate data was termed 'lag 1' meaning 1 month preceding the



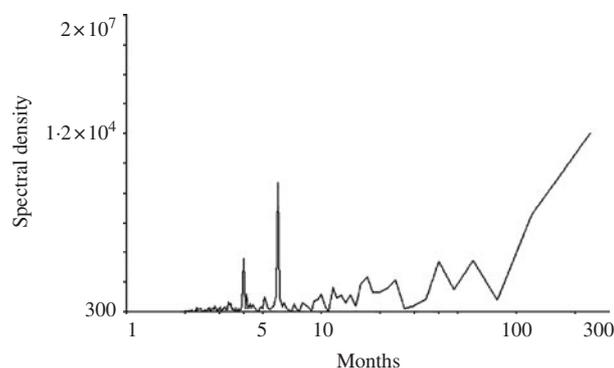
**Fig. 1.** Number of monthly O1 cholera patients aged <10 years in Dhaka City between 1983 and 2002.

original patient data and was independently set for each of the three climate variables. The range of the climatic lag was determined based on the result of the spectral analysis which was carried out before the autoregression analysis. Each of the three independent climate data at different lag times was combined to create the lagged climate datasets (e.g. lag 3 for MAXT, lag 1 for MINT and lag 4 for RAIN). The lagged climate datasets were compared to the patient data for autoregression. Fitness of the autoregression model was evaluated based on the *P* values of the coefficients of the equation and Pearson’s correlation coefficient between the number of patients predicted by the autoregression model and the observed number of patients. A *P* value of  $\leq 0.05$  for the coefficient was considered to be significant. We also evaluated the predictability of the scale of the epidemic as follows. The prediction was arbitrarily defined as ‘well-predicted’ if the number of patients predicted at a time-point by the autoregression model was within the range of 0.8–1.2 times the number of observed patients.

Statistical analyses were performed using the software program SPSS trends 11.5 J (SPSS Inc., Chicago, IL, USA).

**RESULTS**

Figure 1 shows the time-series change in the monthly number of O1 cholera patients aged <10 years from 1983 to 2002. The total sample size of the subjects studied over the 20-year period was 153 775. There was no visually detectable trend in the time-series change during this study period. The unusually large numbers of patients in 1997 and 1998 are considered to be due to the deterioration of sanitary conditions after floods around Dhaka City in these years. The



**Fig. 2.** Result of the spectral analysis for the monthly numbers of O1 cholera patients aged <10 years in Dhaka City between 1983 and 2002.

spectral analysis shows a 6- to 7-month periodicity in the numbers of monthly O1 cholera patients during the study period (Fig. 2). This indicates the peak of O1 cholera epidemics occur nearly every half year. Therefore, we set time lags for the climate variables from 0 to 6, i.e. 0, 1, 2, 3, 4, 5, and 6; and created new time-series data using the combination of each of the lagged climate data. A total of 343 ( $=7^3$ ) combinations of lagged climate data were created and tested against the observed patient data using autoregression. During the autoregression analysis, we noted when we set ‘lead 1’, which means the patient data was shifted 1 month to the past (opposite of the lag), we could obtain very significant Pearson’s correlation coefficients between the numbers of patients predicted by the autoregression model and the real numbers of observed patients. Therefore, we decided to perform the analysis using patient data with lead 1. Both Pearson’s correlation coefficients and *P* values for all of the coefficients of the autoregression equations were significant for the combination of climate variables listed in the Table. Residuals of the autoregression were mostly white noise (data not shown). Of the combinations of climate variables listed in the Table, the combination MAXT (lag2) – MINT (lag4) – RAIN (lag3) was judged to be the most suitable for the prediction model because the Pearson’s correlation coefficient was the highest and the lags were in the range for practical use (discussed in detail below). The equation to predict the monthly number of O1 cholera patients aged <10 years (lead 1) using this lagged climatic combination is as follows:

$$-56.2 [\text{MAXT} (\text{lag}2)] - 23.6 [\text{MINT} (\text{lag}4)] + 0.53[\text{RAIN} (\text{lag}3)] + 2790 + r(t),$$

Table. Combinations of lagged climate data showing significant  $P$  values of the coefficients of the autoregression equation and high Pearson's correlation coefficient between the numbers of patients predicted by the autoregression model and the observed patient numbers set at lead 1

Lag of climate variable			Coefficients of the autoregression equation for				Primary autocorrelation coefficient	Pearson's correlation coefficient
MAXT	MINT	RAIN	MAXT	MINT	RAIN	Constant		
5	4	0	84.0*	-34.8†	0.42†	-1260†	0.646*	0.920
5	1	4	69.2*	31.6†	-0.46†	-2100*	0.660*	0.928
5	0	1	56.8*	48.4*	-0.48†	-2070*	0.654*	0.910
3	5	0	-64.9*	27.3†	0.71†	1930*	0.673*	0.923
3	0	2	-63.6*	24.6†	0.44†	1990*	0.665*	0.923
2	4	3	-56.2*	-23.6†	0.53†	2790*	0.663*	0.952
2	2	0	-108.2*	47.1†	0.56†	2850	0.668*	0.930
0	6	0	69.9*	28.9†	0.56†	-2240*	0.650*	0.919
0	5	0	83.6*	48.0*	0.70†	-3110*	0.661*	0.892
0	4	0	80.2*	23.5†	0.45†	-2420*	0.654*	0.926

MAXT, Maximum temperature; MINT, minimum temperature; RAIN, rainfall.

\*  $P < 0.001$ , †  $P \leq 0.05$ .

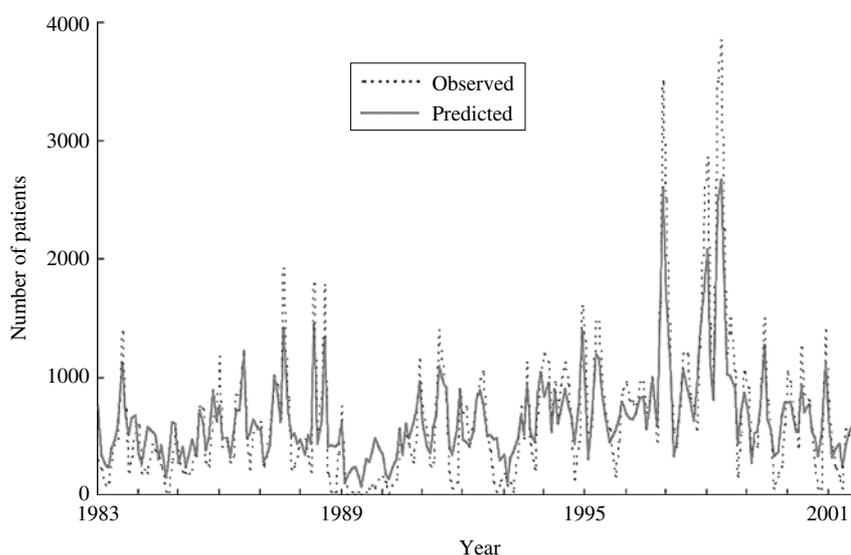
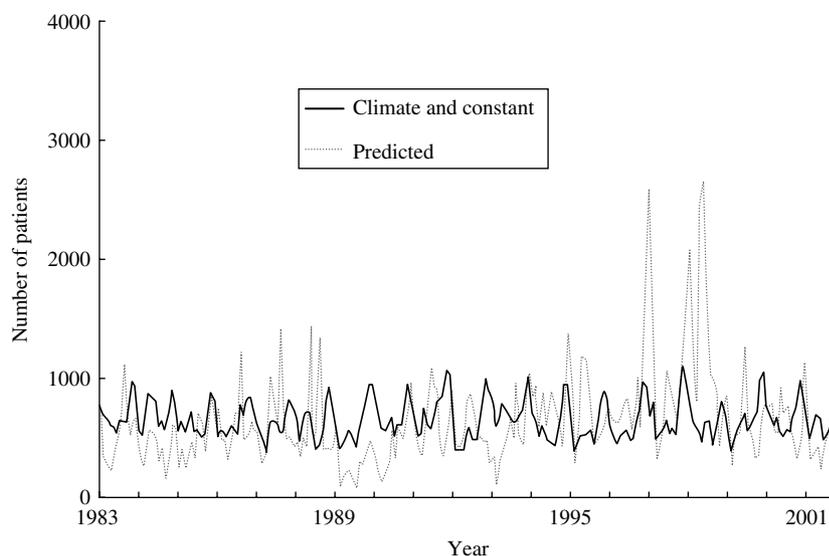


Fig. 3. Comparison of the monthly numbers of predicted and observed O1 cholera patients aged <10 years between 1983 and 2002. The numbers of predicted patients were calculated by the autoregression equation explained in the text.

where  $r(t) = 0.663 [r(t-1)] + u(t)$ ,  $r(t)$  is the residual, and  $u(t)$  is the white noise.

Figure 3 shows a graphic comparison of the numbers of patients predicted by this equation and the actual numbers of observed patients. As shown in this Figure, the tendency of the change in the numbers of predicted patients, i.e. seasonal variation, agrees well with the numbers of observed patients. When we examined the difference between the numbers of the observed and predicted patients at a time-point by the standard we arbitrarily set (explained in the Methods

section), the number of patients was judged to be 'well predicted' on 39.4% of the all time-points. To evaluate the contribution of the climate variables and the effect of the residual in the above equation to the numbers of predicted patients, we graphically compared the number of patients predicted by the equation (Fig. 4, grey dotted line) and those by the modified equation where the residual,  $r(t)$ , was removed from the original equation (Fig. 4, black line). The difference between the two lines is not very small. This is discussed below.



**Fig. 4.** Comparison of the monthly numbers of O1 cholera patients aged <10 years between 1983 and 2002 predicted by the autoregression equation with (.....) and without (—) the residual.

## DISCUSSION

This work demonstrates an approach to formulate a simple autoregression model to predict the numbers of monthly cholera patients using a lag in monthly climate data. By setting the lead to 1 for the patient data, the autoregression model using the combinations of climate variables listed in the Table was found to predict the number of patients significantly (Pearson's correlation coefficient >0.910). If the combination containing lag 0 (climate data collected in the current month) in any of the climate variables is used to predict the number of patients at lead 1 (number of patients occurring next month), this means the epidemic may start 1 month after the climate data are collected. This short prediction gives no time for the authorities to prepare for an epidemic. For this reason, we judged the combination  $MAXT$  (lag2) –  $MINT$  (lag4) –  $RAIN$  (lag3) is practical for prediction. Climate data collected during the preceding 2–4 months could be used to predict the numbers of patients during the next month. There is sufficient lead time to allow practical mobilization of medical resources.

$MAXT$ ,  $MINT$ , and  $RAIN$  values ranged from 22.8 °C to 36.5 °C, from 11.3 °C to 28.1 °C, and from 0 mm to 707 mm (data not shown) during the study period and their coefficient ranges were –108.2 to 84, –34.8 to 48.4, and –0.48 to 0.71, respectively (Table). This shows the contribution of  $RAIN$  for the prediction is

the smallest and  $MAXT$  is the most important of the three climatic variables. There is a debate concerning the role of rainfall contributing to epidemic cholera [11]. Our data suggests the contribution of rainfall is minor or may not affect epidemic cholera. The lag in the climate variables shown in the Table varied considerably. This shows the  $MAXT$  value must be high when the coefficient of  $MAXT$  is positive to predict a large number of patients during an epidemic season. This can happen when the lag is 0 or 5 (Table). If the coefficient of  $MAXT$  is negative, the  $MAXT$  value must be small to predict a large number of patients. This can happen when the lag is 2 or 3 (Table). Our spectral analysis shows, as previously reported, the peak of O1 epidemic cholera occurs nearly every half year (Fig. 2). Assuming the increased atmospheric temperature mediates the ENSO effect to increase the sea surface temperature, then the predicted increase in the number of patients showing a lag 0 or 5 may correspond to the epidemic season and a lag 2 or 3 corresponds to the intermediate between the epidemic seasons.

The equation we selected to closely fit observation was the one that exhibited the highest Pearson's correlation coefficient (0.952). The equation states the lower the  $MAXT$  (2 months before) and  $MINT$  (4 months before) and the higher the  $RAIN$  (3 months before) were, the higher the number of cholera patients 1 month after the present time will be. This equation predicts well the seasonal variation of epidemics

depicted in Figure 3, however, it could only accurately predict the scale of epidemics at 39.4% of all the time-points according to our arbitrarily determined standard.

We conclude the beginning of an epidemic can be predicted using the climate variables; but the climate data are not precise enough to estimate the scale of the epidemics. Figure 4 shows, as previously reported, the fluctuation in the numbers of the patients due to the three climate variables remained within a narrow range (about 500) and suggests the scale of epidemics largely depends on other factors [ $r(t-1)$  in the residual of the equation]. Climate change, e.g. high temperature, triggers proliferation of virulent strains of *V. cholerae* in the natural environment. A virulent strain when propagated to a large number causes infection in a human community where its spread may largely depends on factors not directly related to climate factors. These include poor hygiene, overcrowding, and herd immunity. Abrupt breakdown of sanitation due to civil wars and natural disasters like flood can also lead to large epidemics. As such, it is not easy to predict the disasters in advance by consulting climate variables.

Our investigation has the following limitations. First, we did not include population change dynamics in the Dhaka target population. We used the absolute numbers of O1 cholera patients aged <10 years at a hospital in Dhaka City, but we did not correct it for the population change in this area for the following reasons. We roughly estimated the population change in the children aged <10 years in Dhaka City; and concluded this would not significantly influence the result of our regression analysis during the study period (data not shown). Therefore, we avoided correcting the original data using a population estimation method that may not be reliable. Second, we presumed all epidemics that peak in the spring and autumn were caused by an unknown but identical mechanism; however, we have not identified this mechanism. Finally, we did not include non-climate factors in the prediction model since they are outside of the scope of this study. Other reports suggest it is important to include non-climate factors and factors that are indirectly affected by climate to improve the prediction model [6, 11, 21–23]. These factors include herd immunity, nutrition, seasonal change in community behaviour, sanitary conditions, socioeconomic conditions, and bacterial change by mutation. However, each of these factors is difficult to define in absolute numbers and

thus would be difficult to use in a mathematic formula.

In conclusion, our study does demonstrate simple climate variables such as atmospheric temperature and rainfall may be useful climate variables for early warning of epidemic cholera and may be a good basis for further regression using improved data collection to include the other factors. The most conservative interpretation of this study is that this conclusion applies only to the O1 epidemic among the children aged <10 years in Dhaka City, but this concept may be expanded to include the other aetiological agents (serovar O139), other age groups, and other localities in future studies.

## ACKNOWLEDGEMENTS

This study was supported by Grants-in-Aid for Scientific Research of the Education Exploratory Research ‘Statistical research on the relationship between prevalence of epidemic diseases and meteorological disasters: Case study in Bangladesh (15P-1)’, based on the agreement for joint research between the Disaster Prevention Institute, Kyoto University, and ICDDR,B; and was also funded by The Waksman Foundation of Japan Inc., and The Yakult Research Foundation for Bioscience.

## DECLARATION OF INTEREST

None.

## REFERENCES

1. **World Health Organization.** The World Health Report 2005. ([http://www.who.int/whr/2005/whr2005\\_en/](http://www.who.int/whr/2005/whr2005_en/)). Accessed 12 July 2005.
2. **World Health Organization.** Water and sanitation related diseases fact sheets (cholera), 2004 ([http://www.who.int/water\\_sanitation\\_health/diseases/cholera/en/print.html](http://www.who.int/water_sanitation_health/diseases/cholera/en/print.html)) Accessed 19 May 2006.
3. **Colwell RR.** Global climate and infectious disease: The cholera paradigm. *Science* 1996; **274**: 2025–2031.
4. **Salazar-Lindo E, et al.** El Niño and diarrhea and dehydration in Lima, Peru. *Lancet* 1997; **350**: 1597–1598.
5. **Speelman EC, et al.** Cholera incidence and El Niño-related higher ambient temperature. *Journal of the American Medical Association* 2000; **283**: 3072–3074.
6. **Pascual M, et al.** Cholera dynamics and El Niño–Southern Oscillation. *Science* 2000; **289**: 1766–1769.
7. **Checkley W, et al.** Effects of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *Lancet* 2000; **355**: 442–450.

8. **Lobitz B, et al.** Climate and infectious disease: Use of remote sensing for detection of *Vibrio cholerae* by indirect measurement. *Proceedings of the National Academy of Sciences USA* 2000; **97**: 1438–1443.
9. **Bouma MJ, Pascual M.** Seasonal and interannual cycles of endemic cholera in Bengal 1891–1940 in relation to climate and geography. *Hydrobiologia* 2001; **460**: 147–156.
10. **Rodó X, et al.** ENSO and cholera: a nonstationary link related to climate change? *Proceedings of the National Academy of Sciences USA* 2002; **99**: 12901–12906.
11. **Pascual M, Bouma MJ, Dobson AP.** Cholera and climate: revisiting the quantitative evidence. *Microbes and Infection* 2002; **4**: 237–245.
12. **World Health Organization.** Using climate to predict disease outbreaks: review (<http://www.who.int/globalchange/publications/oeh0401/en/index2.html>). Accessed 12 July 2005.
13. **Longini IM, et al.** Epidemic and endemic cholera trends over a 33-year period in Bangladesh. *Journal of Infectious Diseases* 2002; **186**: 246–251.
14. **Huq A, et al.** Influence of water temperature, salinity, and pH on survival and growth of toxigenic *Vibrio cholerae* serovar O1 associated with live copepods in laboratory microcosms. *Applied and Environmental Microbiology* 1984; **48**: 420–424.
15. **Epstein PR.** Algal blooms in the spread and persistence of cholera. *Biosystems* 1993; **31**: 209–221.
16. **Islam MS, Drasar BS, Sack RB.** Probable role of blue-green algae in maintaining endemicity and seasonality of cholera in Bangladesh: a hypothesis. *Journal of Diarrhoeal Disease Research* 1994; **12**: 245–256.
17. **Colwell R, Huq A.** Marine ecosystems and cholera. *Hydrobiologia* 2001; **460**: 141–145.
18. **Hughes JM, et al.** Epidemiology of eltor cholera in rural Bangladesh: importance of surface water in transmission. *Bulletin of the World Health Organisation* 1982; **60**: 395–404.
19. **Faruque ASG, Fuchs GJ, Albert MJ.** Changing epidemiology of cholera due to *Vibrio cholerae* O1 and O139 Bengal in Dhaka, Bangladesh. *Epidemiology and Infection* 1996; **116**: 275–278.
20. **International Centre for Diarrhoeal Disease Research, Bangladesh.** Trends in aetiologies for diarrhoeal diseases, Dhaka [Editorial]. *Health and Science Bulletin (Bangladesh)* 2002; **1**: 12–15 (<http://202.136.7.26/>). Accessed 6 December 2006.
21. **Huq A, et al.** Critical factors influencing the occurrence of *Vibrio cholerae* in the environment of Bangladesh. *Applied and Environmental Microbiology* 2005; **71**: 4645–4654.
22. **Faruque SM, et al.** Self-limiting nature of seasonal cholera epidemics: role of host-mediated amplification of phage. *Proceedings of the National Academy of Sciences USA* 2005; **102**: 6119–6124.
23. **Koelle K, et al.** Refractory periods and climate forcing in cholera dynamics. *Nature* 2005; **436**: 696–700.