Int J Biometeorol (2008) 52:179–187
DOI 10.1007/s00484-007-0109-4

ORIGINAL PAPER

Climate variations and salmonellosis transmission in Adelaide, South Australia: a comparison between regression models

Ying Zhang · Peng Bi · Janet Hiller

Received: 20 September 2006 / Revised: 6 May 2007 / Accepted: 7 May 2007 / Published online: 11 July 2007
© ISB 2007

Abstract This is the first study to identify appropriate regression models for the association between climate variation and salmonellosis transmission. A comparison between different regression models was conducted using surveillance data in Adelaide, South Australia. By using notified salmonellosis cases and climatic variables from the Adelaide metropolitan area over the period 1990–2003, four regression methods were examined: standard Poisson regression, autoregressive adjusted Poisson regression, multiple linear regression, and a seasonal autoregressive integrated moving average (SARIMA) model. Notified salmonellosis cases in 2004 were used to test the forecasting ability of the four models. Parameter estimation, goodness-of-fit and forecasting ability of the four regression models were compared. Temperatures occurring 2 weeks prior to cases were positively associated with cases of salmonellosis. Rainfall was also inversely related to the number of cases. The comparison of the goodness-of-fit and forecasting ability suggest that the SARIMA model is better than the other three regression models. Temperature and rainfall may be used as climatic predictors of salmonellosis cases in regions with climatic characteristics similar to those of Adelaide. The SARIMA model could, thus, be adopted to quantify the relationship between climate variations and salmonellosis transmission.

Y. Zhang · P. Bi · J. Hiller
Discipline of Public Health, University of Adelaide,
Adelaide, SA, Australia

P. Bi (✉)
Discipline of Public Health, University of Adelaide,
Level 9, 10 Pulteney Street,
Adelaide 5005, SA, Australia
e-mail: peng.bi@adelaide.edu.au

Keywords Climate · Multiple linear regression · Poisson · Salmonellosis · SARIMA · Time-series

Introduction

Salmonellosis is one of the most common and widely distributed food-borne diseases, with millions of cases being reported worldwide every year (WHO 2005). An estimated 1.4 million cases occur annually in the United States (CDC 2004) and *Salmonella* causes more deaths than any other food-borne pathogen in England and Wales (Adak et al. 2002). In Australia, *Salmonella* is one of the most common agents responsible for food-borne disease outbreaks, with 7,917 cases notified to OzFoodNet, the Australian national foodborne diseases surveillance system, in 2002 (OzFoodNet 2002). This was a 10% increase compared with the previous 4-year period. An up-to-date estimation indicated that *Salmonella* may be the cause of 92,000 cases of gastroenteritis annually in Australia (Hall et al. 2005).

More salmonellosis cases occur in the warmer season, which could be due partly to higher temperatures and other climatic variables. Few studies have investigated the relationship between climate factors and enteric infections, using historic surveillance disease datasets. Positive association between temperature and salmonellosis and other enteric infections has been reported in limited studies (Kovats et al. 2004; D'Souza et al. 2004; Louis et al. 2005; Tam et al. 2006). In terms of the relationship between other climatic variables, e.g. rainfall, humidity, and El Nino, with enteric infections, published results are not consistent, with some studies claiming a positive correlation while others have not detected any associations (Curriero et al. 2001; Checkley et al. 2000; D'Souza et al. 2004; Martinez-Urtaza et al. 2004).

There are some statistical issues to account for in time-series analyses, such as autocorrelations, dependent and independent variables, plus the impact of seasonality and potential secular trends of disease. Climatic variables tend to be clustered (e.g. high temperatures occur on continuous days) and the number of cases may also be auto-correlated. Additionally, the number of cases always has a seasonal variation and a long-term trend of incidence may potentially occur over the study period, which could conceal the association between climatic variables and salmonellosis cases. Therefore, in order to detect the relationship between climatic variables and salmonellosis, these factors were considered in this study. The methods and power of the control for these influencing factors varies between different regression models.

An appropriate regression model is essential to detect the real effects of climate variations on disease transmission. Regression methods most commonly used in investigating the relationship between climate and enteric pathogens include Poisson regression, multiple linear regression and seasonal autoregressive integrated moving average (SARIMA) models (Kovats et al. 2004; Bentham and Langford 1995; D'Souza et al. 2004; Singh et al. 2001). All four models were used to investigate the same dataset of salmonellosis in Adelaide. The relative strengths of these models were compared by residual diagnosis, goodness-of-fit statistics and forecasting ability. This is the first study to examine the association between climate variability and salmonellosis using different regression models.

Materials and methods

Data sources

Adelaide, the capital city of South Australia, with a population of 1.1 million in 2005, has a typical Mediterranean climate. Daily weather records, including maximum temperature, minimum temperature, relative humidity and rainfall, over the period 1990 to 2004 were obtained from the Australian Bureau of Meteorology. The weather station used in this study is the Kent Town station in Adelaide, South Australia, located at latitude 138.6E, longitude 34.9S. It has had excellent weather records since the 1970s and, according to the Australian Bureau of Meteorology, adequately represents the local climate variability. No missing data were found in the meteorological records within the study period. Laboratory-confirmed *Salmonella* cases in Adelaide from the 1 January 1990 to 30 November 2004 were provided by the Communicable Diseases Control Branch (CDCB) of the South Australian Department of Health. These included daily, weekly and monthly counts of salmonellosis over the study period. Salmonellosis is a legally notifiable infectious disease in South Australia, with excellent disease surveillance datasets. The onset dates of disease rather than notified dates were used in this analysis. Outbreaks of salmonellosis were identified by South Australia Department of Health and indicated in the original datasets. The trend in the under-reporting of the number of cases in Adelaide is believed to be relatively stable over the study period (South Australian Department of Health 1991). Although under-reporting will not affect the estimate of the relationship between climatic variables and notified cases, the translation of the estimates into numbers of cases in the community must take account of under-reporting to quantify the burden.

Data analyses

Data were analysed on a weekly basis because there were very few cases on most days over the study period. Salmonellosis cases from outbreaks were excluded from the analysis because cases from outbreaks were due mainly to common sources, which could have had different relationships with climatic variables compared with sporadic cases (Kovats et al. 2004; D'Souza et al. 2004).

Spearman correlation analyses between climatic variables and salmonellosis cases were conducted prior to regression analysis. The lagged effects of climatic variables were explored by cross-correlation analysis, and the climatic variables with the maximum correlation coefficient were included in the regression models. Autocorrelation function (ACF) and partial autocorrelation function (PACF) were used to detect autocorrelations of the salmonellosis cases. Data from 1990–2003 were used for parameter estimation, while 2004 data were used to test the forecasting ability. Regression analyses, including standard Poisson regression, autoregressive adjusted Poisson regression, multiple linear regression and the SARIMA model were performed. The lagged effects of climate variables, seasonality and autocorrelations amongst dependent and independent variables were taken into account in the regression models.

Since the outcome variable (the number of cases) follows an approximate Poisson distribution, the square root transformation of the original number of cases was adopted in the multiple linear regression and in the SARIMA model. Moreover, it is better to increase the standard error by square root transformation because the counts are presumably over-dispersed. The numbers of terms in the SARIMA model were determined by ACF and PACF. In Poisson regression, autoregressive adjusted Poisson regression and multiple linear regression, the seasonal variation was modelled using sine [$\mathrm{Sin}(2\pi t/52)$] and cosine [$\mathrm{Cos}(2\pi t/52)$] functions with a period of 52 weeks, which represent seasonal distribution within one single year. Autocorrelation was controlled in autoregressive adjusted Poisson regression, multiple linear regression and the SARIMA models by including autoregressive variates. The detailed functions of the four regression models are given in the Appendix (Brockwell and Davis 1991; Box and Jenkins 1976; Shumway and Stoffer 2000; Weisstein 2004).

The residual diagnosis (R^2), parameter (β) estimation, goodness-of-fit statistics and the forecasting ability of the four models were compared. The goodness-of-fit and forecasting ability was calculated as:

$$\text{Mean Standard Error (MSE)} = \sum_{weeks=1}^{n} \left[(\text{observed} - \text{expected})^2 \middle/ \text{number of weeks} \right].$$

Smaller MSE values show that expected numbers from the models have a better fit for the observed numbers.

The parameters of Poisson regression and Multiple Linear Regression were estimated by Stata 8.0. (StataCorp 2003. Stata Statistical Software: Release 8. College Station, TX) The SARIMA model was fitted with SPSS 11.5 (SPSS for Windows 2002), using the maximum likelihood algorithms function. The significant level in the analysis was α=0.05.

Results

In total, 4,740 salmonellosis cases, including 1,112 outbreak cases, had been notified between 1990 and 2003. Maximum temperature (r=0.33, 95%CI: 0.27, 0.39), minimum temperature (r=0.32, 95%CI: 0.26, 0.38) and rainfall (r=−0.16, 95%CI: −0.21, −0.11), with relevant lag times, were significantly correlated with the number of salmonellosis cases. A 2-week lagged effect of maximum and minimum temperatures on salmonellosis cases was identified but no lagged effect of rainfall was detected. Correlations among the climatic variables show a high correlation between maximum temperature and minimum temperature (r=0.93) and relatively less correlation between maximum temperature and rainfall (r=−0.42), minimum temperature and rainfall (r=−0.25). Additionally, a four-order autocorrelation of the number of salmonellosis cases was detected by the partial autocorrelation function.

Accordingly, maximum and minimum temperatures with 2-week lags and rainfall without lag were chosen to build up regression models by standard Poisson regression, autoregressive adjusted Poisson regression, multiple linear regression and SARIMA models. Weekly maximum and minimum temperatures were included in separate models due to multicolinearity. Only the models having mean maximum temperature are presented in this paper due to the high similarity with the models having mean minimum temperature.

Standard poisson regression

After controlling for seasonal variation, the standard Poisson regression model suggested that maximum temperature was associated with an increase in the number of cases, while increased rainfall was associated with a decrease in the number of cases (Table 1). The goodness-of-fit was relatively poor with MSE=9.23. The forecasting performance (MSE=8.65) of the model is shown in Fig. 1 (1). In the diagnosis of the model residuals, PACF shows autocorrelations among the residuals (Fig. 2 (1)).

Autoregressive adjusted poisson regression

After controlling for autocorrelation and seasonality, this model indicated that the number of salmonellosis cases in any current week was related to the numbers occurring in the preceding 1, 2, 3 and 4 weeks. Maximum temperature 2 weeks prior was also positively associated with the number of salmonellosis cases, while rainfall in the same week was negatively associated with the number of salmonellosis cases (Table 2). The MSE for goodness- of-fit and forecasting of this model were 7.20 and 7.34, respectively (Table 5). Additionally, the fit of the model and PACF of the residuals were plotted in Fig. 1(2) and Fig. 2(2).

Multiple linear regression

As shown in Table 3, the multiple linear regression model suggested that the number of salmonellosis cases were related to the numbers occurring in the previous 1, 2, 3 and 4 weeks. Maximum temperature with 2-week lags was significantly related to the number of cases. The effect of

Table 1 Parameters estimated by standard poisson regression

	Coefficient	Standard error	z	$p>\|z\|$	95% Confidence interval
Lag2_MaxT[a]	0.015	0.006	2.50	0.012	(0.003, 0.027)
Rainfall	−0.022	0.010	−2.29	0.022	(−0.041, −0.003)
Sin(2πt/52)	0.167	0.029	5.70	0.000	(0.108, 0.222)
Cos(2πt/52)	0.126	0.045	2.77	0.006	(0.037, 0.217)
Constant	1.280	0.134	9.54	0.000	(1.017, 1.543)

[a] Maximum air temperature occurring 2 weeks prior

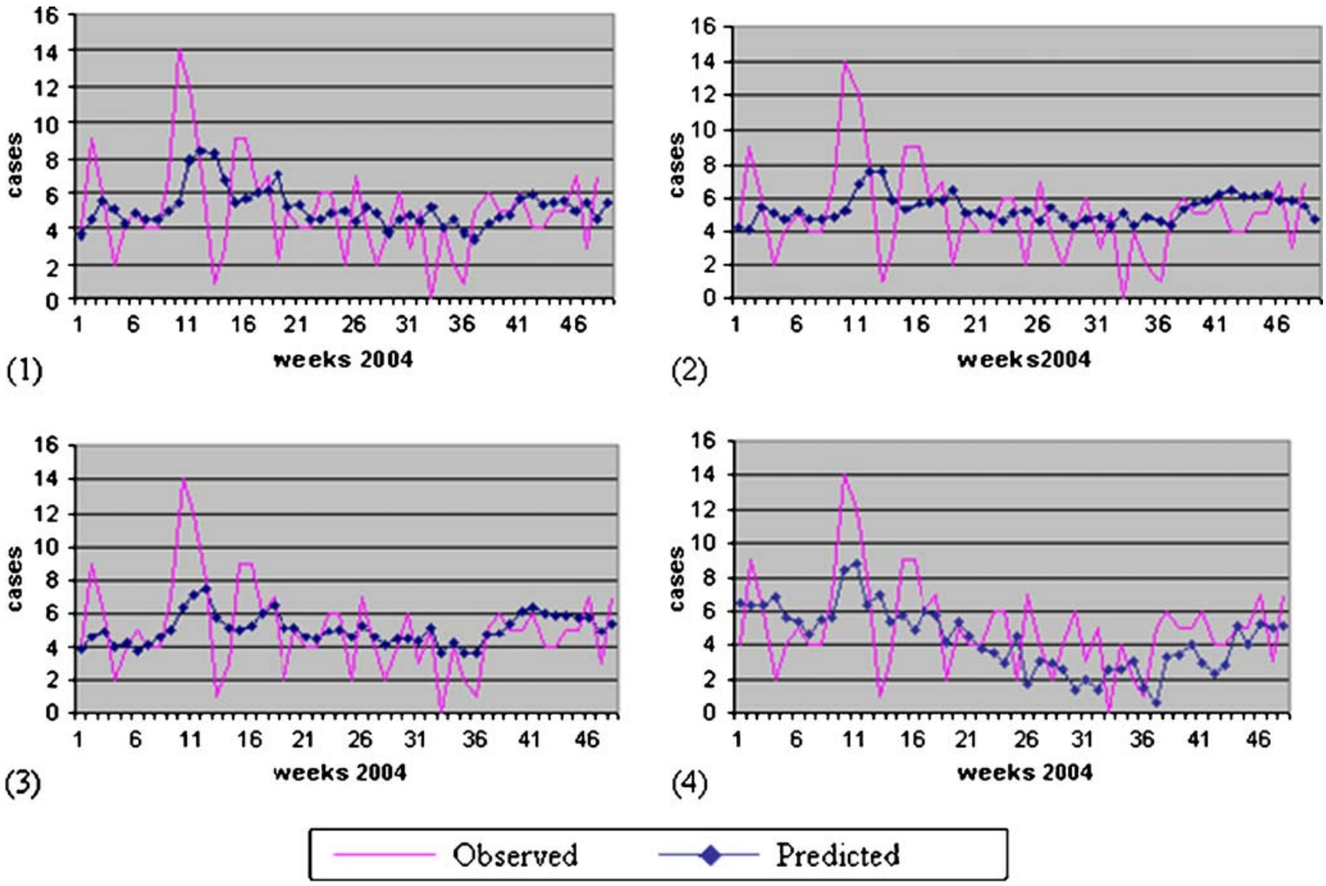

Fig. 1 Predicted vs observed weekly cases in Adelaide in 2004, by standard Poisson regression (*1*), autoregressive adjusted Poisson regression (*2*), multiple linear regression (*3*) and seasonal autoregressive integrated moving average (SARIMA) (*4*)

rainfall was not significant included at a 0.05 level (95% CI), but was at a 0.10 level (90%CI) (Table 3). The MSE for goodness-of-fit and forecasting were 7.52 and 7.35, respectively (Table 5). The fit of the model and PACF of the residuals were plotted in Fig. 1(3) and Fig. 2(3).

Seasonal autoregressive integrated moving average

Based on the results of ACF and PACF, the selected SARIMA model was SARIMA(4,0,0)*$(1,0,0)_{52}$. In addition to a 4-order autoregression, in this model maximum temperature with 2-week lags was significantly related to the number of salmonellosis cases. The association between rainfall and salmonellosis was not significant at a 0.05 level (95%CI) but was at a 0.10 level (90%CI) (Table 4). The MSE for goodness-of-fit and forecasting were 7.17 and 7.20, respectively (Table 5). The fit of the model and PACF of the residuals were plotted in Fig. 1(4) and Fig. 2(4).

Comparison of regression models

The comparison among these regression models, including parameter estimation, residual diagnoses (R^2), goodness-of-fit statistics and forecasting ability (MSE), is summarised in Table 5. All these models showed a significant positive association between temperature (2-week prior) and the number of salmonellosis cases. Rainfall was significantly negatively associated with the number of cases in the two Poisson regression models and was significant at a 0.01 level (90%CI) in the multiple linear regression and SARIMA models. The coefficients of rainfall remained almost the same in the four models. Relative humidity was not significantly included in any of these models. The values of R^2 and MSE indicated that standard Poisson regression had the poorest strength, while the SARIMA model had the best goodness-of-fit and forecasting ability (Fig. 1) among these four regression models. The PACF of the model residuals are shown in Fig. 2.

Discussion

In this study, standard Poisson regression, autoregressive adjusted Poisson regression, multiple linear regression and SARIMA models were performed to quantify the relationship between climate variations and salmonellosis in Adelaide, a temperate city in Australia. All four regression models indicated

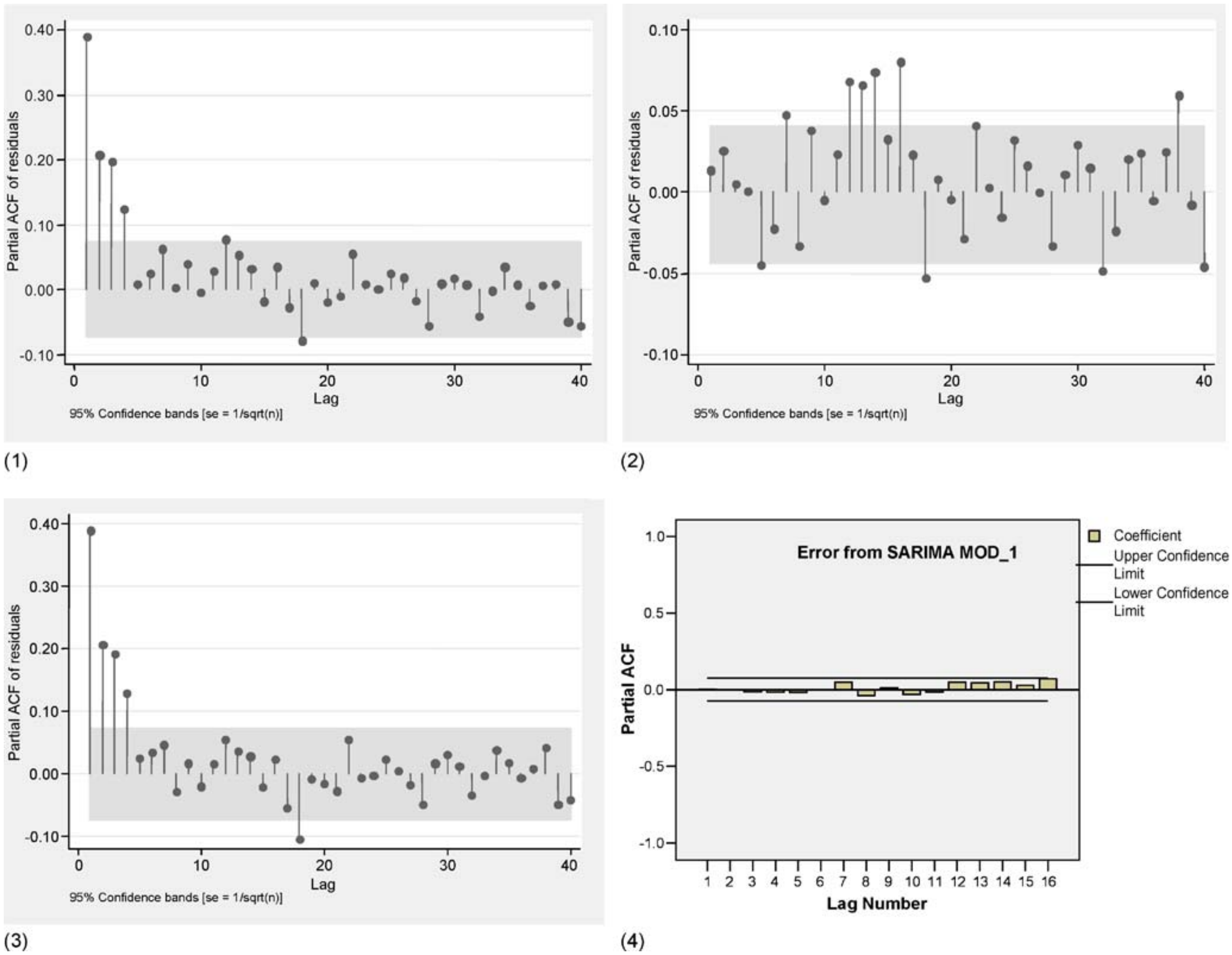

Fig. 2 PACF of the residuals of the regression models: *1* Standard Poisson, *2* autocorrelation adjusted Poisson, *3* multiple linear regression, *4* SARIMA (*1–3 were performed by Stata; 4 performed by SPSS)

that temperature is positively associated with the number of cases of salmonellosis. Additionally, rainfall could affect the transmission of salmonellosis in such a temperate city.

Temperature affects the transmission of food-borne disease in various ways. Temperature directly influences the rate of replication of pathogens and the survival of the pathogens in the environment. Temperature can also affect human behaviour, including eating habits, which influences the transmission of food-borne diseases indirectly. In this study, the four models had a similar estimation of the

Table 2 Parameters estimated by autoregressive adjusted poisson regression

	Coefficient	Standard error	z	$p>\|z\|$	95% Confidence interval
Lag1_cases[a]	0.035	0.005	6.60	0.000	(0.025, 0.046)
Lag2_cases	0.014	0.006	2.41	0.016	(0.003, 0.025)
Lag3_cases	0.020	0.006	3.52	0.000	(0.009, 0.031)
Lag4_cases	0.016	0.005	2.91	0.004	(0.005, 0.026)
Lag2_MaxT[b]	0.017	0.003	5.06	0.000	(0.011, 0.024)
Rainfall	−0.022	0.010	−2.35	0.019	(−0.041, −0.004)
Constant	0.790	0.080	9.84	0.000	(0.633, 0.948)

[a] Cases of salmonellosis occurring 1–4 weeks previously
[b] Maximum air temperature occurring 2 weeks prior

Table 3 Parameters estimated by multiple linear regression

	Coefficient	Standard error	z	$p>\|t\|$	95% Confidence interval
Lag1_Sqrt(cases)[a]	0.180	0.037	4.87	0.000	(0.107, 0.253)
Lag2_Sqrt(cases)	0.097	0.037	2.63	0.009	(0.025, 0.170)
Lag3_Sqrt(cases)	0.160	0.037	4.33	0.000	(0.087, 0.232)
Lag4_Sqrt(cases)	0.105	0.036	2.88	0.004	(0.034, 0.176)
Lag2_MaxT[b]	0.020	0.005	4.06	0.000	(0.010, 0.030)
Rainfall	−0.022	0.013	−1.73	0.083	(−0.047, 0.003)
Constant	0.545	0.134	4.07	0.000	(0.282, 0.808)

[a] Square root of the number of cases with lag values 1–4
[b] Maximum air temperature occurring two weeks prior

association between temperature and the number of cases, which suggested that temperature could be used as a predicting indicator for the number of salmonellosis cases in Adelaide. The temperature occurring 2 weeks prior had the greatest significant association with the number of weekly salmonellosis cases. These findings are consistent with a study in England and Wales from Bentham and Langford (2001), who found an association between temperature and food-poisoning with a lag of 2–5 weeks.

The estimated parameters for rainfall are almost the same in all four regression models, indicating that rainfall could have a negative impact on salmonellosis cases, although rainfall is not significant at 95% CI in all of the models. Unlike other studies, which either did not include rainfall as a climatic variable or did not detect an association between rainfall and disease rates, our study is the first to report that rainfall may affect salmonellosis transmission (Curriero et al. 2001; Nath et al. 1992; Pinfold et al. 1991; Kovats et al. 2004). However, this association may be due to the specific climatic characteristics in Adelaide, which has wet winters (lower temperature and more rainfall) and dry summers (higher temperature but less rainfall). The significant association between rainfall and temperature make the interpretation of the effects of rainfall on the number of cases complex. It is still not clear whether rainfall could have affected the disease transmission independently, or interacted with other climatic variables such as temperature. Moreover, the effect of rainfall on food-borne disease transmission may vary in different climatic areas. Therefore, further research is necessary to fully understand the effect of rainfall on food-borne diseases.

Time-series data are always not independently distributed but demonstrate autocorrelation with adjacent observations. It is important to control for autocorrelation in the regression analysis. If autocorrelation is not controlled for, the regression model will have autocorrelated residuals and unrealistically narrow confidence limits of the estimates (Brockwell and Davis 1991). In this study, the number of salmonellosis cases is an order-4 autocorrelation series, which should be considered in regression analysis. Biologically, the 4-week lag could be due to the growth and development of *Salmonella* in natural reservoirs, the time to transport the food from farm to retail, and the period from shop to consumer’s fridge, as well as the incubation period for salmonellosis to develop within human body.

The assumptions of a standard Poisson model may be infringed because it requires that outcome variables are independent. However, the number of salmonellosis cases assumes the Poisson distribution. Hence, an autoregressive adjusted Poisson regression model is better than a standard Poisson regression model and is more commonly used in epidemiological research because it considers both autocorrelation and the Poisson distribution (Kovats et al. 2004). In multiple linear regression, autoregressive variables were introduced into the model as shown in the appendix. Auto-regression is an intrinsic component of the SARIMA model.

Another challenge for time-series data is how to deal with the effect of potential seasonality. Most cases of enteric infection demonstrate a seasonal distribution, which implies that seasonal changes may affect the spread of infectious diseases by altering host behaviours, contact rates and multiplication of pathogens in the environment (Altizer et al. 2006). Seasonality can be represented as

Table 4 Parameters estimated by seasonal autoregressive integrated moving average (SARIMA) $(4,0,0)*(1,0,0)_{52}$

	B	SEB	*t*	*P*
AR1[a]	0.181	0.037	4.895	0.000
AR2	0.108	0.037	2.906	0.004
AR3	0.164	0.037	4.399	0.000
AR4	0.110	0.037	2.958	0.003
Seasonal_ AR1[b]	0.010	0.038	2.368	0.018
Lag2_MaxT[c]	0.025	0.007	3.456	0.000
Rainfall	−0.022	0.012	−1.746	0.081
Constant	1.585	0.174	9.123	0.000

[a] ~4: 1–4 order auto-regression
[b] 1-order seasonal auto-regression
[c] Maximum air temperature occurring 2 weeks prior

Table 5 Comparison of four regression models

Models	Goodness-of-fit (MSE[a])	Predictive ability for 2004 cases (MSE[a])	Standard deviation	Pseudo R^2	Control of auto-correlation	Control of seasonality	Expected increase in number of cases for 1°C rise in maximum temperature	Coefficients in models (standard error)	
								Lag2_MaxT[b]	Rainfall
Standard Poisson regression	9.23	8.65	2.23	0.05	No	Yes	5.6%~16.3%	0.0149[c] (0.006)	−0.0221[c] (0.010)
Autoregressive adjusted Poisson regression	7.20	7.34	2.22	0.10	Yes	Yes	10.6%~15.2%	0.0172[c] (0.003)	−0.0224[c] (0.010)
Multiple linear regression ($\sqrt{cases}$)	7.52	7.35	0.29	0.22	Yes	Yes	10.2%~17.3%	0.0203[c] (0.005)	−0.0221[d] (0.013)
SARIMA ($\sqrt{cases}$)	7.17	7.20	0.66	0.21	Yes	Yes	10.6%~19.7%	0.0250[c] (0.007)	−0.0216[d] (0.012)

[a] Mean squared error
[b] Maximum air temperature two weeks prior
[c] $P<0.05$
[d] $P<0.10$

simple annual cycles or more complex fluctuations over years, which should be considered in examining the effects of climatic variables on infectious disease. If this is not done, the effect of short-term variables cannot be distinguished from seasonal climatic variations. It is important to control for seasonality before attempting to describe the association between short-term variables, such as maximum temperature and rainfall, and salmonellosis cases. In this study, seasonality was controlled for in all models except in the SARIMA model, by including two trigonometric functions to represent the seasonal fluctuation in the number of cases. No extra variable was included in the SARIMA model because an inherent seasonal function is adopted in the time-series method (Shumway and Stoffer 2000; Tobias and Saez 2004). Although this is rather general, there is no suggestion that the trigonometric functions are inadequate. The two trigonometric functions are appropriate for controlling the seasonal fluctuations in the adjusted Poisson and multiple linear regression. However, the trigonometric functions are not as effective for controlling seasonality in the SARIMA model as demonstrated by the residual plots.

Our study is the first to compare the different regression models used to investigate the association between climate and diarrhoeal diseases. According to our analysis, the SARIMA model could represent the most appropriate model for such analysis, because it has integrated functions controlling seasonal variation, autocorrelation and long-term trend. The standard Poisson regression has not controlled for autocorrelation, which could underestimate the standard errors of the estimates (Weisstein 2004). In addition, it is insufficient simply to estimate the peak values. After adjusting for autocorrelation, a better fit is presented in autoregressive adjusted Poisson regression. In multiple linear regression, the square root of the original number of cases was used to stabilize the variance. The goodness-of-fit of this model was slightly poorer than that with the autoregressive adjusted Poisson regression. Taken together, the results demonstrate that the SARIMA model has both the best forecasting ability and the best goodness-of-fit.

Conclusion

Temperature and rainfall affect the transmission of salmonellosis in Adelaide. All models demonstrate a consistent positive association between temperatures and salmonellosis and a negative association between rainfall and salmonellosis in Adelaide. Compared with other traditional regression models, including Poisson regression and multiple linear regression, the SARIMA model is more effective in epidemiological studies of the association between climatic variations and enteric infections. However, only a few epidemiological studies have used SARIMA models because they have only recently been applied to this type of research (Tong et al. 2002; Bi et al. 2003). Our study suggests that the SARIMA model should be applied to more studies in this field to better understand the effect of climate variability on population health. Further regression models, including other covariates that may also have effects on the transmission of enteric infections, such as socio-economic status, food storage or environmental hygiene, should be developed and applied in other climatic/ecological regions.

Acknowledgements We greatly appreciate Dr Andrew Metcalfe, Senior Lecturer in Statistics at the University of Adelaide, for his kind assistance with statistical issues in this study.

Appendix

1. Standard poisson regression.

In the standard Poisson regression analysis, the dependent variable is weekly salmonellosis cases. The Poisson distribution has only one parameter, ν, which equals the mean (or variance). The distribution is given by the formula (Weisstein 2004):

$$\Pr(\mathrm{Y}) = \frac{\nu^n e^{-\nu}}{n!}$$

The estimation of the parameters is obtained by a maximum likelihood function. The model can be summarized as: ln (ν_t)=α+β_1 temperature$_t$+β_2 rainfall$_t$+sin(2πt/52)+cos(2πt/52)

2. Autoregressive adjusted poisson regression.

According to our previous analysis, there is autocorrelation between both dependent and independent variables. Therefore, the Poisson regression model adjusted for autocorrelation can be given as:

$$\begin{aligned}\ln(\nu_{\mathrm{t}}) = {} & \alpha + \beta_1 \mathrm{Y}_{\mathrm{t}-1} + \beta_2\,\mathrm{Y}_{\mathrm{t}-2} + \beta_3 \mathrm{Y}_{\mathrm{t}-3} + \beta_4 \mathrm{Y}_{\mathrm{t}-4} \\ & + \beta_5 \mathrm{temperature}_{\mathrm{t}} + \beta_6 \mathrm{temperaruep}_{\mathrm{t}-1} \\ & + \beta_7 \mathrm{temperaturep}_{\mathrm{t}-2} + \beta_8 \mathrm{rainfall}_{\mathrm{t}} + \beta_9 \mathrm{rainfall}_{\mathrm{t}-1} \\ & + \beta_{10} \mathrm{rainfall}_{\mathrm{t}-2} + \sin(2\pi \mathrm{t}/52) + \cos(2\pi \mathrm{t}/52)\end{aligned}$$

3. Multiple linear regression.

The regression model is:

$$\begin{aligned}\sqrt{Yt} = {} & \alpha + \beta_1\sqrt{Y_{t-1}} + \beta_2\sqrt{Y_{t-2}} + \beta_3\sqrt{Y_{t-3}} + \beta_4\sqrt{Y_{t-4}} \\ & + \beta_5 \mathrm{temperature}_{\mathrm{t}} + \beta_6 \mathrm{temperaute}_{\mathrm{t}-1} \\ & + \beta_7 \mathrm{temperature}_{\mathrm{t}-2} + \beta_8 \mathrm{rainfall}_{\mathrm{t}} + \beta_9 \mathrm{rainfall}_{\mathrm{t}-1} \\ & + \beta_{10} \mathrm{rainfall}_{\mathrm{t}-2} + \sin(2\pi \mathrm{t}/52) + \cos(2\pi \mathrm{t}/52) + Et\end{aligned}$$

Where Et is independent random error with mean 0.

4. Seasonal autoregressive moving average (SARIMA) models.

An autoregression moving average (ARMA) model predicts the outcome variable from the values of the outcome at previous time points. It is an approach to handle time-series modelling and forecasting based on the landmark work of Box and Jenkins (1976). The theory and application of the models have been described in many papers and books (Shumway and Stoffer 2000; Brockwell and Davis 1991; Tobias and Saez 2004). The ARMA model is suitable only for stationary processes. The SARIMA model allows for a trend and seasonal effects by differencing. The model described here SARIMA(4,0,0)*$(1,0,0)_{52}$, which stands for 4-order autoregression and 1-order seasonal autoregression with a period of 52 weeks, has the following form. Let X_t be the square root of Y_t. Define Wt by $W_t=X_t-X_{t-52}$. Then

$$\begin{aligned}Wt = {} & \alpha_0 + \alpha_1 W_{\mathrm{t}-1} + \alpha_2\,W_{\mathrm{t}-2} + \alpha_3\,W_{\mathrm{t}-3} + \alpha_4\,W_{\mathrm{t}-4} \\ & + \beta \mathrm{E}_{\mathrm{t}-52} + \gamma_1 \mathrm{Temperature}_{\mathrm{t}} + \gamma_2 \mathrm{Temperaute}_{\mathrm{t}-1} \\ & + \gamma_3 \mathrm{Temperature}_{-2} + \gamma_4 \mathrm{rainfall}_{\mathrm{t}} + \gamma_5 \mathrm{rainfall}_{\mathrm{t}-1} \\ & + \gamma_6 \mathrm{rainfall}_{\mathrm{t}-2} E_{\mathrm{t}}\end{aligned}$$

Where E_t is independent random error with mean 0.

The Aikake information criterion (AIC) is used to compare models fit to one same series. The model with the smaller AIC fits the data better. AIC=−2 ln(L)+2 k , where L is the likelihood function and k is the number of free parameters.

References

Adak G, Long SM, O'Brien S (2002) Trends in indigenous foodborne disease and deaths, England and Wales, 1992–2000. Gut 51:832–841

Altizer S, Dobson A, Hosseini P, Hudson P, Pascual M, Rohani P (2006) Seasonality and the dynamics of infectious diseases. Ecol Lett 9:467–484

Bentham G, Langford IH (1995) Climate change and the incidence of food poisoning in England and Wales. Int J Biometeorol 39:81–86

Bentham G, Langford IH (2001) Environmental temperatures and the incidence of food poisoning in England and Wales. Int J Biometeorol 45:22–26

Bi P, Tong S, Donald K, Parton KA, Ni J (2003) Climatic variables and transmission of malaria: a 12-year data analysis in Shuchen County, China. Public Health Rep 118:65–69

Box GEP, Jenkins GM (1976) Time series analysis: forecasting and control. Holden-Day, San Francisco

Brockwell PJ, Davis RA (1991) Time series: theory and methods. Springer, Berlin Heidelberg New York

CDC (2004) Salmonellosis. http://www.cdc.gov/ncidod/dbmd/diseaseinfo/salmonellosis_t.htm. Accessed 2 April 2004

Checkley W, Epstein L, Gilman R, Figueroa D, Cama R, Patz J, Black R (2000) Effect of El Nino and ambient temperature on hospital admissions for diarrhoeal diseases in Peruvian children. Lancet 355:442–450

Curriero F, Patz J, Rose J, Lele S (2001) The association between extreme precipitation and waterborne disease outbreaks in the United States, 1948–1994. Am J Public Heal 91:1194–1199

D’Souza R, Becker N, Hall G, Moodie K (2004) Does ambient temperature affect foodborne disease? Epidemiology 15:86–92

Hall G, Kirk DM, Becker N, Gregory J, Unicomb L, Millard G, Stafford R, Lalor K, the OzFoodNet Working Group (2005) Estimating foodborne gastroenteritis. Emerg Infect Dis 11:1257–1264

Kovats R, Edwards S, Hajat S, Armstrong B, Ebi K, Menne B (2004) The effect of temperature on food poisoning: a time-series analysis of salmonellosis in ten European countries. Epidemiol Infect 132:443–453

Louis V, Gillespie I, O’Brien S, Russek-Cohen E, Pearson A, Colwell R (2005) Temperature-driven Campylobacter seasonality in England and Wales. Appl Environ Microbiol 71:85–92

Martinez-Urtaza J, Saco M, de Novoa J, Perez-Pineiro P, Peiteado J, Lozano-Leon A, Garcia-Martin O (2004) Influence of environmental factors and human activity on the presence of *Salmonella serovars* in a marine environment. Appl Environ Microbiol 70:2089–2097

Nath G, Singh S, Sanyal S (1992) Childhood diarrhoea due to rotavirus in a community. Indian J Med Res 95:259–262

OzFoodNet (2002) The OzFoodNet working group. http://www.ozfoodnet.org.au

Pinfold J, Horan N, Mara D (1991) Seasonal effects on the reported incidence of acute diarrhoeal disease in northeast Thailand. Int J Epidemiol 20:777–786

Shumway RH, Stoffer DS (2000) Time series analysis and its applications. Springer, Berlin Heidelberg New York

Singh R, Hales S, Wet N, Raj R, Hearnden M, Weinstein P (2001) The influence of climate variation and change on diarrheal disease in the Pacific Islands. Environ Health Perspect 109:55–59

South Australian Department of Health (1991) Communicable diseases control branch: csmart computing program. South Australian Department of Health, Adelaide

SPSS for Windows (2002) Rel. 11.5.1. 2002. SPSS, Chicago

StataCorp (2003) Stata statistical software: release 8. College station, TX

Tam CC, Rodrigues LC, O’Brien SJ, Hajat S (2006) Temperature dependence of reported Campylobacter infection in England, 1989–1999. Epidemiol Infect 134:119–125

Tobias A, Saez M (2004) Time-series regression models to study the short-term effects of environmental factors on health. In: http://www.udg.es/fcee/economia/n11.pdf. Accessed 2 April 2004

Tong S, Bi P, Donald K, McMichael AJ (2002) Climate variability and Ross River virus transmission. J Epidemiol Commun Health 56:617–621

Weisstein EW (2004) Poisson distribution. In: Wolfram A (ed) Mathworld. Web Resource. http://mathworld.wolfram.com/PoissonDistribution.html

WHO (2005) Drug-resistant Salmonella. http://www.who.int/mediacentre/factsheets/fs139/en/ Accessed 2 April 2005