J HEALTH POPUL NUTR 2004 Mar;22(1):27-33 
ISSN 1606-0997 $ 5.00+0.20

# Time Series Analysis of Deaths Due to Diarrhoea in Children in Rio de Janeiro, Brazil, 1980-1998

**Pauline Lorena Kale[1], Valeska Lima Andreozzi[2], and Flávio Fonseca Nobre[3]**

*[1]Medicine School/Institute of Studies of Public Health-NESC, Federal University of Rio de Janeiro, Brazil, [2]National School of Public Health, Oswaldo Cruz Foundation, Brazil, and [3]Biomedical Engineering Program/COPPE, Federal University of Rio de Janeiro, Brazil*

## ABSTRACT

Diarrhoea is still a considerable public-health problem in developing countries, especially among children aged less than five years. The well-known relationship between seasonal variation and aetiological agents of diarrhoeal diseases helps inform the decisions about the prevention and control measures. The aim of this study was to identify the temporal patterns of deaths due to diarrhoea in children, aged less than five years, in Rio de Janeiro State from 1980 to 1998. The study analyzed data on monthly deaths due to diarrhoea; the data were supplied by the Natural Mortality Information System (NMIS, Ministry of Health of Brazil). Auto Regressive Integrated Moving Average modelling was applied to the data. Using this model, the huge decline and the winter peaks could be highlighted, and they may express the predominance of rotavirus as the aetiology of diarrhoeal deaths, suggesting that an effective vaccine is the main measure for the prevention and control of severe diarrhoea.

***Key words***: Diarrhoea; Times series; Epidemiology; Seasonality; Brazil

## INTRODUCTION

Diarrhoea still remains one of the most common diseases during childhood. Approximately, an estimated one billion episodes and 3.3 million deaths occur each year worldwide among children aged less than five years (1). Mortality due to diarrhoea is to a great extent preventable. Since the early 1980s, mortality and hospital admissions due to diarrhoea in Brazil have been decreasing due to the extensive use of oral rehydration therapy (ORT) within the primary healthcare network. Changes in factors other than ORT, such as improvements in socioeconomic conditions, water supply and vaccine coverage, nutritional status and breast-feeding duration, led to an almost 21% reduction in infant mortality due to diarrhoea (2). Despite this, diarrhoeal disease is still a considerable public-health problem in developing countries, particularly among children aged less than five years (3).

The relationship between specific aetiologic agents of diarrhoeal diseases and seasonal rainfall has been described worldwide. While rotavirus infections are commonly associated with ‘winter’ diarrhoea, most bacterial diarrhoea predominates during warm and wet seasons (4). In the USA, the peak season of rotavirus infection starts in the Southwest in autumn and moves to the Northeast by Spring (5). The pronounced reduction in mortality due to diarrhoea was accompanied by changes in its seasonal patterns following the shift of the mean enteric pathogen, possibly from *Escherichia coli* to rotavirus (6).

Rotavirus is the most common cause of fatal acute watery diarrhoea among children in the world (7) and particularly in developing countries where it accounts for 600,000-870,000 infant deaths per year (8). Undernutrition seems to play a large role in the mortality rates due to rotavirus infection as distressed children face more severe symptoms when infected (9). In the USA,

Correspondence and reprint requests should be addressed to: Dr. Pauline Lorena Kale
Rua Professor Miguel Couto
344/302 Bloco B
Icaraí–Niterói, Rio de Janeiro
Brazil
CEP: 24.230.240
Email: pkale@nesc.ufrj.br
Fax: 55 21 26115313

approximately 2.7 million children aged less than five years are affected by diarrhoea due to rotavirus each year, resulting in 500,000 physician-visits and 50,000 hospitalizations at an estimated cost of US$ 274 million in medical care and more than US$ 1 billion in societal costs (9). Its seasonality closely resembles that of other viruses causing childhood illnesses, such as measles, that are spread by the respiratory route (10). Since the magnitude of the disease associated with rotavirus infections and public-health interventions, such as clean water and improved sanitation, are unlikely to decrease its incidence, vaccines are considered the main strategy for prevention (11). Effective vaccines have the potential to reduce morbidity and mortality due to rotavirus infection. In developed countries, the tetravalent reassortant vaccine (RR-TV) had 65-100% protection efficacy against severe disease and, in less-developed countries, it has been more variable (11). In northern Brazil, RR-TV was shown to be 80% protective against very severe gastroenteritis (8). The tetravalent rotavirus vaccine was licensed in the USA in 1998, but its routine usage has been suspended because of its association with a rare type of bowel obstruction, called intussusception (12).

Particular temporal patterns of occurrence of diarrhoea may be associated with different prevailing aetiological agents; such information may be appropriately used for supporting the decisions about prevention and control measures (10). Effects of trend, seasonality, and cycle variations on diarrhoeal disease can be estimated through analysis of time series or historical series—a sequence of data obtained at regular intervals of time during a specific period. The aim of this research was to identify the temporal patterns of mortality due to diarrhoea among children, aged less than five years, in Rio de Janeiro State, Brazil, from January 1980 to December 1998.

## MATERIALS AND METHODS

### Data

All deaths due to diarrhoea from 1980 to 1998—International Classification of Diseases (ICD), 9th revision, codes 001-009 up to 1995 and codes A01-A09 ICD 10th revision from 1996 towards—occurred among children, aged less than five years, who lived in the State of Rio de Janeiro, Brazil, were selected for this study. However, 99% of deaths had non-specific aetiology (cause unspecified, but presumed infectious). The time series analysis corresponds to monthly number of deaths due to diarrhoea from January 1980 to December 1998. Data were supplied by the Natural Mortality Information System (NMIS) through the Internet (http://www.saude.rj.gov.br) and CD-ROM (DATASUS, Ministry of Health of Brazil). It is expected that the data represent the total number of deaths from severe diarrhoea among children aged less than five years in Rio de Janeiro State, since the NMIS coverage in Rio de Janeiro State is almost 100%.

### Modelling of time series

Auto Regressive Integrated Moving Average (ARIMA–Box and Jenkins) modelling was applied to the mortality time series. This method is based on the existence of autocorrelation within the series and tries to approach the random process, which generates the observations, and to isolate some deviations, which are not directly observable. It can be applied to discrete and continuous data, equally spaced time intervals. Construction of an adequate ARIMA model requires a minimum of about 50 observations and also requires a stationary time series (the mean and variance of observations remain constant over time).

An ARIMA model can include two sets of autoregressive (AR) and moving average (MA) terms (or parameters) of any order. AR terms relate the observation made at time $t$ in the series to the observation made at time $t$-1 (first order), $t$-2 (second order), and so on. MA terms relate the error—the difference between the observed and estimated values—at time $t$ to the error at times $t$-1, $t$-2, etc. Both the sets can also include seasonal terms. For instance, an observation at month $t$ can be associated with an observation at the same month one year earlier.

The Box-Jenkins models have four steps: identification of the ARIMA model, estimation of the parameters laid down in the identification, diagnostic checking of the model, and once diagnostic checking has been deemed satisfactory, prediction step.

### Identification process

For adequate ARIMA modelling, a time series should be stationary with respect to mean and variance. If the mean increases or decreases over time, or if the variance does (as indicated by the excursions around the mean becoming smaller or larger over time), the series may need to be transformed to make it stationary, before being modelled. If necessary to stabilize the mean, the series can be ‘differenced’. In the presence of a secular trend in the time series, regular differencing is indicated: each observation is replaced by the difference between it and the previous observation. In the presence of clear seasonal variations, seasonal differencing is indicated: each observation is replaced by the difference between it and the observation a year before.

The autocorrelation function (acf) and the partial autocorrelation function (pacf) of the transformed series are estimated to check the existence of temporal dependence between ordered pairs of observations separated by lags (the time between two observations).

Autocorrelations are the correlations between each observation and the previous ones (lag one) or the previous observation but one (lag two) and so on. Partial autocorrelations are the same, except that the effect of the intervening observation(s) is removed. Clearly, for lag 1, since there is no intervening observation, the autocorrelation and the partial autocorrelation are the same. The size of autocorrelations and partial autocorrelations of various lags guides the selection of terms to include in the initial ARIMA model. When a strong serial correlation at seasonal lag is presented, a seasonal ARIMA model (SARIMA) can be derived.

### Estimation

The method of maximum likelihood can be used for estimating the parameters of a SARIMA model. It provides a unified and practicable estimation tool for models of arbitrary order. Several statistical packages can be used for estimating the parameters of a Box-Jenkins model using this method.

### Diagnostic checking

If the fitting model is satisfactory, their residuals (the difference between observed values and fitted values) may not be correlated and may follow a normal distribution with zero mean and equal variance. Histogram and quantile-quantile norm plots of the residuals are used for checking the assumption of normality. The residuals acf is estimated to test the independence assumption. The goodness fit statistic Portmanteau test (13) was used for testing the magnitudes of the residual autocorrelations as a group.

Details of Box-Jenkins modelling have been reported by Box-Jenkins (13), Chatfield (14), and Diggle (15).

## RESULTS

Figure 1 shows the huge decline in the number of deaths due to diarrhoea over time. An overall 93% reduction in the number of monthly deaths was observed from January 1980 to December 1998. The peaks in the time series plot suggest a strong seasonal pattern. Three large peaks were identified: the first in February 1980 and the other two in June 1984 and June 1990 (Fig. 1).

**Fig. 1.** Time series of deaths due to diarrhoea in Rio de Janeiro state, January 1980–December 1998

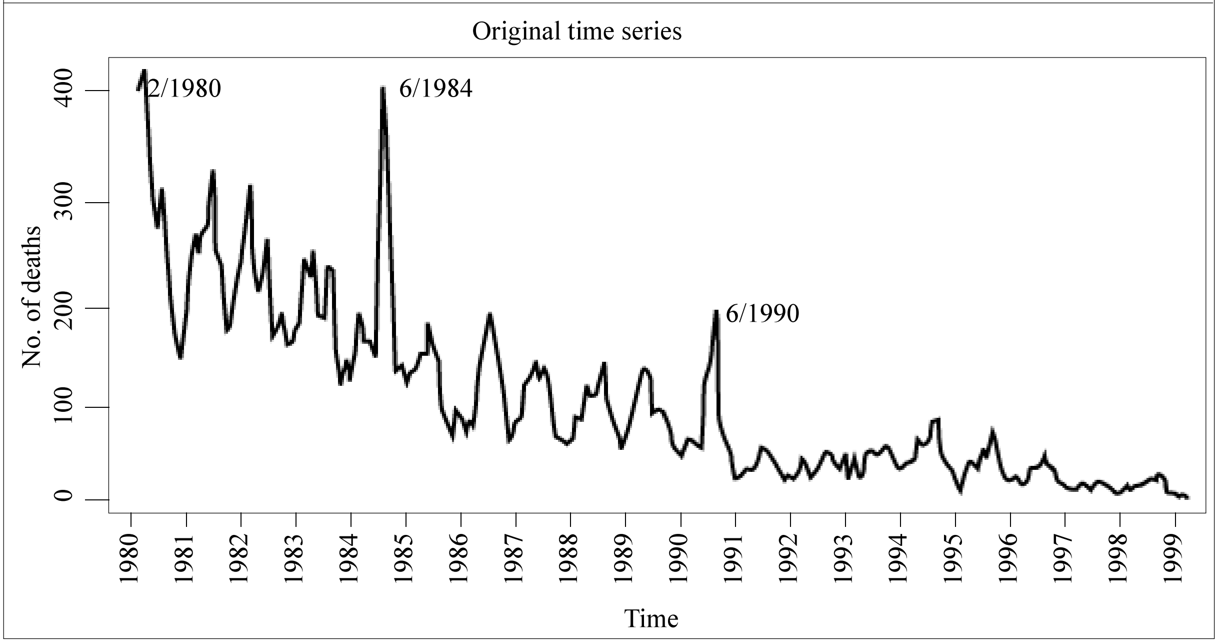


Table 1 shows the total number, the mean, and the respective standard deviations of deaths occurring each month. May and June had the largest total number and monthly average of deaths, suggesting seasonal peaks.

**Table 1.** Total number and monthly mean and standard deviation of deaths due to diarrhoea in Rio de Janeiro State, 1980-1998

| Month | Total no. of cases (n=22,650) | Mean | Standard deviation |
|---|---|---|---|
| January | 2,159 | 113.63 | 106.00 |
| February | 2,125 | 111.84 | 99.50 |
| March | 2,152 | 113.26 | 94.43 |
| April | 2,189 | 115.21 | 78.26 |
| May | 2,420 | 127.37 | 91.77 |
| June | 2,504 | 131.79 | 98.55 |
| July | 2,051 | 107.95 | 83.59 |
| August | 1,579 | 83.11 | 61.62 |
| September | 1,300 | 68.42 | 50.73 |
| October | 1,298 | 68.32 | 55.98 |
| November | 1,320 | 69.47 | 60.01 |
| December | 1,553 | 81.74 | 73.30 |

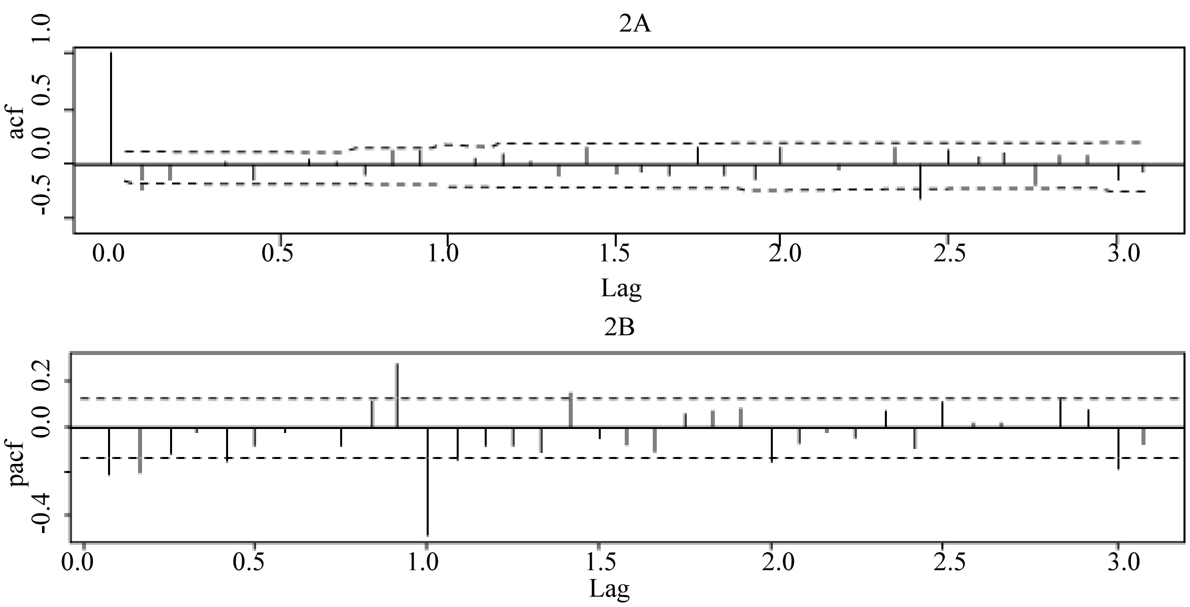
As the mean number of deaths each month and the respective variance decreased over the years (linear relationship), a natural logarithmic transformation was necessary to induce constant variance. This transformation also brings the time series closer to normality (an important assumption of the Box-Jenkins approach). To remove the trend and seasonal components of the logarithmic time series, one regular differencing and one seasonal differencing were made respectively. All these transformations were elaborated to ensure the stationarity of the time series.

The next step was the estimation of acf and pacf of the differencing logarithmic time series (transformed time series) to evaluate the direction and strength of the relationship among observations. Figure 2A shows the correlation coefficients different from zero at lags 1, 11, and 12, suggesting, respectively, a non-seasonal moving average term (first order) and a seasonal autoregressive term (first order). In the pacf correlogram (Fig. 2B), there was an exponential decay in the first nine lags and two correlations different from zero corresponding to the lags 11 and 12. Based on acf and pacf, some SARIMA models were identified. According to the Akaike information criteria (AIC) and p values of the Portmanteau test (13), the SARIMA model chosen can be written as follow:

$$w_t = a_t + 0.86a_{t-1} - 0.53y_{t-1} + 0.60y_{t-12} \qquad (1)$$

where the logarithm of the non-seasonal and seasonal differencing observations at time $t$ ($w_t$) depend on the error at time $t$ ($a_t$) and 0.86 of the previous error ($a_{t-1}$) and –0.53 of the observations at time $t$-1 ($y_{t-1}$) and 0.60 of the observations at time $t$-12 ($y_{t-12}$).

**Fig. 2.** Autocorrelation and partial autocorrelation correlogram of the transformed series

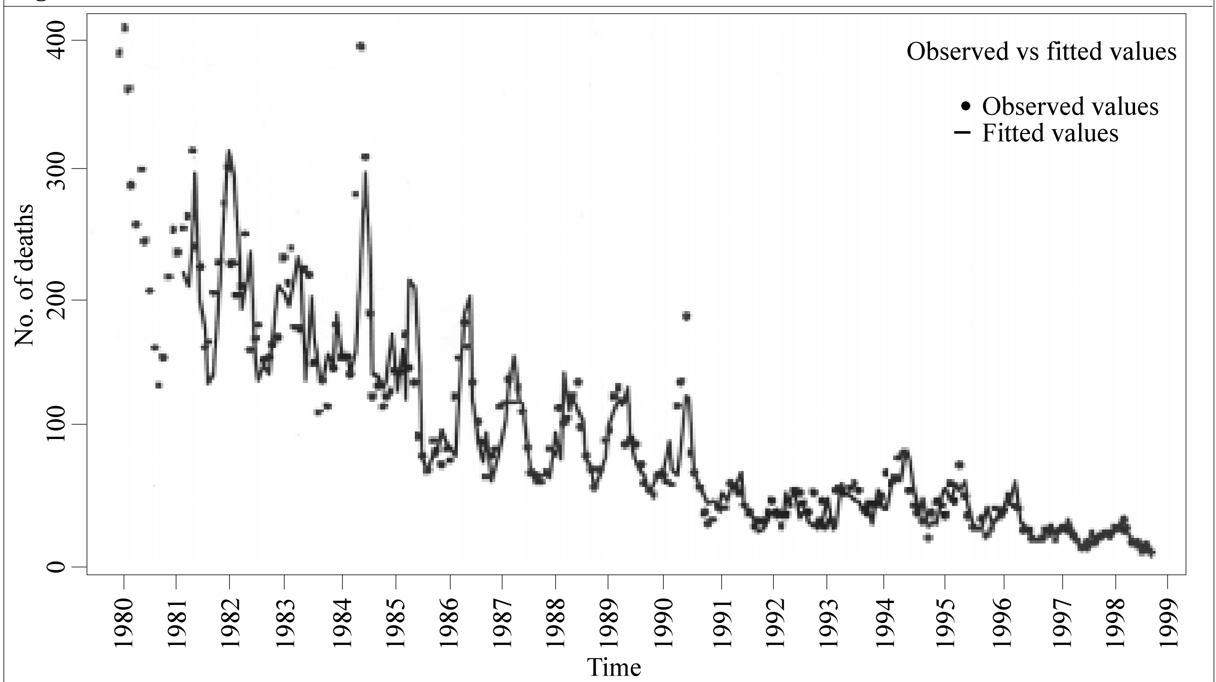



acf=Autocorrelation function
pacf=Partial autocorrelation function

The residual analysis of the chosen model was used for checking the inference assumption of ARIMA modelling (independence and normality of the residuals). The standard residual time series of the selected model did not present any clear pattern. The autocorrelation function of the residuals was estimated, and no temporal dependent structure could be observed. The residuals had a symmetric distribution around zero and a linearity of the quantile-quantile plot indicating normality.

The estimated SARIMA model, showed above (equation 1), was a reasonable fit for the data as shown in Figure 3 (fitted values close to the observed values). The observed increase in the number of deaths in February 1980, June 1984, and June 1990 may be atypical peaks which caused some harmful interference in the modelling process. During the analysis period, ICD changed only once—in 1996, with the adoption of its 10$^{th}$ revision—after the occurrences of these peaks. Other appropriate statistical techniques may be needed to check if these peaks were actually epidemics. Some climate changes may account for the increase in the number of diarrhoea cases. Checkley *et al.* found a direct relationship between the increase in hospital admissions due to diarrhoeal diseases and the increase in temperature caused by El Niño in 1997 and 1998 among Peruvian children (17). Modelling of deaths due to diarrhoea, including climate explanatory variables, could help explain these findings.

**Fig. 3.** SARIMA model




The temporal structure of hospitalization series due to rotavirus-associated diarrhoea was identified by José *et al.* when analyzing hospital admission data in Melbourne, Australia (18). They described annual and inter-epidemic cycles and, for the first time, a biannual cycle was identified. Purohit *et al.* described a pure seasonal autoregressive model in the analysis of hospitalizations due to rotavirus-associated diarrhoea in Pune, India (19).

## DISCUSSION

The huge decline in the number, and the annual peaking, of deaths due to diarrhoea in May and July characterized by a more dry climate in Rio de Janeiro State—over the 1980-1998 period—could be evidenced in the analysis. These findings may suggest rotavirus as the predominant aetiology of diarrhoeal deaths among children in Rio de Janeiro State. Kilgore *et al.*, using the same ICD codes, described a marked decline in mortality due to diarrhoea in the USA from 1968 to 1985 and a stabilization thereafter when the winter peaks decreased (16). The authors associated this behaviour of diarrhoeal deaths with aetiology of rotavirus.

Box-Jenkins analysis, used as one of the main tools to describe the temporal patterns of time series, provides accurate short-term forecasts. The decision about which terms to retain in the model and which to exclude is arcane, and the choice of the best model is, usually, a difficult task. Here, the SARIMA model chosen seemed to be appropriated and to fit the data. ARIMA modelling is based only on the mathematical properties of the series and not on the dynamic of transmission of infectious diseases. However, an epidemiological interpretation, regarding the transformations, could be that the fitted model suggests that the number of deaths due to diarrhoea in month $t$ depends on the number of deaths in the previous month $t$-1 (autoregressive component–AR 1) and on the frequency of the number of deaths in the same month of the previous year $t$-12 (seasonal autoregressive component–$AR_{12}$ 1). The number of deaths at month $t$ also depends on the variability of the frequency of deaths in the previous month (moving average component–MA 1).

There was no important departure from the assumptions of normal distribution and dependence structure in the respective residuals. The AR and MA terms may represent a confounding structure in the presence of other explanatory variables as, for example climate-related variables, but it does not interfere in the recognition of the temporal structure of the time series, which was the aim of this study. The inclusion of further years of the time series with the expected small number of deaths may cause a departure from the normality assumption and hinders the application of Box-Jenkins modelling.

In conclusion, the temporal pattern of the time series of diarrhoea in this study suggests that a large portion of deaths of children, aged less than five years, due to diarrhoea in Rio de Janeiro State may be attributed to rotavirus, which is in agreement with findings of other studies (1,4,5,8). Hospital-based studies have shown that rotavirus infections are associated with 12-42% of acute diarrhoea cases in Brazil—11.6% in Rio de Janeiro in 1993—and there are peaks of occurrence of diarrhoea due to rotavirus in dry months (May-September) in the southern part of the country where Rio de Janeiro State is situated (7). The decision to use future approved vaccines will be based on an expected reduction in moderate and severe rotavirus infections and the predicted cost-effectiveness of such a proposed programme (20). Linhares *et al.* believe that further vaccine trials in Brazil would probably involve rotavirus candidate vaccines rather than the tetravalent rotavirus vaccine (7).

## REFERENCES


1. Bern C, Martines J, de Zoysa I, Glass RI. The magnitude of the global problem of diarrhoeal disease: a ten-year update. *Bull World Health Organ* 1992;70:705-14.
2. Victora CG, Olinto MT, Barros FC, Nobre LC. Falling diarrhoea mortality in northeastern Brazil: did ORT play a role? *Health Policy Plan* 1996;11:132-41.
3. Waldman EA, Barata RCB, Moraes JC, Aquemi G, Timenetsky MCST. Gastroenteritis and acute respiratory infections among children up to 5 years old in an area of southeastern Brazil, 1986-1987. II. Diarrhea. *Rev Saúde Pública* 1997;31:62-70.
4. Guerrant RL, Kirchhoff LV, Shields DS, Nations MK, Leslie J, de Sousa MA *et al.* Prospective study of diarrheal illnesses in northeastern Brazil: patterns of disease, nutritional impact, etiologies, and risk factors. *J Infect Dis* 1983;148:986-97.
5. Ho M-S, Glass RI, Pinsky PF, Anderson LJ. Rotavirus as a cause of diarrheal morbidity and mortality in the United States. *J Infect Dis* 1988;158:1112-6.
6. Robins-Browne RM. Seasonal and racial incidence of infantile gastroenteritis in South Africa. *Am J Epidemiol* 1984;119:350-5.
7. Clemens J, Keckich N, Nacify A, Glass R, Rao M. Public health considerations for the introduction of new rotavirus vaccines for infants: a case study of tetravalent rhesus rotavirus-based ressortant vaccine. *Epidemiol Rev* 1999;21:24-42.
8. Linhares AC. Rotavirus infection in Brazil: epidemiology and challenges for control. *Cadernos de Saúde Pública* 2000;16:629-46.
9. Parashar UD, Bresee JS, Gentsch JR, Glass RI. Rotavirus. *Emerging Infect Dis* 1998;4:561-70. (http://www.cdc.gov/ncidod/eid/vol4no4/parashar.htm).
10. Cook SM, Glass RI, LeBaron CW, Ho M-S. Global seasonality of rotavirus infections. *Bull World Health Organ* 1990;68:171-7.
11. Bresee JS, Glass RI, Ivanoff B, Gentsch JR. Current status and future priorities for rotavirus vaccine development, evaluation and implementation in developing countries. *Vaccine* 1999;17:2207-22.

12. Shann F, Steinhoff MC. Vaccines for children in rich and poor countries. *Lancet* 1999;354(Suppl 2):7-11.
13. Box GEP, Jenkins G. Time series analysis. Rev. ed. San Francisco: Holden-Day, 1976. 575 p.
14. Chatfield C. The analysis of time series: an introduction. 4th ed. London: Chapman & Hall, 1989. 241 p.
15. Diggle PJ. Time series: a biostatiscal introduction. 1st ed. Oxford: Oxford University Press, 1990. 257 p.
16. Kilgore PE, Holman RC, Clarke MJ, Glass RI. Trends of diarrheal disease-associated mortality in US children, 1968 through 1991. *J Am Med Assoc* 1995;274:1143-8.
17. Checkley W, Epstein LD, Gilman RH, Figueroa D, Cama RI, Patz JA *et al.* Effects of El Niño and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. *Lancet* 2000;355:442-50.
18. José MV, Bobadilla JR, Bishop RF. Oscillatory fluctuations in the incidence of rotavirus infections by serotypes 1,2,3, and 4. *J Diarrhoeal Dis Res* 1996;14:194-200.
19. Purohit SG, Kelkar SD, Simha KV. Time series analysis of patients with rotavirus diarrhoea in Pune, India. *J Diarrhoeal Dis Res* 1998;16:74-83.
20. Ford-Jones EL, Wang E, Petric M, Corey P, Moineddin R, Fearon M. Rotavirus-associated diarrhea in outpatient settings and child care centers. *Arch Pediatr Adolesc Med* 2000;154:586-603.