A STATISTICAL METHOD FOR ESTIMATING UNDER-REPORTED INCIDENCE RATES WITH APPLICATION TO CHILD DIARRHEA IN THAI PROVINCES BORDERING CAMBODIA

Sulawan Yotthanoo[1] and Chamnein Choonpradub[2]

[1]Department of Statistics, School of Science and Technology, Naresuan University, Phayao; [2]Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani, Thailand

Abstract. Diarrhea is a major health problem in Thailand, but reported data of disease incidence are known or suspected to be under-reported. This study aimed to develop a statistical model for estimating the annual incidence of hospital diarrhea cases among children under five years. Data regarding diarrhea patients 0-4 years old were collected for the National Notifiable Disease Surveillance (Report 506) about Thai provinces bordering Cambodia during 1999-2004 by the Ministry of Public Health. A log-linear regression model based on the prevailing seasonal-trend pattern was used for diarrhea incidence as a function of quarter, year and district, after imputing rates where under-reporting was evident, using populations obtained from the 2000 population census. The model also takes any spatial correlation between districts into account, using the generalized estimating equation (GEE) method. Diarrhea incidence had seasonal peaks in the first quarter (January to March) and the trend steadily increased from 1999 to 2004. Results from such studies can help health authorities develop prevention policies.

Key words: child diarrhea, statistical model, BEE method, linear regression

INTRODUCTION

Diarrhea is one of the world's top five infectious disease causes of death (Brownlie *et al*, 2006) and remains a major cause of morbidity and mortality among children in developing countries (Carlos and Saniel, 1990; Parashar *et al*, 2003). Children under five years of age have an average of 3.3 diarrhea episodes per year, and more than one-third of all deaths in this age group are associated with diarrhea. Approximately 1.5 billion diarrhea episodes and 4 million deaths occur annually among children age less than five years (Vargas *et al*, 2004).

As in other developing countries, diarrhea in Thailand is a major health problem and accounts for approximately 50% of all hospital-reported infectious diseases (Thai Working Group on Burden of Disease and Injuries, 2002). The Bureau of Epidemiology (2002) reported that while diarrhea-related mortality declined from 1.11 per 100,000 in 1988 to 0.23 per 100,000 in 2002, morbidity increased from 1,488 cases per 100,000 in 1993 to 1,687 cases per 100,000 in 2002. In 2002, there were 1,055,393 cases of diarrhea in

---

Correspondence: Chamnein Choonpradub, Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkla University, Pattani Campus, Pattani 94000, Thailand.
Tel: 66 (0) 89 4660 803; Fax: 66 (0) 7331 2729
E-mail: cchamnein@bunga.pn.psu.ac.th

Thailand, of which one-third occurred among children under five years of age and 12% required hospitalization (Jiraphongsa *et al*, 2005).

Given that diarrhea morbidity remains high in Thailand, there is a need to improve treatment to prevent the disease, especially in provinces bordering Cambodia where cross-border migration may be a factor (Thimasarn *et al*, 1995; Konchom *et al*, 2003).

We investigated a statistical model based on linear regression for estimating the extent of under-reporting. We then classified patterns of child diarrhea in Thai provinces bordering Cambodia using the GEE model.

MATERIALS AND METHODS

Study area and data source

The border between Thailand and Cambodia is approximately 800 km long, stretching along the provinces of the lower northeast area of Thailand from a point known as "Chong Bog" in Ubon Ratchathani Province and ending in the Had Lek Subdistrict of Khlong Yai in Trat Province (Fig 1). Due to its stronger economic growth, Thailand attracts many migrant workers from Cambodia. Cross-border migration has been connected with health problems, including infectious diseases (Thimasarn *et al*, 1995; Konchom *et al*, 2003).

Data regarding diarrhea cases from 1999 to 2004 in the border provinces of interest were taken from the National Notifiable Disease Surveillance Report (506), Bureau of Epidemiology, Ministry of Public Health. Each record contains the type of infectious disease, age, gender, subdistrict of residence, date of hospitalization, and disease severity of the patient. The resident population denominator used to compute the annual incidence was obtained from the Population and Housing Census of 2000, performed by the National Statistics Office of Thailand.

Data analysis

Although the registry included the village of residence and date of hospitalization of the patient, we used districts (statistical regions containing up to hundreds of villages ranging in population from 795 to 21,409). We did this to substantially reduce correlations between annual incidence outcomes in successive periods of time and in neighboring locations, while still enabling trends for place and time to be identified. Data from the Thai infectious disease registry are known, or suspected, to be seriously under-reported (Lumbiganon *et al*, 1990; Saengwonloey *et al*, 2003; Intusoma *et al*, 2008).

We first calculated disease incidence in cells defined by district (i) and month (j) of year (t) as the ratio of the number of reported cases (n_{ijt}) to the district population in 1,000s (P_i). For reasons evident from a detailed study of monthly disease counts (Table 2), any occurrence of zero cases in a cell was considered as a possible instance of under-reporting, and an additive linear model was fitted to the logarithms of the remaining incidence rates, namely,

$$ln\left(\frac{n_{ijt}}{P_i}\right) = y_{ijt} = \mu + \alpha_i + \beta_j + \gamma_t . \qquad (1)$$

In this model μ is a constant and α_i, β_j and γ_t are the effects of district $i=1,2,...,106$, month $j=1,2,...,12$ and year $t=1,2,...,6$, respectively, with zero means. After examining the plot of standardized residuals, this model was refitted after further omission of cells corresponding to residuals below a specified cut-off value. Having thus obtained an acceptable fit, the omitted occurrences were then imputed using the model. Next, we aggregated the monthly data into quarterly incidence rates and fitted a model similar to (1) with j now representing quarter instead of month, and the residuals from this model were used to compute correlation coefficients between different districts. The averages of the correlation coefficients within each province were then used to fit a generalized estimating equation (GEE) model (Liang and Zeger, 1986; Yan and Fine, 2004) having a fixed block-diagonal correlation structure with the blocks corresponding to provinces. Residuals from this model were examined using normal quantile plots after filtering to remove their estimated correlation.

To obtain unbiased estimates of incidence rates in cells we used the formula

$$\hat{r}_{ijt} = \exp(\hat{y}_{ijt} + c), \qquad (2)$$

where $\hat{y}_{ijt}$ is the fitted value of y_{ijt} and c is a constant chosen to match the total number of observed cases with the total given by the model. After fitting the model, unbiased incidence rates for levels of each factor adjusted for other factors were calculated similarly. Standard errors for these adjusted incidence rates were obtained using sum contrasts (Venables and Ripley 2002) to compare the incidence rates for each level of a factor with the overall mean incidence rate.

Since the confidence intervals for factor-specific incidence rates and proportions obtained from this model (using the sum contrasts) divide naturally into three groups according to their location entirely above the mean, around the mean, or entirely below the mean, we used this trichotomy to create schematic maps of districts according to their estimated diarrhea annual incidence rates and under-reporting percentages.

We also estimated the extent of under-reporting data by district, quarter and year, by fitting a simple logistic regression model (Hosmer and Lemeshow, 2000; Kleinbaum and Klein, 2002) to the corresponding proportions imputed using the method described above. These estimated proportions for levels of each factor after adjusting for the other factors in the model were again computed by requiring the weighted average of the adjusted proportions match the overall proportion, using a Newton-Raphson iteration procedure with Marquardt damping, and standard errors for differences between individual proportions and the overall mean.

All statistical analysis was carried out using the R program (R Development Core Team, 2007).

RESULTS

Preliminary analysis

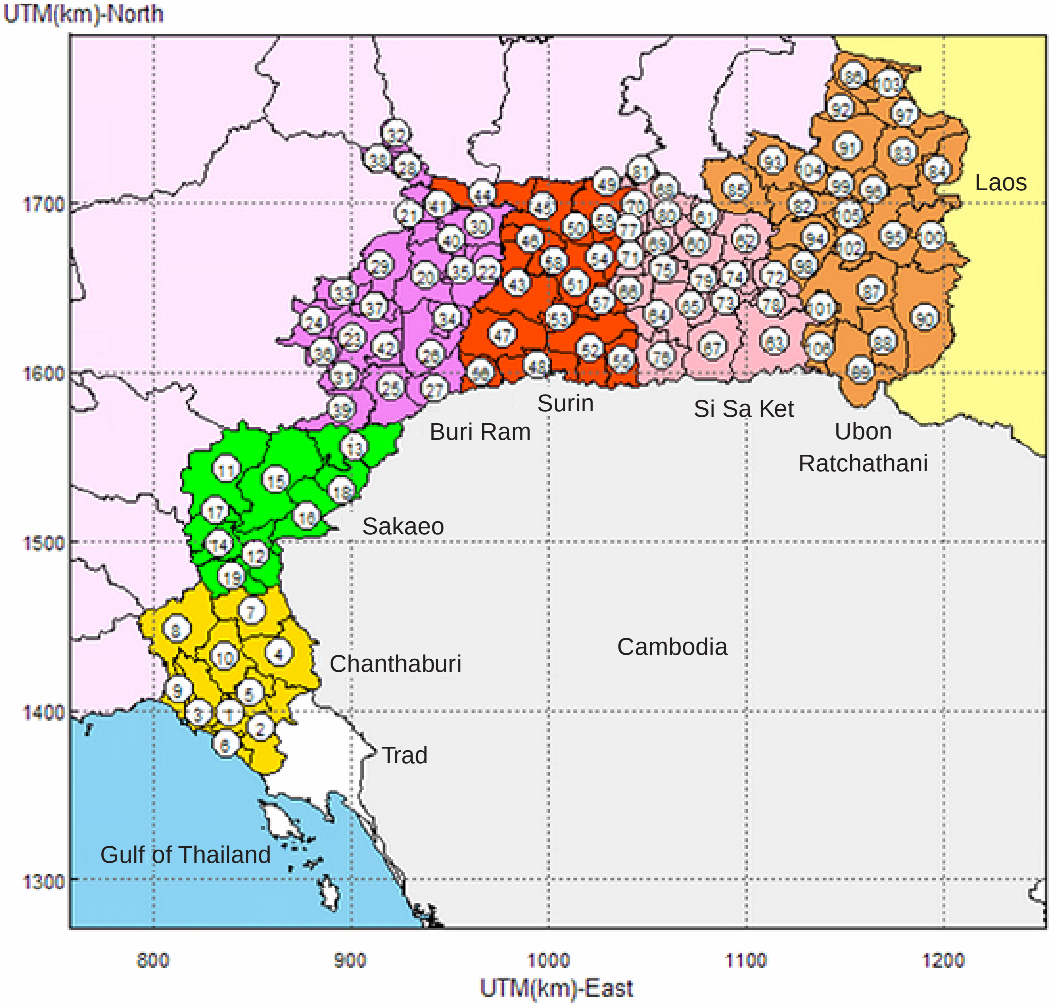
Preliminary analysis indicated gender was not a major factor influencing diarrhea incidence (Ardkeaw and Tongkumchum, 2009), therefore it was not included in the model. Children less than five years old had the highest age-specific annual incidence rate (Bureau of Epidemiology, 2007), so this group was selected for study.

Data from Trat Province were not available for 2002 and so were excluded from the analysis.

During the study period from January 1999 to December 2004, 260,522 cases of diarrhea were reported from district hospitals in the Thai-Cambodia border provinces among children less than 5 years old. The number of cases reported in a month for each district varied from zero to 578 (average annual incidence rate 69.4 per 1,000). Among the six provinces, Surin (78.8 per 1,000) and Buri Ram (78.2 per 1,000) had relatively high rates (Table 1).

Under-reporting

Table 2 shows the number of monthly reported diarrhea cases in ten selected districts. Six of these districts Pho Si Suwan, Mueang Chan, Ban Mai Chaiyaphot, Sila Lat, Nam Kliang and Chalermphrakiet, had



1 Mueang Chantchaburi
2 Khlung
3 Tha Mai
4 Pong Nam Ron
5 Makham
6 Laem Sing
7 Soi Dao
8 Kaeng Hang Maeo
9 Na Yai Am
10 Khao Khitchakut
11 Mueang Sa Kaeo
12 Khlong Hat
13 Ta Phraya
14 Wang Nam Yen
15 Wattana Kakhon
16 Aranyaprathet
17 Khao Chakan
18 Khok Sung
19 Wang Sombun
20 Mueang Buri Ram
21 Khu Mueang
22 Krasang
23 Nang Rong
24 Nong Ki
25 Lahan Sai
26 Prakhon Chai
27 Ban Kruat
28 Phu Thai Song
29 Lam Plai Mat
30 Satuek
31 Pakham
32 Na Pho
33 Nong Hong
34 Phlapphla Chai
35 Huai Rat
36 Non Suwan
37 Chamni
38 Ban Mai Chaiyaphot
39 Non Din Daeng
40 Ban Dan
41 Khaen Dong
42 Chaloem Phra Kiet
43 Mueang Surin
44 Chumphon Buri
45 Tha Tum
46 Chom Phra
47 Prasat
48 Kap Choeng
49 Rattanaburi
50 Sanom
51 Sikhoraphum
52 Sangkha
53 Lamduan
54 Samrong Thap
55 Buachet
56 Phanom Dong Rak
57 Si Narong
58 Khwao Sinarin
59 Non Narai
60 Mueng Si Sa Ket
61 Yang Chum Noi
62 Kanthararom
63 Kantharalak
64 Khukhan
65 Phrai Bueng
66 Prang Ku
67 Khun Han
68 Rasi Salai
69 Uthumphon Phisai
70 Bung Bun
71 Huai Thap Than
72 Non Khun
73 Si Rattana
74 Nam Kliang
75 Wang Hin
76 Phu Sing
77 Mueang Chan
78 Benchalak
79 Phayu
80 Pho Si Suwan
81 Sila Lat
82 Mueang Ubon Ratchathani
83 Si Mueang Mai
84 Khong Chiamk
85 Khueang Nai
86 Khemarat
87 Det Udom
88 Na Chaluai
89 Nam Yuen
90 Buntharik
91 Trakan Phuet Phon
92 Kut Khaopun
93 Mueang Sam Sip
94 Warin Chamrap
95 Phibun Mangsahan
96 Tan Sum
97 Pho Sai
98 Samrong
99 Don Mot Daeng
100 Sirindhorn
101 Thung Si Udom
102 Na Yai
103 Na Tan
104 Lao Suea Kok
105 Sawang Wirawong
106 Nam Khun

Fig 1–Districts of Thai provinces bordering Cambodia (excluding Trad).

Table 1
Number of reported diarrhea cases and population in each province.

Province	No. districts	Diarrhea cases	Population	Annual Incidence Rate
Chanthaburi	10	12,552	33,108	63.2
Sa Kaeo	9	18,849	42,560	73.8
Surin	23	64,574	136,634	78.8
Buri Ram	17	58,940	125,642	78.2
Si Sa Ket	22	42,798	130,875	54.5
Ubon Ratchathani	25	62,809	153,410	68.2

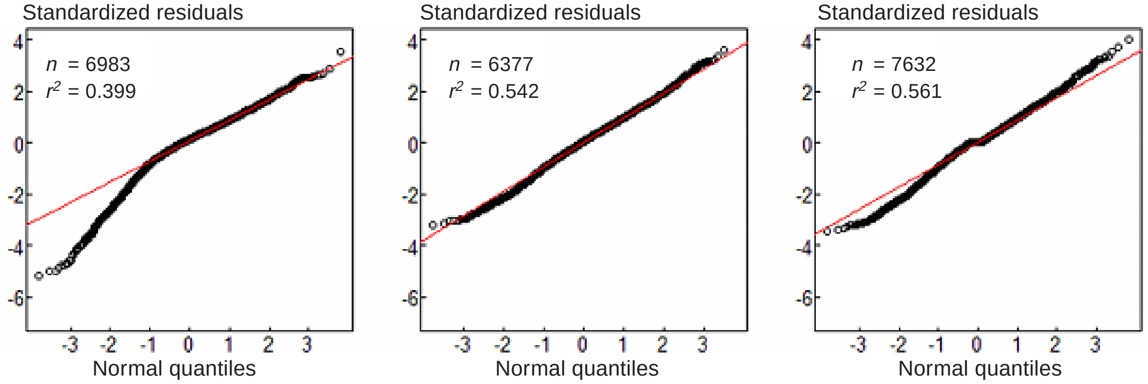


Fig 2–Plots of standardized residuals against normal quantiles with zero cell counts omitted (left panel), further cells with low residuals omitted (middle panel) and imputed using the fitted model (right panel).

stretches of successive unusually low case counts. In contrast, the four remaining selected districts (Na Yai Am, Khwao Sinarin, Khlung, and Chom Phra) showed no such evidence of under-reporting. Any outcome of zero or an extremely low number of reported cases was considered as a possible case of under-reporting.

The method involved first aggregating diarrhea cases by district counts for month and year from 260,522 individual cases into 7,632 records by cross-tabulating 106 districts over 72 months. The left panel of Fig 2 shows the plot of standardized residuals against normal quantiles after omitting the 649 cells with zero counts (8.5%) based on model (1). The residuals indicated a poor fit, which improved substantially when a further 606 cells with residuals less than -1.4 were omitted (7.9%), as the middle panel shows. We thus used this latter model to impute cell counts for the omitted data, obtaining the plot shown in the right panel of Fig 2.

Table 3 lists the estimated proportions of under-reported cases. These proportions were highest in the October to December quarter and lowest in the April to June quarter. Similarly the estimated under-reporting rate was lowest in the year 1999 and gradually increased to 15.2% in 2001 and then decreased to 7.7% in 2004.

Based on the criterion we used, there were only four districts with no evidence of under-reporting (Mueang Sa Kaeo, Ban Kruat, Chom Phra and Phrasat). Chom Phra

Table 2
Number of monthly reported diarrhea cases in ten selected districts.

Month ID	Pho Si Suwan	Mueang Chan	Ban Mai Chaiyaphot	Sila Lat	Nam Kliang	Chaloem Phra Kiet	Na Yai Am	Khwao Sinarin	Khlung	Chom Phra
1	0	0	10	18	0	53	57	88	59	193
2	0	0	10	7	0	19	24	22	48	141
3	0	0	0	19	0	34	26	15	37	88
4	0	0	3	12	0	11	42	18	37	44
5	0	0	1	16	0	9	50	18	32	59
6	0	0	4	24	0	25	35	33	32	53
7	0	0	1	8	0	26	21	16	33	43
8	0	0	4	0	0	16	20	20	33	47
9	0	0	1	18	5	9	33	21	28	45
10	0	0	3	4	15	16	30	27	33	58
11	0	0	2	5	1	10	17	27	22	73
12	0	0	0	7	0	1	18	45	31	121
13	5	13	3	0	88	74	41	55	75	101
14	9	25	10	0	64	57	48	51	58	112
15	25	2	4	8	66	62	39	30	41	92
16	18	0	3	12	53	48	34	31	44	45
17	13	0	6	21	55	68	35	47	49	69
18	37	0	2	4	72	69	23	29	54	115
19	0	1	7	10	35	63	23	45	51	110
20	0	0	4	0	21	40	18	27	31	60
21	0	0	2	0	65	31	25	26	38	53
22	0	0	1	0	17	37	20	25	48	62
23	0	0	1	0	28	48	35	29	36	41
24	0	0	0	0	0	59	21	42	31	79
25	17	20	10	0	21	77	34	56	53	91
26	9	17	5	0	15	61	43	31	63	74
27	6	10	7	0	16	60	52	48	34	72
28	2	3	2	0	37	64	59	34	29	49
29	0	0	1	0	108	54	26	33	27	57
30	0	0	6	0	160	71	32	53	42	84
31	0	0	2	0	81	32	32	67	37	111
32	0	0	0	0	56	33	39	53	31	118
33	0	0	0	0	45	22	26	33	30	85
34	7	0	0	0	37	20	9	37	27	36
35	9	0	0	0	0	11	34	19	31	39
36	14	0	0	1	0	0	13	37	11	39
37	25	42	3	91	0	0	42	67	55	381
38	51	20	6	60	1	0	70	58	87	173
39	34	4	1	66	1	1	63	49	94	187
40	26	30	5	54	2	3	33	48	61	91
41	2	30	0	16	0	4	38	32	37	113
42	1	24	26	17	0	5	53	38	47	153
43	0	19	31	30	0	7	36	42	46	178
44	2	12	14	31	2	7	37	48	39	178
45	2	22	7	33	5	7	30	41	26	94
46	5	12	3	24	3	0	22	45	27	118
47	6	0	24	20	1	0	24	68	38	136
48	5	0	9	30	2	0	28	28	55	104
49	5	2	13	2	2	18	30	100	68	187
50	16	0	12	4	4	44	65	77	46	217
51	14	1	19	2	6	34	74	46	51	171
52	3	1	14	8	1	48	26	27	32	103
53	1	0	17	5	3	32	38	17	54	105

Table 2 (Continued).

Month ID	Pho Si Suwan	Mueang Chan	Ban Mai Chaiyaphot	Sila Lat	Nam Kliang	Chaloem Phra Kiet	Na Yai Am	Khwao Sinarin	Khlung	Chom Phra
54	0	1	16	4	7	13	47	36	48	91
55	4	0	14	4	2	7	34	37	48	117
56	5	2	10	6	3	9	27	29	51	123
57	1	1	46	7	1	8	32	22	73	56
58	4	0	52	2	2	0	26	36	54	76
59	2	1	87	1	4	0	17	25	45	99
60	5	1	6	0	3	0	15	48	11	106
61	51	85	111	54	98	0	47	111	60	124
62	48	43	12	48	60	58	45	57	69	116
63	28	81	19	50	57	28	33	60	69	170
64	8	45	12	41	53	27	9	27	45	101
65	10	47	23	47	51	41	24	29	46	115
66	35	31	22	25	32	47	35	29	59	182
67	73	41	35	55	113	62	28	60	45	274
68	64	60	39	54	53	30	20	72	36	219
69	36	31	12	26	0	51	20	56	31	131
70	2	7	17	16	0	8	11	43	42	103
71	0	1	17	2	1	54	22	76	40	125
72	0	0	11	2	0	72	13	39	33	129
Total	745	788	880	1,131	1,734	2,115	2,348	3,011	3,164	7,905
Population	12,306	9,654	14,226	10,926	24,336	20,526	13,128	16,506	21,480	32,658

and Phrasat are located in the Surin Province. The highest estimates were found in Mueang Chan, Nam Kliang and Ban Mai Chaiyaphot Districts where the percentages exceeded 50%. Mueang Chan and Nam Kliang are located in the Si Sa Ket Province (see the left panel in Fig 5).

Diarrhea incidence

After aggregating the monthly cell counts (including those imputed from the under-reporting model for monthly data), we used the generalized estimating equation (GEE) model with a fixed correlation structure to account for spatial correlations between districts. In this structured correlation matrix, correlations between different districts within a given province were specified as the common mean of the corresponding residual correlation coefficients after fitting the linear model, unless this value was less than 0.1, in which case it was taken to be 0. Correlations between districts in pairs of different provinces were similarly fixed at the means of the corresponding residual correlation, provided these means exceeded 0.1 in magnitude.

Table 3
Estimates of under-reporting percentages by quarter and year.

Factor	Percent
Quarter	
1 : Jan-Mar	8.1
2 : Apr-Jun	7.3
3 : Jul-Sep	9.6
4 : Oct-Dec	21.8
Mean	11.7
Year	
1999	6.4
2000	13.0
2001	15.2
2002	14.4
2003	8.9
2004	7.7
Mean	10.9

Table 4
Means and standard errors of residual correlations between districts within and between provinces.

Province	Chanthaburi	Sa Kaeo	Surin	Buri Ram	Si Sa Ket	Ubon Ratchathani
Chanthaburi	0.17 (0.04)					
Sa Kaeo	0.07 (0.03)	0.13 (0.05)				
Surin	0.03 (0.02)	0.05 (0.02)	0.03 (0.02)			
Buri Ram	-0.06 (0.02)	-0.05 (0.02)	0.03 (0.01)	0.17 (0.03)		
Si Sa Ket	-0.06 (0.02)	-0.02 (0.02)	0.00 (0.02)	0.09 (0.01)	0.12 (0.01)	
Ubon Ratchathani	0.02 (0.02)	0.09 (0.02)	0.03 (0.01)	0.03 (0.01)	0.01 (0.01)	0.12 (0.02)

Table 4 shows the means and standard errors of the residual correlation coefficients between different districts in each province. These correlations were generally quite small, ranging from 0.03 in Surin to 0.17 in Chanthaburi Province.

The results obtained by fitting the GEE model are shown in Fig 3. The left plot shows observed counts versus expected counts. The middle plot shows observed annual incidence per 1,000 versus the corresponding model-fitted values. Since both the cell counts and the corresponding incidence rates were strongly right-skewed, they were plotted on a cube root scale, which gave a squared correlation of 0.65 between the observed and fitted rates on this scale. The right plot shows the residuals (after filtering out the estimated spatial correlations) versus normal quantiles. Apart from a slight tilt, this plot shows little reason to doubt the normality assumption.

Fig 4 shows the fitted annual diarrhea incidence rates based on the GEE model. The graphs show 95% confidence intervals for annual diarrhea incidence/1,000 by quarter (left panel), year (middle panel) and district (right panel), each adjusting for the effects of the other two factors in the model. The dotted horizontal lines on each graph represent the overall mean annual incidence rate (19.4 per 1,000). The seasonal pattern of diarrhea incidence clearly indicates a peak in the January to March quarters with an adjusted annual incidence rate of 29.8 per 1,000 (95% CI 28.6 - 31.0). The trend steadily increased from 17.7 (95% CI 16.5 - 18.9) in 1999 to 20.0 (95% CI 18.9 - 21.0) in 2002, then dropping slightly to 18.6 (95% CI 17.8 - 19.4) in 2003 and increasing to 23.4 (95% CI 22.4 - 24.4) in 2004. The variation between districts was greater, ranging from 3.6 (95% CI 3.0 - 4.4) in Kantharalak District to 58.2 (95% CI 50.8 - 66.7) in Bueng Bun District. The annual diarrhea incidence rate was general higher than the mean in the districts of Buri Ram and Surin Provinces, generally lower than the mean in Chantaburi and Si Sa Ket Provinces, and typical of the whole in Sa Kaeo and Ubon Ratchathani Provinces.

The particular statistical and graphics methods used to produce Fig 5 enable under-reporting and incidence relativities between districts to be clearly illustrated. The right panel shows a thematic map of districts with diarrhea incidence coded according to whether the confidence interval exceeds, crosses, or is below the overall mean. Higher disease incidence occurred mainly in Buri Ram (14 of 23 districts) and Surin (8 of 17 districts) in the middle of the region.

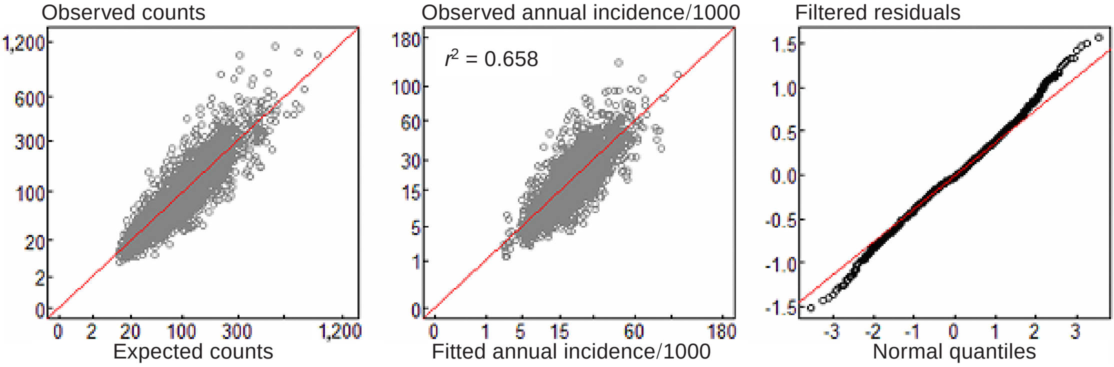
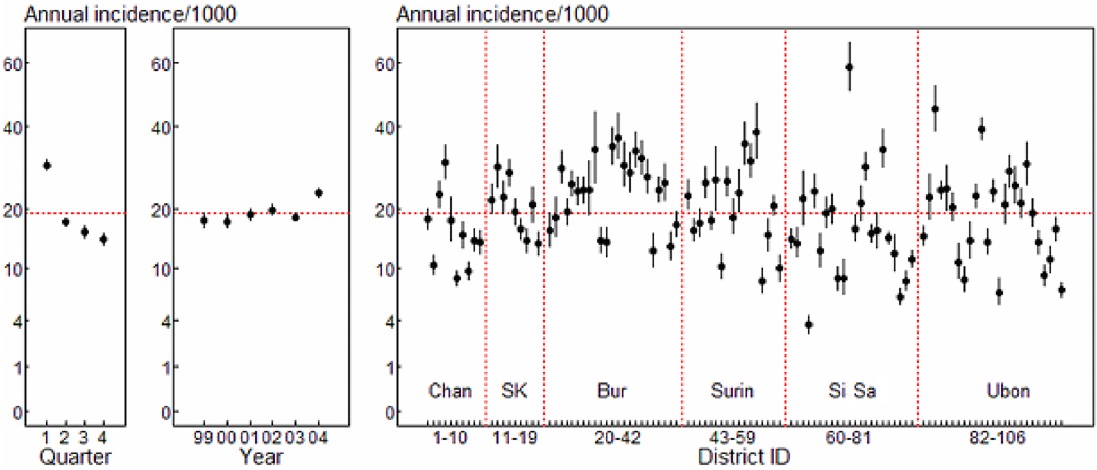


Fig 3–Plots of observed versus fitted counts and incidence rates (left panels) and residuals versus normal quantiles after fitting the GEE model.



Chan, Chanthaburi; SK, Sakaeo; Bur, Buri Ram; Si Sa, Si Sa Ket; Ubon, Ubon Ratchathani

Fig 4–Confidence interval plots of annual incidence rates for each factor (quarter, year and district).

DISCUSSION

The results show the diarrhea incidence in the Thai provinces bordering Cambodia is a serious health problem (Staff of the Department of Planning and Health Information, 2008). The log-linear model was used to impute cell counts for the omitted data and the generalized estimating equation (GEE) model with a fixed correlation structure based on quarter, year and district which were used for analysis. The use of the GEE method for modeling spatial correlation is discussed in detail in a recent review by Dormann *et al* (2007). Generalized linear models (GLMs) provide powerful statistical modeling (Aitkin *et al*, 1989) and the application of the GLMs to model epidemiological data was recommended by Flanders and Kleinbaum (1995). This method has also been applied to modeling diarrhea diseases by Kale and Hinde (2004), HIV/AIDS and other infectious disease mortality rates by Lim and Choonprabub (2007).

The estimated under-reporting of diarrhea case (Table 3) was relatively high dur-

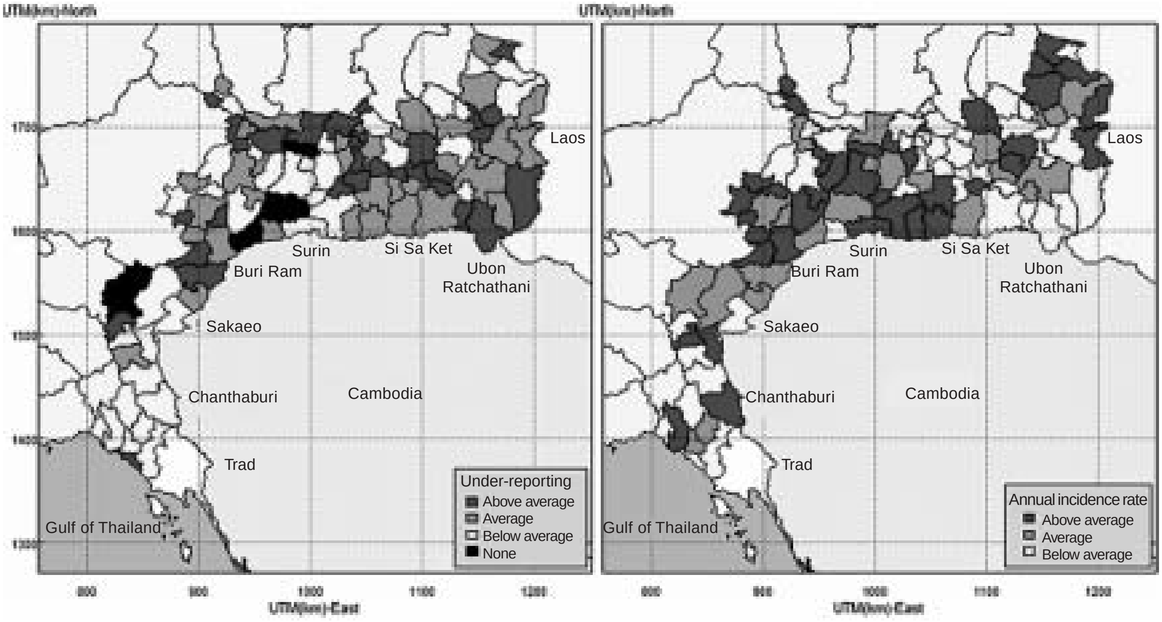


Fig 5–Schematic map of under-reporting (left panel) and annual diarrheal incidence rates (right panel) in districts of Thailand bordering Cambodia.

ing the last quarter and low during the second quarter. One reason for this pattern might have been due to the low number of actual cases during the last quarter of each year, but may also have been due to health worker failure to record and report actual cases for various reasons, including computer system problems. During the period 1999 to 2004, estimated under-reporting was lowest in 1999 and steadily increased in 2001 and then decreased by 2004. The highest estimates were found in the three districts Mueang Chan, Nam Kliang and Ban Mai Chai Phot, where our estimates exceed 50%. Evidence of under reporting was also found in most districts of Si Sa Ket Province. Possible causes include overworked government personnel, inadequate coordination with private hospitals, and inadequate supervision at the central public health level (Saengwonloey *et al*, 2003). However, there were four districts with no evidence of under-reporting.

The highest diarrhea incidence occurred among children under five years old, which is consistent with the findings by Pinfold *et al* (1991) who conducted a study of seasonal effects on the reported incidence of acute diarrhea in Northeast Thailand. They found young children were the most affected group, with a reported annual incidence of 2,952 per 100,000 for children less than five years old. Kosek *et al* (2003) found that diarrhea accounted for 21% of all deaths of children less than five years old. Early intervention for this age group could reduce the incidence.

According to this study, the diarrhea incidence during the study period was relatively high from January to March. This period mainly overlaps the winter and summer seasons and is associated with a high risk of diarrhea (Pinfold *et al*, 1991). Diarrhea disease was influenced by El Niño (Patz, 2005) in Peru, the number of diarrhea cases increased with an ambient temperature increase (Checkley, 2000). Moreover, we discovered the overall trend of diarrhea incidence from 1999 to 2004 was slightly higher. In contrast, the Bureau of Policy and Strategy has shown that the diarrhea trend among the overall Thai population remained stable based on statistics from the Ministry of Public Health (2001). The high, and increasing trends of diarrhea incidence occurred in the districts in the middle of the region. Possible reasons are as follows: residents in this area have a relatively low Gross Regional and Provincial Product (GRP) per capita (Office of the National Economic and Social Development, 2006); residents are more likely to consume raw meats and fermented foods (Lee *et al*, 1993; Somnasang *et al*, 1998) and long droughts occur in the region.

In conclusion, the data analysis model enables adjustments to be made to compensate for under-reporting and should thus provide more accurate estimates of disease incidence. The model may be useful for health planning in countries similar to Thailand where routine epidemiological reports of diarrhea and other disease cases are provided at the district level, because it provides a simple method based on readily available demographic data. The model can also be used to identify an unusually high annual incidence within the period of its occurrence, and thus enable health authorities to reduce the severity of ensuing epidemics by implementing preventative measures put in place for the demographic group at risk.

ACKNOWLEDGEMENTS

This study was funded by the Graduate School, Prince of Songkla University. We would like to thank the Ministry of Public Health for giving permission to use the data. We are grateful for Prof Don McNeil and Dr Vorasith Sornsrivichai for their helpful advice and suggestions. We also thank Greig Rundle and Dr Shanley Chong who corrected the English for our paper.

REFERENCES

Aitkin M, Anderson D, Francis B, Hinde J. Statistical modelling in GLIM. Oxford, England: Claredon Press, 1989.

Ardkeaw J, Tongkumchum P. Statistical modelling of childhood diarrhea in the north-eastern border region of Thailand. *Southeast Asian J Trop Med Public Health* 2009; 40: 807-15.

Brownlie J, Peckham C, Waage J, *et al*. Foresight. Infectious diseases: preparing for the future. Future threats. London: Office of Science and Innovation, 2006.

Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health. Thailand: Annual epidemiology surveillance report 2001. Bangkok: War Veterans Organization of Thailand, 2002.

Bureau of Epidemiology, Department of Disease Control, Ministry of Public Health, Thailand: Annual epidemiology surveillance report 2006. Bangkok: War Veterans Organization of Thailand, 2007.

Carlos CC, Saniel CM. Etiology and epidemiology of diarrhea. *Phil J Microbiol Infect Dis* 1990; 19: 51-3.

Checkley W. Effect of El Niño and ambient temperature on hospital admissions for diarrhoeal disease in Peruvian children. *Lancet* 2000; 355: 442-50.

Dormann CF, McPherson JM, Arauújo MB, *et al*. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography* 2007; 30: 609-28.

Flanders WD , Kleinbaum DG. Basic models for disease occurrence in epidemiology. *Int J Epidemiol* 1995; 24: 1-7.

Hosmer DW, Lemeshow S. Applied logistic regression. 2nd ed. New York: John Wiley and Sons, 2000.

Intusoma U, Sornsrivichai V, Jiraphongsa C, Varavithaya W. Epidemiology, clinical presentations and burden of rotavirus diarrhea in children under five seen at Ramathibodi Hospital, Thailand. *J Med Assoc Thail* 2008; 91: 1350-5.

Jiraphongsa C, Bresee JS, Pongsuwanna Y, *et al*. Epidemiology and burden of rotavirus diarrhea in Thailand: results of sentinel surveillance. *J Infect Dis* 2005; 192: S87-93.

Kale PL, Hinde JP. Modeling diarrhea disease in children less than 5 years old. *Ann Epidemiol* 2004; 14: 371-77.

Kleinbaum DG, Klein M. Logistic regression - A self learning text. 2nd ed. New York: Springer-Verlag, 2002.

Konchom S, Singhasivanon P, Kaewkungwal J, *et al*. Trend of malaria incidence in highly endemic provinces along the Thai borders, 1991-2001. *Southeast Asian J Trop Med Public Health* 2003; 34: 486-94.

Kosek M, Bern C, Guerrant RL. The magnitude of the global burden of diarrhoeal disease from studies published 1992-2000. *Bull World Health Organ* 2003; 81: 197-204.

Lee HC, Steinkraus HK, Reilly PJ. Fish fermentation technology. United Nations University Press, 1993: 155-66.

Liang KY , Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; 73: 13-22.

Lim A, Choonpradub C. A Statistical method for forecasting demographic time series counts, with application to HIV/AIDS and other infectious disease mortality in southern Thailand. *Southeast Asian J Trop Med Public Health* 2007; 38: 1029-40.

Lumbiganon P, Panamonta M, Laopaiboon M, Pothinam S, Patithat N. Why are Thai official perinatal and infant mortality rates so low?. *Int J Epidemiol* 1900; 19: 997-1000.

Ministry of Public Health. Public health statistic. Nonthaburi: Ministry of Public Health, 2001.

Office of the National Economic and Social Development. Gross regional and provincial product. Bangkok: The Office of the National Economic and Social Development, 2006.

Parashar UD, Bresee SJ, Glass RI. The global burden of diarrhoeal disease in children. *Bull World Health Organ* 2003; 81: 236.

Patz JA, Lendrum DC, Holloway T, Foley JA. Impact of regional climate change on human health. *Nature* 2005; 438: 310-17.

Pinfold JV, Horan NJ, Mara DD. Seasonal effects on the reported incidence of acute diarrhea disease in northeast Thailand. *Int J Epidemiol* 1991; 20: 777-86.

R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2008. (Cited 2008 Jun 28). Available from: URL: http://www.R-project.org

Saengwonloey O, Jiraphongsa C, Foy H. Thailand report: HIV/AIDS surveillance 1998. *J Acquir Immune Defic Syndr* 2003; 32: S63-7.

Somnasang P, Moreno G, Chusil K. Indigenous knowledge of wild food hunting and gathering in north-east Thailand. *Food Nutri Bull* 1998; 19: 359-65.

Staff of the Department of Planning and Health Information, Ministry of Health, the Reproductive Health Association of Cambodia, and PRB, No date. (Cited 2008 Dec 10). Available from: http://www.prb.org/Articles/2002/ChildreninCambodiaFaceHighMortalityRate.aspx

Thai Working Group on burden of disease and injuries. Burden of disease and injuries in Thailand: priority setting for policy. Nonthaburi: Ministry of Public Health, 2002.

Thimasarn K, Jatapadma S, Vijaykadga S, Sirichaisinthop J, Wongsrichanalai C. Epidemiology of malaria in Thailand. *J Travel Med* 1995; 2: 59-65.

Vargas M, Gascon J, Casals C, *et al*. Etiology of diarrhea in children less than five years of age in Ifakara, Tanzania. *Am J Trop Med Hyg* 2004; 70: 536-9.

Venables WN, Ripley BD. Modern applied statistics with S. 4th ed. New York: Springer-Verlag, 2002: 144-51.

Yan J, Fine J. Estimating equations for association structures. *Stat Med* 2004; 23: 859-80.