PREDICTING CLINICALLY DIAGNOSED DYSENTERY INCIDENCE OBTAINED FROM MONTHLY CASE REPORTING BASED ON METEOROLOGICAL VARIABLES IN DALIAN, LIAONING PROVINCE, CHINA, 2005-2011 USING A DEVELOPED MODEL

Qingyu An, Wei Yao and Jun Wu

Dalian Center for Disease Control and Prevention, Liaoning Province, PR China

Abstract. This study describes our development of a model to predict the incidence of clinically diagnosed dysentery in Dalian, Liaoning Province, China, using time series analysis. The model was developed using the seasonal autoregressive integrated moving average (SARIMA). Spearman correlation analysis was conducted to explore the relationship between meteorological variables and the incidence of clinically diagnosed dysentery. The meteorological variables which significantly correlated with the incidence of clinically diagnosed dysentery were then used as covariables in the model, which incorporated the monthly incidence of clinically diagnosed dysentery from 2005 to 2010 in Dalian. After model development, a simulation was conducted for the year 2011 and the results of this prediction were compared with the real observed values. The model performed best when the temperature data for the preceding month was used to predict clinically diagnosed dysentery during the following month. The developed model was effective and reliable in predicting the incidence of clinically diagnosed dysentery for most but not all months, and may be a useful tool for dysentery disease control and prevention, but further studies are needed to fine tune the model.

Keywords: dysentery, predicting model, meteorological variable

INTRODUCTION

Intestinal infection causing diarrhea may be caused by bacteria, viruses or parasites. One study found shigellosis was the most common cause of diarrhea in China (Wang *et al*, 2006). Dysentery occurs worldwide to an estimated 164.7 million people with 1.1 million deaths per year (Kotloff *et al*, 1999). In China, the incidence of intestinal infectious diseases has declined considerably in recent years, but bacillary dysentery is still common because of its high incidence (Guan *et al*, 2008), and remains a serious public health problem in China (Yan *et al*, 2010). In this study, we attempted to establish a reliable forecasting model for dysentery incidence in Dalian.

Correspondence: Qingyu An, Dalian Center for Disease Control and Prevention, Dalian, Liaoning Province, PR China, 116021.
Tel: 86 0411 84335830; Fax: 86 0411 84335901
E-mail: anqingyu@163.com

MATERIALS AND METHODS

Study area

Dalian is the main coastal city of Liaoning Province, China and a major tourist city located at 38°43′-40°10′N latitude and 120°58′-123°31′E longitude. It had a population of 6.69 million in May 2011. Dalian has a warm continental monsoon climate and is in a marine temperate zone (Wang, 2013). The average temperature is 10.5°C with a maximum of 37.8°C, and a minimum of -19.1°C (SINA, 2008). The average rainfall is 550-950 mm and the average annual sunshine is 500-2800 hours (SINA, 2008).

Data collection

The incidence of clinically diagnosed dysentery for 2005 to 2011 was obtained from the Dalian Center for Disease Control and Prevention, Liaoning Province, China. The cases were diagnosed based on clinical symptoms, such as diarrhea with blood or mucus and tenesmus. Only 15.26% were laboratory diagnosed. In China, all dysentery cases must be reported through the China information system within 24 hours. Dysentery is commonly diagnosed in China. The meteorological data for the same period for Dalian was obtained from the China meteorological data sharing service system. These data consist of atmospheric pressure, air temperature, relative humidity, rainfall, hours of sunlight and wind speed.

Data analysis

The relationship between the meteorological variables and the incidence of clinically diagnosed dysentery was analyzed using Spearman correlation analysis observing for a lag effect. We used the Box-Jenkins approach to seasonal autoregressive integrated moving average (SARIMA) model for time series. The model was as follows: given a stationary time series of data Y′ = (Y_1,Y_2, ..., Y_n), a SARIMA model with S observations per period, denoted by SARIMA (p,d,q)$(P,D,Q)_s$, is given by

$$\phi(B^s)\phi(B)(1-B)^d(1-B^s)^D Y_t = \theta(B^s)\theta(B)E_t \qquad (1)$$

where

$$\phi(B^s) = 1 - \phi s_{,1}B^s - \phi s_{,2}B^{2s} - \cdots - \phi s_{,p}B^{ps} \qquad (2)$$

$$\theta(B^s) = 1 + \theta_{s,1}B^s + \theta_{s,2}B^{2s} + \ldots + \theta_{s,Q}B^{Qs} \qquad (3)$$

$\phi(B^s)$ and $\theta(B^s)$ are seasonal polynomial functions of order P and Q, respectively; $\phi 1$, ϕ_2,...,ϕ_p are vectors for autoregressive coefficients, θ_1, θ_2,...,θ_Q are vectors of moving average coefficients; E_t is an error term assumed to be independent; d and D are the values for non-seasonal and seasonal cases of differencing order, respectively; p and P are the non-seasonal and seasonal cases of autoregression order, respectively; q and Q are the non-seasonal and seasonal cases of moving average order, respectively (Martinez and Silva, 2011).

The criterion for comparing the predictive ability of the models was the average relative error defined as:

$$e = \frac{1}{n}\sum_{t=1}^{n}\left[\frac{(x_t - \hat{x}_t)}{x_t} \times 100\%\right] \qquad (4)$$

where x_t and $\hat{x}_t$ denote the observed and fitted values for that point in time. The preferred model was the one with the lowest average relative error. SPSS version 11.5 (SPSS, Chicago, IL) was used for data analysis. A p-value < 0.05 was considered statistically significant.

RESULTS

Descriptive analysis

From January 2005 to December 2010, 15,990 cases of clinically diagnosed dysentery were reported (8,621 males and 7,369 females) from Dalian, Liaoning Province, China. One person died. The median age of the patients was 25 years (range: 2 days-96 years).

The highest incidence occurred in 2006 with 70.133/100,000 population and the lowest incidence occurred in 2010 with 25.961/100,000 population. The incidences

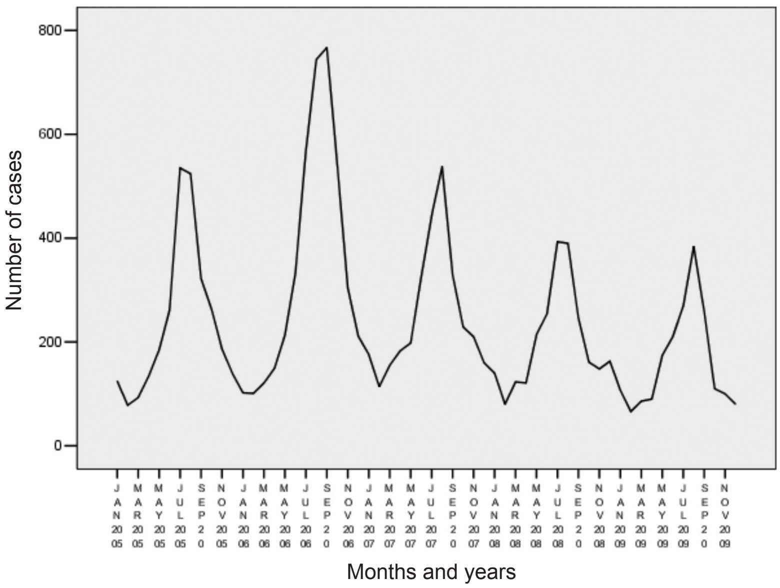

Fig 1–Monthly cases of dysentery in Dalian 2005-2009.

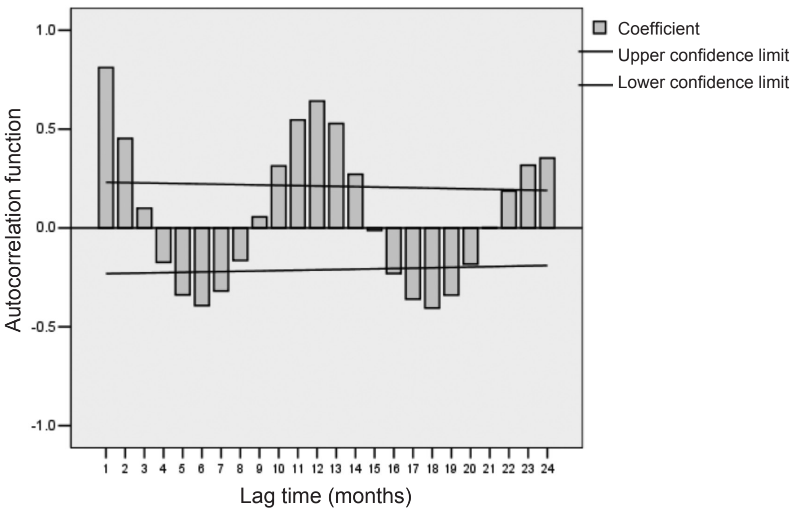

Fig 2–Autocorrelation functions of dysentery cases without differencing.

of bacillary dysentery cases reported during 2005, 2007, 2008 and 2009 were 48.085, 51.097, 41.394 and 32.023 per 100,000 population, respectively.

The monthly cases of dysentery reported from Dalian during 2005-2009 are shown in Fig 1. July, August and September each year were the months with the highest incidence of dysentery.

The incidences of dysentery varied considerably by sub-district during the study period. The three sub-districts with the highest incidences of dysentery during 2005 were: Zhong Hua Road subdistrict (226.33 per 100,000), Hu Tan subdistrict (193.06 per 100,000) and Pao Ya subdistrict (178.22 per 100,000); during 2006 were: Hu Tan subdistrict (293.08 per 100,000), Bai Shan subdistrict (248.22 per 100,000) and Quan Shui subdistrict (247.81 per 100,000); during 2007 were: Quan Shui subdistrict (206.50 per 100,000), Pao Ya subdistrict (201.38 per 100,000) and the airport subdistrict (196.80 per 100,000); during 2008 were: the airport subdistrict (179.31 per 100,000), Hu Tan subdistrict (176.78 per 100,000) and Quan Shui subdistrict (144.55 per 100,000); during 2009 were Quan Shui subdistrict (206.50 per 100,000), Hu Tan subdistrict (151.19 per 100,000) and the airport subdistrict (118.08 per 100,000); during 2010: were Quan Shui subdistrict (191.02

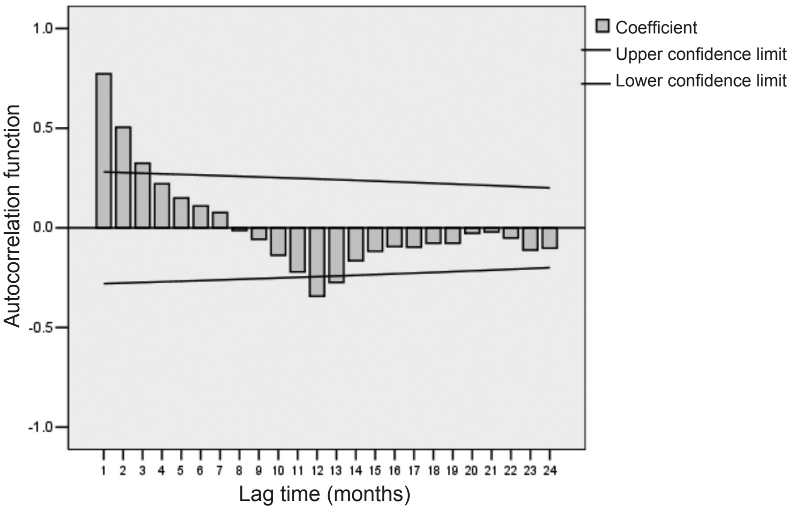

Fig 3–Autocorrelation functions of dysentery cases with natural logarithm and seasonal differencing.

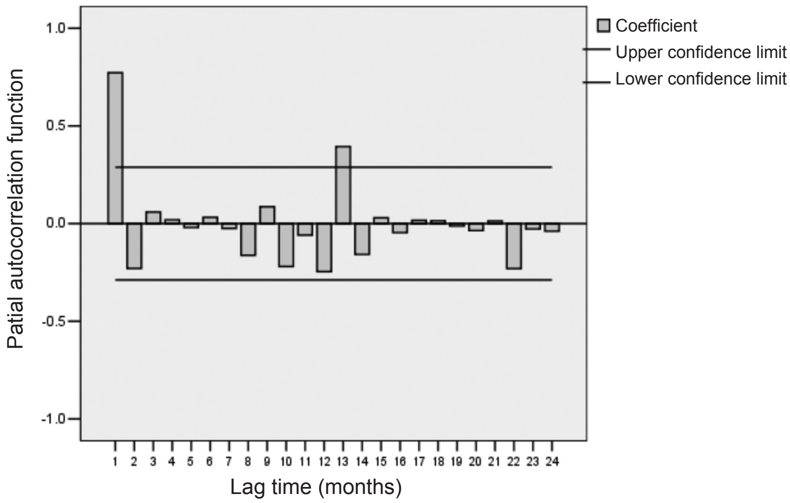

Fig 4–Partial autocorrelation functions of dysentery cases with natural logarithm and seasonal differencing.

per 100,000), Ren Min Square subdistrict (91.44 per 100,000), and Hu Tan subdistrict (86.06 per 100,000).

Spearman correlation analysis

Meteorological variables affected the incidence of reported dysentery cases with a lag time. Spearman correlation analysis was conducted using meteorological variables from the same month, and one and two months before. The relationships between the incidence of reported dysentery cases and the major meteorological variables are shown in Table 1. Most climate variables were significantly associated with the incidence of reported dysentery cases except hours of sunlight. The correlation coefficients ranged from 0.515 to 0.729 for temperature, -0.737 to -0.253 for atmospheric pressure, -0.629 to -0.223 for wind speed and 0.351 to 0.688 for relative humidity and rainfall.

The best-fitting SARIMA model

The series of clinically diagnosed reported dysentery case notifications had a nonstationary mean (Fig 2), so it was necessary to stabilize the mean for dysentery incidence by taking seasonal first order differencing (Fig 3). All further statistical conclusions were made using the transformed dysentery incidence. Based on distribution characteristics (Figs 3 and 4) and the relationship with meteorological variables, we developed 60 models and selected the average temperature, average minimum temperature, extreme minimum temperature, average atmospheric pressure and maximum atmospheric pressure

Table 1
Spearman correlation coefficients between monthly meteorological variables and dysentery cases.

Variables	Same month		One month before		Two months before	
	Correlation coefficient	*p*-value	Correlation coefficient	*p*-value	Correlation coefficient	*p*-value
Average wind speed	-0.629	0.000	-0.521	0.000	-0.223	0.064
Maximum average wind speed	-0.399	0.001	-0.393	0.001	-0.285	0.017
Maximum instantaneous wind speed	-0.355	0.002	-0.360	0.002	-0.274	0.022
Sunshine hours	-0.002	0.984	0.147	0.220	0.392	0.001
Average atmospheric pressure	-0.544	0.000	-0.728	0.000	-0.684	0.000
Maximum atmospheric pressure	-0.599	0.000	-0.737	0.000	-0.640	0.000
Minimum atmospheric pressure	-0.253	0.032	-0.502	0.000	-0.563	0.000
Average temperature	0.697	0.000	0.699	0.000	0.523	0.000
Average maximum temperature	0.687	0.000	0.695	0.000	0.537	0.000
Average minimum temperature	0.711	0.000	0.701	0.000	0.515	0.000
Extreme maximum temperature	0.644	0.000	0.671	0.000	0.540	0.000
Extreme minimum temperature	0.709	0.000	0.729	0.000	0.561	0.000
Average relative humidity	0.688	0.000	0.666	0.000	0.379	0.001
Minimum relative humidity	0.552	0.000	0.448	0.000	0.081	0.506
Rainfall	0.503	0.000	0.531	0.000	0.351	0.003

as covariates to use in the model (Table 2). Of the models tested (Table 2 and Fig 5, 6), the SARIMA(1,0,0)(1,1,0)$_{12}$ and SARIMA(1,0,0)(0,1,1)$_{12}$ models with the average temperature one month before as the covariate fit the data best. Table 3 shows the parameter estimates for the two models.

The equation for the SARIMA(1,0,0)(1,1,0)$_{12}$ model was:

$$(1-0.792B)(1+0.459B^{12})(1-B)^{12}Y_t = E_t\,.$$

The equation for the SARIMA(1,0,0)(0,1,1)$_{12}$ model was:

$$(1-0.776B)(1-B)^{12}Y_t = (1+0.468B^{12})E_t\,.$$

The model's fitted and observed values are shown in Figs 7 and 8. The observed and fitted values share the same incidence trend, except for July 2006, August 2006 and July 2008 for the SARIMA(1,0,0)(1,1,0)$_{12}$ model and July and August 2006 for the SARIMA(1,0,0)(0,1,1)$_{12}$ model.

Table 4 shows the number of predicted cases and 95% confidence intervals obtained from the SARIMA(1,0,0)(1,1,0)$_{12}$ and SARIMA(1,0,0)(0,1,1)$_{12}$ models with the average temperature one month before as the covariate. The observed and predicted values were relatively close to each other except for July and August. The average relative error values for the SARIMA(1,0,0)(1,1,0)$_{12}$ and SARIMA(1,0,0)(0,1,1)$_{12}$ models were 26.53% and 22.39%, respectively. The SARIMA(1,0,0)(0,1,1)$_{12}$ model using the average temperature one month previously as the covariate was the best-fitting SARIMA model.

DISCUSSION

Many models such as the generalized regression, seasonal autoregressive integrated moving average and artificial neural networks (ANN) models, have been

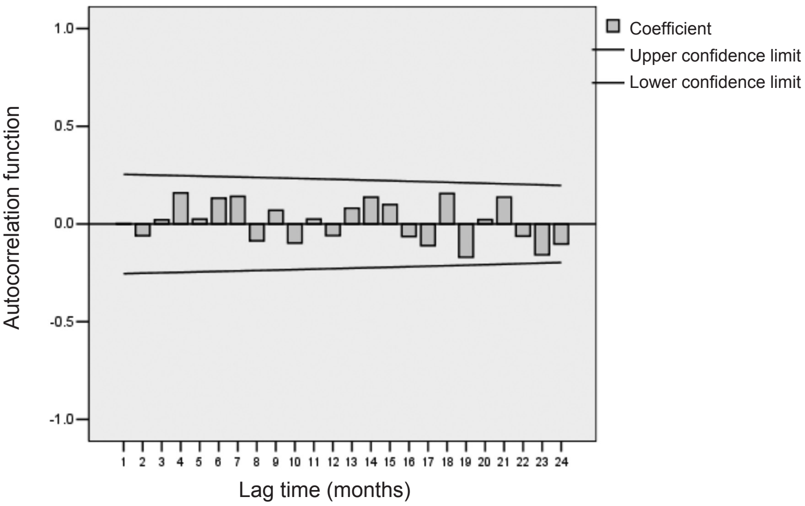

Fig 5–Autocorrelation function of residuals for the SARIMA(1,0,0)(1,1,0)$_{12}$ model.

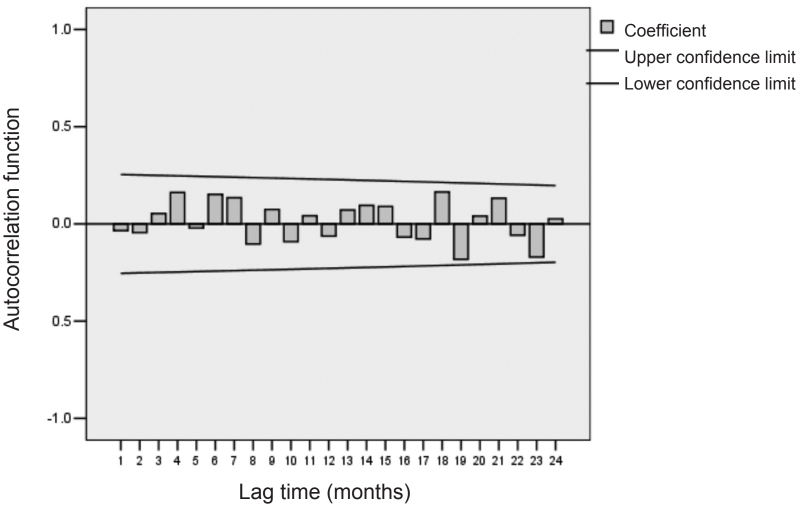

Fig 6–Autocorrelation function of residuals for the SARIMA(1,0,0)(0,1,1)$_{12}$ model.

applied to infectious disease forecasting research (Urashima *et al*, 2003b; Yan *et al*, 2010; Pasomsub *et al*, 2010). The SARIMA model has been used for time-series modeling and predicting (Box and Jenkins, 2008). It takes into account the impact of seasonality and autocorrelations. The SARIMA model has been successfully used in epidemiology to predict infectious diseases, such as dengue fever (Luz *et al*, 2008), malaria (Kinley *et al*, 2010) and bacillary dysentery. Guo *et al* (2012) developed a SARIMA model based on the data of the monthly incidence of bacillary dysentery from 2004 to 2010 in Nanning, Guangxi Province, China, to predict trends in bacillary dysentery in Nanning. Li *et al* (2010) used a SARIMA model to predict the monthly number of bacillary dysentery cases in Guangxi. Mu *et al* (2009) used a SARIMA model to predict the number of bacillary dysentery case from 1980 to 2007. Cui *et al* (2009) used an ARIMA model to predict the weekly incidence of bacillary dysentery in Chao Yang District, Beijing, China for 2004 to 2008. All these studies used a SARIMA model based on monthly incidence data and an ARIMA model based on weekly incidence data. In this study, we developed a SARIMA model based on monthly incidence and meteorological data during the same period.

In order to develop a stable and effective SARIMA model, different prediction models were explored by fitting covariates to the time series data. These covariates were obtained from the results of Spearman correlation analysis and used

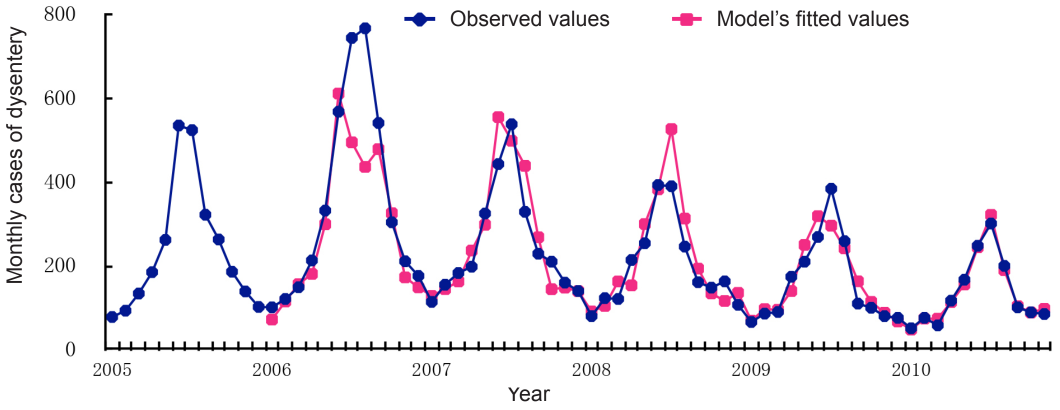

Fig 7–Observed number of notified dysentery cases during 2005-2010 in Dalian and numbers of cases estimated by the SARIMA $(1,0,0)(1,1,0)_{12}$ model.

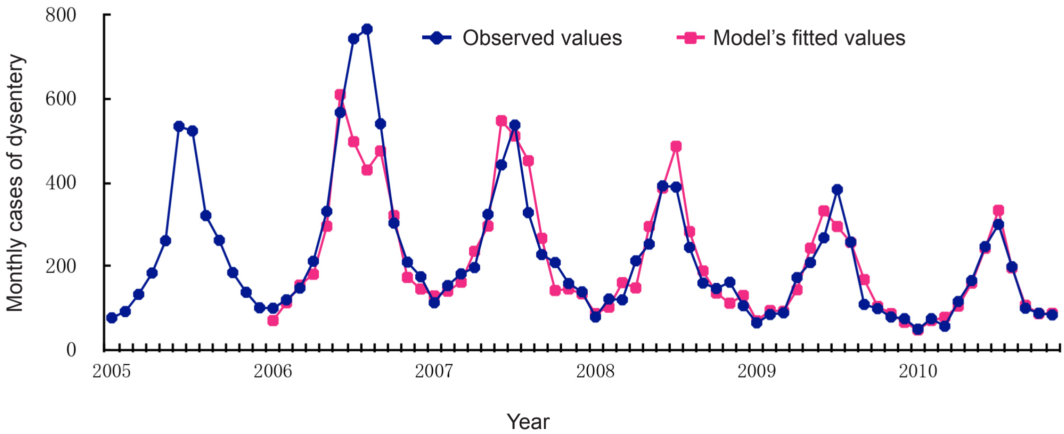

Fig 8–Observed number of reported dysentery cases during 2005-2010 in Dalian and numbers of cases estimated by the SARIMA $(1,0,0)(0,1,1)_{12}$ model.

to explore the relationship between meteorological variables and the incidence of clinically diagnosed dysentery. Our finding suggests climate variability may have played a significant part in the transmission cycle of reported dysentery cases. A higher temperature can increase the growth of bacteria. A lower atmosphere pressure may reduce the partial pressure of oxygen, impairing the body's resistance (Qu *et al*, 2009) and increasing the risk for infection.

The SARIMA model we developed gave a fair estimate of the number of clinically diagnosed reported dysentery cases for most, but not all the study months for Dalian, Liaoning Province, China. The model failed to correctly estimate the number of dysentery cases during June to September and especially during July and August. The predicted values were about half the observed values. The reason for this error could be due to outbreaks of dysentery during those two months. When developing a model and an outbreak occurs during the study period, those intervals containing excess cases need to be excluded to retain the seasonality

Table 2
Akaike Information Criterion (AIC) values for different SARIMA models.

Model	AIC	Model	AIC
Average temperature as covariate		Extreme minimum temperature as covariate	
ARIMA(1,0,0)(1,1,0)$_{12}$	-19.423	ARIMA(1,0,0)(1,1,0)$_{12}$	-13.855
ARIMA(1,0,0)(1,1,1)$_{12}$	-17.824	ARIMA(1,0,0)(1,1,1)$_{12}$	-12.956
ARIMA(1,0,0)(0,1,0)$_{12}$	-10.533	ARIMA(1,0,0)(0,1,0)$_{12}$	-8.369
ARIMA(1,0,0)(0,1,1)$_{12}$	-19.260	ARIMA(1,0,0)(0,1,1)$_{12}$	-14.959
ARIMA(1,0,1)(1,1,0)$_{12}$	-17.628	ARIMA(1,0,1)(1,1,0)$_{12}$	-12.682
ARIMA(1,0,1)(1,1,1)$_{12}$	-16.252	ARIMA(1,0,1)(1,1,1)$_{12}$	-12.365
ARIMA(1,0,1)(0,1,0)$_{12}$	-9.292	ARIMA(1,0,1)(0,1,0)$_{12}$	-7.458
ARIMA(1,0,1)(0,1,1)$_{12}$	-18.061	ARIMA(1,0,1)(0,1,1)$_{12}$	-14.369
ARIMA(1,0,2)(1,1,0)$_{12}$	-16.716	ARIMA(1,0,2)(1,1,0)$_{12}$	-10.96
ARIMA(1,0,2)(1,1,1)$_{12}$	-15.323	ARIMA(1,0,2)(1,1,1)$_{12}$	-10.612
ARIMA(1,0,2)(0,1,0)$_{12}$	-7.506	ARIMA(1,0,2)(0,1,0)$_{12}$	-5.446
ARIMA(1,0,2)(0,1,1)$_{12}$	-17.022	ARIMA(1,0,2)(0,1,1)$_{12}$	-12.651
Average minimum temperature as covariate		Average atmospheric pressure as covariate	
ARIMA(1,0,0)(1,1,0)$_{12}$	-1.627	ARIMA(1,0,0)(1,1,0)$_{12}$	-14.168
ARIMA(1,0,0)(1,1,1)$_{12}$	-1.672	ARIMA(1,0,0)(1,1,1)$_{12}$	-13.165
ARIMA(1,0,0)(0,1,0)$_{12}$	1.996	ARIMA(1,0,0)(0,1,0)$_{12}$	-8.371
ARIMA(1,0,0)(0,1,1)$_{12}$	-3.527	ARIMA(1,0,0)(0,1,1)$_{12}$	-15.129
ARIMA(1,0,1)(1,1,0)$_{12}$	-0.095	ARIMA(1,0,1)(1,1,0)$_{12}$	-12.752
ARIMA(1,0,1)(1,1,1)$_{12}$	-0.387	ARIMA(1,0,1)(1,1,1)$_{12}$	-12.26
ARIMA(1,0,1)(0,1,0)$_{12}$	3.457	ARIMA(1,0,1)(0,1,0)$_{12}$	-7.471
ARIMA(1,0,1)(0,1,1)$_{12}$	-2.39	ARIMA(1,0,1)(0,1,1)$_{12}$	-14.29
ARIMA(1,0,2)(1,1,0)$_{12}$	1.887	ARIMA(1,0,2)(1,1,0)$_{12}$	-11.216
ARIMA(1,0,2)(1,1,1)$_{12}$	1.121	ARIMA(1,0,2)(1,1,1)$_{12}$	-10.618
ARIMA(1,0,2)(0,1,0)$_{12}$	5.54	ARIMA(1,0,2)(0,1,0)$_{12}$	3.571
ARIMA(1,0,2)(0,1,1)$_{12}$	-0.55	ARIMA(1,0,2)(0,1,1)$_{12}$	-12.679
Maximum atmospheric pressure as covariate			
ARIMA(1,0,0)(1,1,0)$_{12}$	-15.617		
ARIMA(1,0,0)(1,1,1)$_{12}$	-14.601		
ARIMA(1,0,0)(0,1,0)$_{12}$	-9.001		
ARIMA(1,0,0)(0,1,1)$_{12}$	-16.462		
ARIMA(1,0,1)(1,1,0)$_{12}$	-14.156		
ARIMA(1,0,1)(1,1,1)$_{12}$	-13.502		
ARIMA(1,0,1)(0,1,0)$_{12}$	-7.928		
ARIMA(1,0,1)(0,1,1)$_{12}$	-15.524		
ARIMA(1,0,2)(1,1,0)$_{12}$	-12.893		
ARIMA(1,0,2)(1,1,1)$_{12}$	-12.16		
ARIMA(1,0,2)(0,1,0)$_{12}$	-6.052		
ARIMA(1,0,2)(0,1,1)$_{12}$	-14.173		

Table 3
Parameters for the SARIMA $(1,0,0)(1,1,0)_{12}$ and SARIMA $(1,0,0)(0,1,1)_{12}$ models.

Model		Estimates	Std error	t	p-value
SARIMA$(1,0,0)(1,1,0)_{12}$					
Non-seasonal lags	AR1	0.792	0.080	9.940	0.000
Seasonal lags	Seasonal AR1	-0.459	0.114	-4.007	0.000
Regression coefficients	Average temperature	0.003	0.001	2.421	0.019
Constant		-0.124	0.082	-1.509	0.137
SARIMA$(1,0,0)(0,1,1)_{12}$					
Non-seasonal lags	AR1	0.776	0.082	9.419	0.000
Seasonal lags	Seasonal MA1	0.468	0.142	3.293	0.002
Regression coefficients	Average temperature	0.003	0.001	2.072	0.043
Constant		-0.125	0.068	-1.838	0.071

Table 4
Number of dysentery cases observed during 2011 and predicted values obtained from the SARIMA$(1,0,0)(1,1,0)_{12}$ and SARIMA$(1,0,0)(0,1,1)_{12}$ models.

Month	Number of observed cases	Number of predicted cases and 95% confidence intervals			
		SARIMA $(1,0,0)(1,1,0)_{12}$ model		SARIMA $(1,0,0)(0,1,1)_{12}$ model	
		Cases	95%CI	Cases	95%CI
1	76	65.5	44.3~96.9	69.0	46.6~102
2	63	40.1	24.1~66.8	42.8	25.8~71.3
3	57	61.2	34.8~107	65.5	37.5~115
4	90	57.7	31.7~105	61.3	33.9~111
5	141	113	60.5~210	114	61.6~211
6	219	150	79.3~284	156	83.3~293
7	419	200	105~384	218	115~413
8	453	261	135~504	264	138~505
9	221	181	93.3~351	182	95.0~349
10	141	84.2	43.3~164	92.9	48.3~179
11	127	77.7	39.8~151	84.6	43.9~163
12	70	72.5	37.1~142	78.6	40.7~152

of the data (Allard, 1998). However, further studies need to be conducted when there is not an outbreak to confirm the model, otherwise the model is invalid.

The SARIMA model we developed gave the best fit of the models we examined. The model is suited for short term prediction for some months. The average temperature one month previously, when used as covariate, is helpful in predicting the number of dysentery cases during the following month, but errors occurred when predicting cases during July and August, suggesting further studies are needed to refine the model. Once improved, this model could inform intervention measures.

An ARIMA model is not accurate unless it is based on at least 30 cases repeating at equal time intervals (*eg*, day, week, month) (Gao, 1997). Allard (1998) reported the longer the series, the better, but the preceding interval should not be too long. In our study, the longer time series did not give a better fitting model. We tried using different length of time data (1992-2009, 2001-2009, 2005-2010) to develop the SARIMA model, but 2005-2010 period gave the smallest error. Data based on periods further back in time might be affected by different case definitions and other factors, such as meteorological factors, human hygiene habits, immunity and prevention and control measures which can differ over time (Wu and Pu, 2006), leading to time series not being stationary in respect to means and variances.

This study had several limitations. First, there might be other factors affecting dysentery incidence, such as living conditions. The majority of the cases were diagnosed clinically and could have another etiology besides bacterial infection. A case definition could vary by clinician. The etiology of dysentery is complex. Many factors, such as the organism, host, and environmental conditions, are involved in the transmission cycle of dysentery. Temperature, humidity, human behavior, and population immunity can all contribute to and interact in the dysentery transmission cycle (Li, 2005); however, the availability of these data is limited. Another limitation of dysentery incidence data is that it was obtained from passive surveillance, resulting in a potential underreporting of cases, influencing the precision of our analysis. The average relative error was greater than 20% and the predicted values for July and August were only half the observed values.

With further studies, this model might be useful for estimating dysentery trends using meteorological factors. Non-climate factors may also impact dysentery organism transmission. More accurate predictions may require introducing non-climatic variables into the model, such as sociological and economic factors.

REFERENCES

Allard R. Use of time-series analysis in infectious disease surveillance. *Bull World Health Organ* 1998; 76: 327-33.

Box G, Jenkins G. Time series analysis: forecasting and control. 4th ed. New York: John Wiley & Sons, 2008.

Cui SF, Ma JX, Li SM. [The application of the prediction of the reported weekly incidence of bacillary dysentery in Chao yang district using the time series model]. *Chin J Health Stat* 2009; 26: 583-6.

Gao HX. SAS System·SAS/ETS Software Manual. Beijing: China Statistics Press, 1997; 83 (in Chinese).

Guan P, Huang DS, Guo JQ, *et al*. Bacillary dysentery and meteorological factors in northeastern china: a historical review based on classification and regression trees. *Jpn J Infect Dis* 2008; 61: 356-60.

Guo ZQ, Lin JY. Prediction of epidemiological tendency of bacillary dysentery in Nanning by ARIMA model. *J Prev Med Inform* 2012; 28: 806-8.

Kinley W, Singhasivanon P, Silawan T, Lawpoolsri S, White NJ, Kaewkungwal J. Development of temporal modelling for forecasting and prediction of malaria infections using time-series and ARIMAX analyses: a case study in endemic districts of Bhutan. *Malar J* 2010 Sep 3; 9: 251.

Kotloff KL, Winickoff JP, Ivanoff B, *et al*. Global burden of Shigella infections: implications for vaccine development and implementation of control strategies. *Bull World Health Organ* 1999; 77: 651-66.

Li LM. Epidemiology. 5th ed. Beijing: People's

Medical Publishing House, 2005: 181 (in Chinese).

Li YH, Lin M, Dong BQ, *et al*. Application of ARIMA model in the forecasting of bacillary dysentery. *Modern Prev Med* 2010; 37: 1203-10.

Luz PM, Mendes BV, Codeco CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *Am J Trop Med Hyg* 2008; 79: 933-9.

Martinez EZ, Silva EA. Predicting the number of cases of dengue infection in Ribeirão Preto, São Paulo State, Brazil, using a SARIMA model. *Cad Saude Publica* 2011; 27: 1809-18.

Mu J, Xie X, Li Y, *et al*. Feasibility study on ARIMA model in the prediction of the key notifiable communicable diseases in Shenzhen city from 1980 to 2007. *Prev Med Tribune* 2009; 15: 1051-5.

Pasomsub E, Sukasem C, Sunkanuparph S, Kijsirikul B, Chantratita W. The application of artificial neural networks for phenotypic drug resistance prediction: evaluation and comparison with other interpretation systems. *Jpn J Infect Dis* 2010; 63: 87-94.

Qu B, Guo HQ, Guan P, *et al*. Influence and prediction of meteorological factors on epidemic situation of digestive system infectious diseases in drought area. *World Chin J Digestol* 2009; 17: 1443-7.

SINA. Characters of climatic in Dalian. (on line). Shanghai: Sina Corp, 2008. [Cited 2008 Aug 28]. Available from: URL: http://sina.com.cn/china/2008-08-28/094621480.shtml

Urashima M, Shindo N, Okabe N. A seasonal model to simulate influenza oscillation in Tokyo. *Jpn J Infect Dis* 2003b; 56: 43-7.

Urashima M, Shindo N, Okabe N. Seasonal models of herpangina and hand-foot-mouth disease to simulate annual fluctuations in urban warming in Tokyo. *Jpn J Infect Dis* 2003a; 56: 48-53.

Wang XP. Climatic characteristics of thunderstorm in Dalian from 1951 to 2010. *J Meteorol Environ* 2013; 29: 80-3.

Wang XY, Tao FB, Xiao DL, *et al*. Trend and disease burden of bacillary dysentery in China (1991-2000). *Bull World Health Organ* 2006; 84: 561-8.

Wu JH, Pu LY .Comparison of two predicting methods of bacillary dysentery incidence rate in Zhoushan, China. *Chin Prev Med* 2006; 7: 69-70.

Yan WR, Xu Y, Yang XB, *et al*. A hybrid model for short-term bacillary dysentery prediction in Yichang City, China. *Jpn J Infect Dis* 2010; 63: 264-70.