MINISTRY OF EDUCATION AND TRAINING

MINISTRY OF HEALTH

NATIONAL INSTITUTE OF HYGIENE & EPIDEMIOLOGY

----------*----------

PHAN DANG THAN

EPIDEMIOLOGICAL CHARACTERISTIC AND APPLICATION OF MATHEMATICAL MODELS TO PREDICT DIARRHEA IN HATINH PROVINCE

Major: EPIDEMIOLOGY

Code: 62 72 03 01

SUMMARY OF PhD THESIS ON MEDICINE

HANOI – 2020

THESIS PERFORMED AND COMPLETED AT THE
NATIONAL INSTITUTE FOR HYGIENE & EPIDEMIOLOGY

Scientific supervisors:

1. **Assoc.Prof. Nguyen Ngoc Sang**
2. **Assoc.Prof. Le Thi Phuong Mai**

Counter arguer 1: ……..

Counter arguer 2: …….

Counter arguer 3: ……..

This doctoral thesis will be defended at the Dissertation Committee of Institutional level at:

National Institute of Hygiene & Epidemiology

on 2020

This doctoral thesis can be found at:

1. The National Library
2. The Library of National Institute of Hygiene and Epidemiology

INTRODUCTION

In 2017, 1.6 million people died from diarrhea globally, of which 1.7 million children died from diarrhea in the last three decades among the group of under five years old. In Southeast Asia, the global burden of diarrhea ranks second place after the Shahara region.

In Vietnam, the trend of diarrhea in recent years has not decreased, with 8.6% of children under 5 years old having diarrhea in the community within two weeks before the survey, 3% higher than this rate of acute respiratory infections in the same period. Early warning of diarrhea and endemic diarrhea in Vietnam is not proactive and has not used regular monitoring system data and weather data to build forecasting models for the early warning system. Ha Tinh is a province with a harsh climate and high endemic diarrhea with an averaging 8,600 cases of diarrhea per year in the period 2000 - 2010, cross-sectional surveys in the areas frequently flooded showed this rate of 14.2%.

The revealed question was what are the epidemiological characteristics of diarrhea in the population? Is the number of diarrhea cases related to weather factors? And can we apply a mathematical model to diarrheal data and weather factors for the short-term prediction of this disease? To answer the above questions, we conducted a research on "Epidemiological characteristics and application of mathematical models to predict diarrhea in Ha Tinh province" with the following objectives:

1. To describe the epidemiological characteristics of diarrhea in a commune of a drought-affected area in Ha Tinh province in 2014 – 2015.

2. To analyze the relationship between some weather factors and diarrhea in Ha Tinh during the period of 1992 - 2015.

3. To apply mathematical models based on climatic factors to predict diarrhea in Ha Tinh

New contributions of the thesis

The thesis has described the epidemiological characteristics by the new incidence rate per person - time unit, cumulative rate (%) per month/year, an average number of cases/year, recurrence rate per person – time unit. The thesis has identified new incidence rate of 40% person/year for diarrhea among the population in the drought area of Ha Tinh province (3.3% person/month). The cumulative/monthly prevalence of diarrhea was 6.3% and 13.1% among children under 5 years old. The average number of diarrhea was 0.78 times/person/year, this rate for the group of children under 5 years old was 1.52 times/child/year. Children under 5 years old showed to have higher new incidence rate, higher average number of diarrhea/year, higher number of days with diarrhea, higher cumulative rate per month and higher recurrence rate than these of other age groups.

Result of analyzing the relationship between some climate factors and diarrhea showed that average temperature, average maximum temparature, average minimum temperature, absolute humidity, total sunny hours of 1 - 2 months prior were correlated positively with the number of

current cases/monthly. While the increase in total rainfall and sea surface temperature index at NINO3 area in 2 - 3 months prior was correlated to the increase in diarrhea case number currently.

A short-term predictive model of diarrhea has been developed using routine monitoring data on diarrhea and weather data of 24 years from 1992 to 2015 in Ha Tinh. The SARIMA-X model $(1,1,1)\ (0,1,1)_{12,}$ where X is the mean maximum temperature factor, was considered as a suitable model for the forecast. The simulation forecast results showed that the prediction error before 1 month was 6.1% and before 12 months was 9.7% compared with the observed data in 2016.

The structure of thesis

Thesis consisted of 140 pages, of which, 2 pages of the Introduction; 37 pages of Literature overview (Chapter 1); 19 pages of Subjects and Research methods (Chapter 2); 51 pages of Research Results (Chapter 3); 36 pages of Discussion (Chapter 4); 2 pages of Conclusion and 1 page of Recommendation. The thesis included 40 tables, 5 figures, 35 charts and 152 references (14 Vietnamese documents, 138 English documents).

Chapter 1. LITERATURE OVERVIEW

1.1. Diarrhea and its causative

Diarrhea is the second leading cause of death among children under 5 years old. Every year, diarrhea caused death for 1.5 million children worldwide.

In Vietnam, according to research data on gastrointestinal disease trends for the 2000-2010 period, diarrhea tended to decrease slightly with the incidence of diarrhea decreased from 12,369 cases per 100,000 population in 2000 to 9,588 cases/100,000 people in 2010. The mortality rate also decreased significantly from 0.1 cases/100,000 people in 2001 to 0.01 cases/100,000 people in 2010. The incidence rate was found high in May - July (104 - 111 cases/100,000 people) and lowest in January (84 cases/100,000 people) annually. Each year, on average, in Ha Tinh, there was about 8,600 diarrhea cases/100,000 people in the period 2000 - 2010, of which, during 2014 - 2015, on average, each month there were 731 cases/100/000 population, equivalent to 0.731% of cases having diarrhea by month.

1.2. The factors related to diarrhea

There are many factors that can have a direct or indirect effect on diarrhea, that it can be divided into several groups including: socio-economic status; infrastructure, living and sanitary conditions; nutrition; hygienic behavior; infection with microorganisms that cause intestinal diseases.

1.3. Relation between weather factors and diarrhea

According to WHO estimates in 2000, changes in climate and weather factors were responsible for nearly 2.4% of diarrhea among middle-income countries worldwide. Studies have shown the correlation between weather factors and fertility, survival rate of bacteria and food-borne microorganisms that cause diarrhea. Weather factors, especially extreme weather, can have

an indirect effect on behavioral aspects such as increased water use and poor sanitation, thus increasing the likelihood of disease transmission.

1.4. Predictive mathematical model and time series model predict diarrhea based on weather factors

Accurate and timely prediction for infectious disease can support the measures to develop proactively plan for infectious disease control. Currently, 6 types of models are applied to predict infectious diseases:

- Agent-based models
- Compartmental models
- Ensemble models
- Metapopulation models
- Time series models

The time series model uses data collected, stored and observed with time increments, in which, the time series usually denoted as (T) and Xt random variables derived from a probability distribution where t is natural numbers.

The Autoregressive Integrated Moving Average model (ARIMA model) is the method of predicting research factors independently (forecasts by time series) by algorithms using latency, giving the appropriate prediction model.

The typical linear time series models such as moving average process (MA), autoregression processes (AR), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) model can be applied in disease analysis and prediction. If the data series having seasonal character, it needs to include the seasonal factor in the model, then the model is written as SARIMA - Seasonal Autoregressive Integrated Moving Average)

Predictive models are used in the communicable disease early warning system to monitor disease and risk factors for disease control.

Chapter 2. STUDY SUBJECTS AND METHODS

2.1. For Objective No. 1

To describe the epidemiological characteristics of diarrhea in a commune of drought-affected area in Ha Tinh province in 2014 – 2015

2.1.1. Study subjects

Predictive models are used in the early warning system of communicable diseases to monitor disease and risk factors for disease control.

2.1.2. Study time and place

Time: From 1/07/2014 till 30/06/2015.

Place: One commune of a drought-affected area in Ha Tinh province

2.1.3. Study design

The cohort research method was used.

2.1.4. Sample size and sampling method

Entire sample size was applied.

2.1.5. Methods and tools for data collection

Method: Cases of diarrhea were determined by self-reporting.

Tools and methods of data collection:

+ Tools: use "Diarrhea Diary"

+ Method: Self-recorded daily under the supervision of investigators.

2.1.6. Research variables

- Total number of people - year and person - month of recorded diarrhea cases.
- Incidence rate/100 people - year and 100 people - month
- Percentage of new accumulated diarrhea cases by month and week
- Number of diarrhea/person/year
- Rate of recurrence/100 people - year and 100 people – month

2.1.7. Data analysis

- Incidence rate and confidence interval using the survival analysis method for future cohort studies
- Analysis of average diarrhea episodes and days and its standard deviation

2.2. For Ojective No. 2

To analyze the relation between some weather factors and diarrhea ta Hatinh province during the period of 1992 - 2015

2.2.1. Study subjects

- Reports, records on diarrhea cases/100.000 population/month and weather data of 24 years of Hatinh during 1992– 2015.

2.2.2. Methods

Time series analysis was used.

2.2.3. Data analysis

Poisson regression analysis was used and a delayed effect model was setting up with diarrhea data series used as the dependent variables and the weather factors controlled with seasonal character and trends as independent variables and trends by the Flexible spline functions using Stata 14.2 software.

2.2.4. Variables

Variables of disease (Dependent variables)

- Number of diarrheal cases/month/100,000 population during 24 years.

Variables of weathers (Independent variables)

- Temperature: including mean temperature/month (T_{mean}), average minimal temperature/month (Tmin) and average maximal temperature/month (Tmax)
- Total rainfall by months (RR)
- Average absolute humidity/month (AH)
- Total sunny hours/month (SH)
- Anomalous sea surface temperature (SAT) at 4 areas: NINO1+2, NINO), NINO4 and NINO 3.4.

2.3. For Objective No. 3

To apply mathematical models based on weather factors to predict diarrhea in Ha Tinh

2.3.1. Study subjects

- Number of diarrhea cases/100.000 population/month during past 24 years.
- Temperature, total rainfall, humidity, number of sunny hours by month in 24 years
- Number of diarrhea cases/100,000 people/month in 12 months of 2016

2.3.2. Methods

Application of ARIMA modeling method with controlled seasonal factors, trends and prediction factors that are weather factors that have a linear relationship with diarrhea

2.3.3. Steps to build SARIMA-X model and forecast

Building and forecasting by SARIMA model

- Step 1: Identify the SARIMA model
- Step 2: Select forecasting model
- Step 3: Test the model
- Step 4: Predict and evaluate the model

2.3.4. Research index groups

- Number of cases of diarrhea/100.000 population/month and weather data/month
- Number of cases of diarrhea monthly/100.000 population in 2016.

2.4. Ethics in research

This study was conducted after being approved by scientific committee and ethics committee of the National Institute of Hygiene and Epidemiology.

Chapter 3. RESULTS

3.1. The epidemiological characteristics of diarrhea in a commune of drought-affected area in Ha Tinh province in 2014 – 2015

A total of 2642.1 persons – year, equivalent to 31655.8 person – month has been followed up for diarrhea among 2961 cases participated the study, of which female was rated 54.3% and male was of 45.7%. The group under 5 years old composed of at least 293 children, accounting

for 9.9%, the group aged from 25-40 years old were accounted for the most (24.1%). The age group from 45 to 60 years old and the group over 60 years old were rated 18.7% and 15.3%, respectively.

Table 3.1 Diarrheal incidence rate/100 people-year by sex and by age group (n = 2642,1)

Characteristics	N (people - year)	Number of cases	Rate %	Confidence interval (95% CI)	
Total	2642.1	1058	40.0	37.7	42.5
Sex					
Female	1,452.0	585	40.3	37.2	43.7
Male	1,190.1	473	39.7	36.3	43.5
Age group					
Under 5 y.o.	246.5	150	60.9	51.9	71.4
From 5 to 15 y.o.	473.6	227	47.9	42.1	54.6
From 15 to 25 y.o.	347.8	125	35.9	30.2	42.8
From 25 – 45 y.o.	637.6	228	35.8	31.4	40.7
From 45 – 60 y.o.	512.1	191	37.3	32.4	43.0
Above 60 y.o.	424.7	137	32.3	27.3	38.1

The incidence rate of diarrhea in the community was 40% people - year (95% CI: 37.3 - 42.5) with person unit - month was 3.3% (95% CI: 3.1 - 3.6). The diarrhea incidence rate was the highest among group of subjects aged under 5 years old (60.9% person-year, equivalent to 5.1% of person-month) and had a risk of being 1.6 times higher than that of the other community age group ($p < 0.001$).

Table 3.2 Average number of diarrhea in a community per follow-up year by sex and by age group (n = 2961)

Characteristics	Average	SD	Confidence interval (95% CI)		p- value
Total (n=2961)	0.78	1.4	0.73	0.83	
Sex					
Female (n= 1607)	0.77	1.33	0.71	0.84	0.80
Male (n=1354)	0.78	1.49	0.70	0.86	
Age group					
< 5 y.o. (n=293)	1.52	2.29	1.25	1.78	<0.001
From 5 y.o. (n=2668)	0.70	1.24	0.65	0.74	

The average number of diarrhea in the community was 0.78 times/person/year. This rate among the group under 5 years old was statistically significant higher (1.52 times/child/year, p<0.001). The average total days of diarrhea was 5.41 days/patient/year and this was higher among the group under 5 years old with 7.7 days/patient/year.

The rate of diarrhea recurrence in the community was 22.4% people - year (95% CI: 20.7 - 24.3). This rate was the highest in group aged under 5 years old [38.5% person - year (95% CI: 31.5 - 47.1)] and was 28.5% person - year (95% CI: 24.1 - 33.7) in the group from 5 to 15 years old.

During the period of recording diarrheal cases, an average of 6.3% of people in the community were infected each month, increased in the period from May and reached the highest rates of 7.7% and 7.1%, respectively, in June and July.

3.2. The relationship between some weather factors and diarrhea in Ha Tinh during the period of 1992 - 2015

The average total number of diarrhea cases was 827.5 cases/100,000 people, in the highest month it was 2,715.4 cases/100,000 people. The average temperature in the period 1992 - 2015 was 24.4^{0}C with the average highest temperature of the month of 36.7^{0}C and the average lowest of 9.3^{0}C. The total average monthly rainfall was 207.1mm, of which the highest average month rainfall was 1859mm and the lowest average month rainfall was only 3mm. The average absolute humidity was 25.8%. The lowest sea surface temperature of NINO1 + 2 region was 23.2^{0}C and the highest was 28.6^{0}C found at NINO4 region. During this period, climatic factors were determined seasonal, but the trend was not clear shown.

Table 3.3. The relation between average temperature and diarrhea cases at different time latency (n = 288)

Weather factors	**IRR**	**z**	**P>z**	**Confidence interval** (95% CI)	
T_{mean} currently	1.027	10.01	<0.001	1.022	1.032
T_{mean} 1 month late	1.024	8.990	<0.001	1.019	1.029
T_{mean} 2 months late	1.013	4.710	<0.001	1.007	1.019
T_{mean} 3 months late	1.000	0.35	0.729	0.995	1.001

Temperature factor 1 - 2 months prior expressed positive relationship with the number of diarrhea cases of the same time. Increase Tmean for 1^{0}C possed the risk of increasing the number

of cases at the same time from 2.4% to 2.7%. Similar to the factors Tmax and Tmin increased 1^0C caused the increasing of diarrhea cases from 2% to less than 3% with p <0.001.

Table 3.4. The relation between the total rainfall and the number of diarrhea cases by month, in 1992 – 2015 at Hà Tĩnh (n = 288)

Weather factors	**IRR**	**z**	**P>z**	**Confidence interval** (95% CI)	
RR currently	0.997	-0.810	0.418	0.989	1.005
RR of 1 month late	0.984	-2.940	0.003	0.973	0.995
RR of 2 months late	1.005	0.940	0.350	0.994	1.021
RR of 3 months late	1.171	5.360	0.000	1.192	1.231

Total rainfall of 2-3 months prior was found propotional related to the diarrhea, if the total rainfall increased 10 mm, the risk for diarrhea also increased by 17.1% of cases with p<0.0001 and confidence intervals of between 1,192 - 1.23.

The absolute humidity increased by 1 g/m3 in the previous month was found ralated to the the risk of increasing the diarrhea in the current month from 2.8% to 3.3% with p <0.001.

The sea surface temperature index at the NINO3 region in the 3-month delay step had the strongest correlation with the risk of an increase in the number of diarrhea cases up to 13.8% if this index increased by 1^0C (p<0.001).

3.3. Applying the ARIMA models based on climatic factors to predict short term diarrhea in Ha Tinh

3.3.1. Building and forecasting the SARIMA model

The Augmented Dickey-Fuller test of diarrhea data series in Ha Tinh in the period 1992 to 2015 with p = 0.875 showed the unstationary of the data series. Three models are ARIMA (0,1,1), (1,1,0) and (1,1,1) satisfying the condition that the AR and MA function has p<0.05 in both values, the tissues Figure.

Three models of ARIMA (0,1,1), (1,1,0) and (1,1,1) were found satisfied the condition of the AR and MA function with p <0.05 for both values and for all 3 models.

Table 3.5. ARIMA model controlled by seasonal factor (n=287)

Model	**Index**	**SARIMA model**		
		$(0,1,1)(0,1,1)_{12}$	**$(1,1,0)(0,1,1)_{12}$**	**$(1,1,1)(0,1,1)_{12}$**
ARMA	L1.ar	-	-0.41**	0.22*

	L1.ma	-0.64**	-	-0.78**
$ARMA_{12}$	L1.ma	-0.82**	-0.85*	-0.83**
AIC		3892.8	3916.2	3889.4
BIC		3907.3	3930.8	3907.7

*** p<0.01 * p<0.05*

Seasonality and trend control showed all three models of SARIMA model (0,1,1) (0,1,1) 12, (1,1,0) (0,1,1) 12, (1,1,1) (0,1,1) 12 statistically significant with the function AR, MA, MAR, MMA with p <0.05.

The residual checking for the white noise calculation of the SARIMA model (0,1,1) $(0,1,1)_{12}$ by the Portmanteau (Q) test showed a value of 20.5 and p = 0.045. Meanwhile, the SARIMA model (1,1,0) $(0,1,1)_{12}$ with the Portmanteau test (Q) gave the value of 23.1 with p = 0.026 and the SARIMA model (1,1,1). $(0,1,1)_{12}$ with the Portmanteau (Q) test gave the value of 13.5 and p = 0.31. Thus, the model SARIMA (1,1,1) $(0,1,1)_{12}$ has the smallest AIC = 3889.4 and BIC = 3907.7, so was the most suitable for forecasting diarrhea in Ha Tinh.

The model's forecast data showed the overall short-term forecast error of 7.6% before 1 month, thus could reach to the highest value (15.7%) in January 2016 and minimized to the lowest value (2.5%) in October 2016. Meanwhile, the medium-term forecast (before 12 months) showed to have an overall error of 10.1%, with the maximal error of 15.7% in January and minimal error of 2.9% in April.

3.3.2. Building SARIMA-X model with predict variables of weather factors

Linear regression analysis using the least squares method showed a linear correlation between diarrheal disease data series and the temperature, humidity and sunshine hour factors at Ha Tinh province with p <0.05.

Table 3.37. SARIMA (1,1,1)$(0,1,1)_{12}$ model with predict factors of temperature, humidity and total sunny hours at Ha Tinh province (n= 288)

Model	**Index**	**SARIMA (1,1,1)$(0,1,1)_{12}$**				
		Tmean	Tmax	Tmin	AH	SH
ARMA	L1.ar	0.22**	0.22**	0.22**	0.22**	0.22**
	L1.ma	-0.78*	-0.78*	-0.78*	-0.78*	-0.78*
$ARMA_{12}$	L1.ma	-8.3**	-8.2**	-8.2**	-8.2**	-8.2**

DS12.Tmean	16.59	-	-	-	-
DS12.Tmax	-	18.37*	-	-	-
DS12.Tmin	-	-	9.34	-	-
DS12.AH	-	-	-	11.0	-
DS12.SH	-	-	-	-	0.72
AIC	3887.4	3735.0	3889.5	3888.7	3889.6
BIC	3909.4	3756.7	3911.4	3910.7	3910.3

*** p<0.01 * p<0.05*

The model was built with the factors Tmean, Tmax, Tmin, AH and SH with seasonal character of a 12-month cycle showed only SARIMA (1,1,1) $(0,1,1)_{12}$ with average maximum temperature factor having the AR, MA and MMA functions with p <0.05. This model also has AIC value of 3735.0 and BIC value of 3756.7, smaller than the model without participation of weather factors.

The residuals of the model showed that the ACF and PACF functions of the model residuals were in the range and have p>0.05 value. Bartlett's test showed a value (B) of 0.81 with p = 0.52; Portmanteau test (Q) had a value of 14.8 with p = 0.22. Thus, the SARIMA model (0,1,1) $(0.1,1)_{12}$ with average maximum temperature has white noise and AIC and BIC of the smallest values, therefore is the most suitable model for predicting diarrhea in Ha Tinh province.

The error for the short-term forecast by this model was 6.1%, the month with the least error of 0.8% was October 2016 and the month with the most error (14.9%) was September 2016. Meanwhile, the error for the 12-month forecast was 9.7% and the month with the most error (16.0%) was April 2016 and month with the least error (1.9%) was January 2016.

Chapter 4. DISCUSSION

4.1. The epidemiological characteristics of diarrhea in a commune of drought-affected area in Ha Tinh province in 2014 – 2015

Our study using cohort follow-up method and using indicators such as incidence rate, average number of cases, average number of days infected, and recurrence rate to describe the overall burden of diarrhea in the residential community. Used the survival analysis method and

based on the denominator of the total number of people - follow-up time would ensure accurate sample size in the study condition that some subjects could be absent or were not followed for enough time or gave up for some reasons. It is also the method showing the speed of generating a health problem, especially infectious diseases, because of time and place.

This study showed that the incidence of diarrhea in Hatinh province was 40% per year, equivalent to 3.3% of people - month compared with the vertical follow-up study done by Katherine L. Anders et al in the South of Vietnam. Male infants born up to 12 months old had an incidence rate of 27.1% - year, however, because these infants were under 12 months of complementary feeding period, the incidence of diarrhea was lower. Meanwhile, according to data reported on statistics of infectious diseases in Ha Tinh in the period 2014 - 2015, the monthly diarrhea rate was 0.731% (731/100,000 people/month), compared with the cumulative incidence rate in the study community of 6.3% monthly. As the data recorded in the surveillance system was almost 10 times lower than the actual incidence in the community, this showed that a large proportion of the population in the community got diarrhea but was not statistically recorded in the infectious disease reporting system. The reason might be that they self-treated or did not need any treatment or the monitoring system was not sensitive enough to cover all cases. The 2014 Vietnam National Child and Women Targeted Survey Report also found that only 55.1% of children under 5 years old went to health facilities or received advice from health workers and 12.6% treated with zinc and oresol tablets. Diarrhea is a disease that can recur, in this study we have considered a new episode of diarrhea as that ending at least 2 days before the appearance of next diarrhea.

For children under 5 years old, the 2014 survey and assessment of the child and women's goals (MICS) nationwide using suitable sample size, survey methods, data processing and analysis showed that the rate of diarrhea among children under 5 years old in the North Central region was 7.7% and that of the whole country was 8.6%. Compared with results of our studies, Ha Tinh province has a higher rate of diarrhea among children under 5 years old (13.1%) than that in other regions of Vietnam. Our study results showed that the average number of diarrhea/ year/subjects was 0.78 and among the children under 5 years old this was 1.52 times. Thus, this data is equivalent to result of a multisite study of 195 countries showing that Southeast Asia has a diarrhea rate of 0.72/person/year, of which, among children under 5 years old this was 1.6/person/year and among people aged over 70 years old this was 1.28 times/person/year. Meanwhile, the highest area was sub-Saharan which has the rate of 1.21 times/year and 2.28 times/child/years among children under 5 years old.

For other ages, compared with other cohort studies, Pham Duc Phuc et al. had conducted a 1 year follow-up study on adults, the obtained results showed only 0.28 diarrhea

episodes/person/year. The age group of 16 to 35 years old was found with 25.2 to 28.4% person – year while among the age over 55 years old this rate was 39.6% person - year. Meanwhile, the other 18-month community follow-up study showed that the incidence rate of new diarrhea of rural areas of Hanoi was 28% person - year for the age group of 15-70 years old. Thus, for the group over 5 years old, the incidence rate in different regions was found different, but not much. Especially the incidence rate of the group aged over 70 years old, not high compared to the results of studies conducted in West Asia, India, and Pakistan region, might be because that these regions have frequent endemic diseases related to bacteria such as cholera, dysentery, typhoid.

When compared with the results of the study on diarrhea in community at delta region, results obtained by the cross-sectional investigation method showed that the highest diarrhea incidence month was June 2015 (7.7%). The results of a cross-section survey study on diarrhea conducted by the National Institute of Hygiene and Epidemiology in 2012 showed the incidence rate of diarrhea in Ha Tinh of 14.2% in the flooded area and 11.6% in the non-flood area. Thus, the method of self-identified diarrhea could overcome the error of the cross-sectional survey and gave more accurate data.

Our study showed that the summer months of from June to August had a higher incidence of diarrhea than other months. This is also consistent with studies carried out in the North Central region and nationwide that indicated more cases of diarrhea during the summer.

For the average diarrhea episodes, according to a systematic review of diarrhea incidence rates in low-income and middle-income countries, on average, each child under 5 years old had 2.1 diarrhea episodes. Our research also showed that children under 5 years old have 1.5 times/child/year, on average, lower than result of one overview study in Southeast Asia (2.4 times/year (1.5 - 3.3). Our study showed that 32.4% of children under 5 years old suffered from 2 or more diarrhea episodes/year, this age group was also the age that having the highest incidence and recurrence. Results obtained by our study were consistent with the results of a national survey of the General Statistics Office and UNICEF conducted in 2014 with the 2-week prevalence rate of 8.6%. Since these studies were conducted on inpatients of clinics, hospitals or data supplied by health care workers and the subjects were only certain selected populations, others may be omitted. On the other hand, among children under 5 years old, those aged of 12 - 24 months old showed to have the highest rate of diarrhea because they’re not matured yet in immunity. On the other hand, children are more exposed to pathogens in nature environment and home environment.

The October 2014 cross-sectional survey in one area of Shanghai city showed the diarrhea incidence rate per month of 4.1% with an average episode number of 0.5 times/person/year and among them, only 21.2% received the advices or treatment from medical staff.

In this study, the incidence/month is the number of new diarrhea cases occurred in the month of the study population, so this rate could be higher than the monthly average of the incidence/year. Our study showed the incidence rate/month of about 6%, depending on the season. From May to July and September to November, the incidence rate was about 7% and that of the first quarter was 5%. Thus, the incidence rate/month was also consistent with seasonal studies in the North Central region. In the summer months or months with heavy rain, the prevalence of diarrhea increased in the community.

4.2. The relationship between some weather factors and diarrhea in Ha Tinh during the period of 1992 - 2015

Regarding the cycle of disease, it was not clear. The cumulative incidence rate by month was found highest in May, June, July and in October in the year. The findings were consistent with the data reported regularly throughout the year. In the period of 2003 – 2013, diarrhea in Ha Tinh province was found with higher prevalence in June and July (67.2 - 67.7 episodes/100,000 people) and the lowest was determined in January (45.3 cases/100,000 inhabitants) annually. The incidence was high because after the dry season, the rains increased transmission of disease causatives, contaminated water, food and this is the factor that increases the risk of diarrhea. This has also been found by Carlton et al. when studied the diarrhea in the community by weeks and by heavy rain phenomenon. However, the seasonal trend of this is not clear, in consistent with the found trend of diarrhea in Ha Tinh, where climatic factors such as average temperature, average rainfall, tend to increase and decrease not clearly according to the months of the year.

Concerning temperature factors, both univariate and multivariate analysis showed that temperature positively correlated with the number of diarrhea cases. The results of this study were consistent with the results obtained by a study of Hall et al. in 2011 using self-reported 1-year diarrhea data, that showed that with a 1^0C increase, the number of diarrhea cases increased by 2.48%. Bennett et al. in 2012 conducted a study based on community diarrhea data, the obtained results showed higher rates of diarrhea in El Niño period. The increased diarrhea in case number in June and July of this study was also consistent with the temperature trend of Ha Tinh in the period 2003-2013, the average temperature in the province was the highest in June and July with unclear trends.

The relationship between rainfall and the number of diarrhea cases in this study was found in positive correlation, this result is similar to the results obtained by Carlton’s study in 2014, which indicated that the heavy rains happened in the dry season increased diarrhea by 1.39 times and decreased during the rainy season. Meanwhile, Krumkamp's 2015 study with reported diarrhea per week for 1 year showed that heavy rains increased diarrhea. Thus, the rain can spread pathogens within the community, especially after the dry season, caused the growth of

microorganisms. On the other hand, heavy rains can directly affect infrastructure, living conditions, especially clean water and toilets. In Ha Tinh, the average monthly rainfall was about 204 mm and variated largely. Thus, the results of this study showed the high rate of diarrhea in October in Hatinh, this was also the month with the highest rainfall in the period 2003 – 2013, while in the last months of the dry season (February and March), the rate of diarrhea was low in community.

4.3. Applying the ARIMA models based on climatic factors to predict short term diarrhea in Ha Tinh

We used the data of infectious disease monitoring system and weather monitoring data to build a short-term predictive model of infectious diseases, risk factors in order to enhance the effectiveness of the infectious disease warning system. The infectious disease surveillance system is a very important system for collecting, analyzing and taking action to control infectious diseases.

Various statistical methods have been used to predict the incidence of infectious diseases, of which, the time series modeling has been used to modelling the epidemiological factors to predict future diseases based on the data collected over time. Many researchers have applied several time series methods such as exponential smoothing method, generalized regression method, decomposition methods, multilevel time series models. However, ARIMA (Autoregressive integrated moving average) method is widely used to predict many infectious diseases such as seasonal flu, cholera, dysentery, tuberculosis ... Thus, the application of time series model using seasonality and weathering are relevant for infectious disease monitoring systems with availability, ease of access and ease to use. In addition, a database of weather factors is available and collected according to monthly averages similar to the diarrhea data system.

Results of the study showed that the series of data on temperature, absolute humidity and hours of sunshine were linearly related to the series of diarrhea data in the studied period and these served as the base for incorporating the factors into a predictive model. These data can be collected and available at the Hydrometeorology Station and formatted in excel software so it can be easily combined with the data of the monitoring system before being included in the statistical software for analysis.

Develop, test and predict diarrhea with SARIMA-X model

The inclusion of weather properties in the infectious disease prediction model is a WHO recommendation and of many studies as it is available, has a quantitative correlation to various infectious diseases such as dengue, malaria, cholera. These are climate-sensitive diseases. For model ARIMA (1,1,1) $(0,1,1)_{12}$ with the predictor factor being the average maximum temperature, it showed the appropriateness and statistical significance.

The temperature factor has been also used a lot in building predictive models for dengue fever, flu, malaria, cholera. This model was developed used available data, can be useful in short-term monitoring and predicting diarrhea in Ha Tinh. One month's forecast can help promote communication measures to prevent diarrhea such as hand washing with soap, food safety, properly using the latrines and clean water, and especially vaccination, to reduce the incidence of diarrhea. The 1 year prediction can support health managers in developing suitable plans for more effective control of this disease.

4.4. Limitation of the thesis

For Onjective 1: To determine the incidence of new diarrhea in a commune, we used the method of disease self-recording by diary. This is a method that has been used in many countries with many different diseases/conditions/symptoms. However, during the application, the training of enumerators as well as guidance for villagers to record according to WHO's diarrhea standard still revealved some inaccurate cases. On the other hand, due to the large number of subjects that need to be followed up in the community population, there was an error in the denominator of the total number of people - time. Diarrhea is a common disease so it is easy to recur, in this study we apply the definition of new diarrhea case for subjects who have recured diarrhea in 2 days after the previous diarrhea in the community and only the rated the occurrence if an object have from 2 times or more. To overcome this error, investigators visited the household once a week to check and validate information to avoid omitting or identifying incorrect cases.

During data analysis for Objective 2, we found that diarrhea data in Ha Tinh were recorded by paper reports, or by excel software or by the infectious disease management software. Therefore, it is possible that the data will decrease compared to reality or being lost during storage. Because diarrhea related to many factors, especially clean water, hygienic conditions and recently vaccines, this is a confounding factor that has not been controlled. On the other hand, when analyzing time series, it is necessary to control seasonality and cycle with different techniques, in this study we only used *time function* so it cannot be compared with other methods.

For the application of mathematical modeling in disease prediction, we only have weather factors to include in the model without economic, social factors, medical interventions that related to the disease, there will therefore be the errors in the forecasting process. To overcome these factors, in the analysis process, algorithms and evaluation indicators were used to ensure that the parameters were large enough and to ensure the white noise for the residuals.

CONCLUSION

Through the research results and discussions mentioned above, we have some following conclusions:

1. The epidemiological characteristics of diarrhea in a commune of drought-affected area in Ha Tinh province in 2014 – 2015

The diarrhea incidence rate was 3.3% person-month (40% person-year) in general and 5.1% child-month among the under-5-year-old group, 1.6 times higher than other age group.

The cumulative rate of new infections was 35.6%/year, this among the group of under 5 years old was 48.7%, higher than of other age groups.

The average number of diarrhea episodes was 0.78 times/person/year and raised up to 1.52 times/person/year among group of under 5 years old.

The recurrence rate was 1.87% person - month (22.4% person - year), this rate was higher among children under 5 years old, higher than that of other age groups.

The rate of cumulative diarrhea by month of the community was 6.3% and it was 13.1% among children under 5 years old.

Diarrhea occured all year round with the highest incidence rate in the summer months (June, July and August).

2. The relationship between some weather factors and diarrhea at Ha Tinh in the period of 1992 - 2015

Temperature factor at present and in 1-2 months prior was positively correlated with the number of diarrhea cases. An increase of 1^0C in average temperature/maximum temperature or average minimum temperature might pose the risk of increase diarrhea cases by 2-3%.

10 mm increase in total rainfall in 2-3 months prior might pose the risk of increasing the current number of diarrhea cases by 17.1%.

An increase of 1g/m3 in absolute humidity in 1-2 months prior ago was found related to the risk of increasing the case number of diarrhea from 2.8 to 3.3%.

An increase of 1^0C of the sea surface temperature in NINO3 region in the 3-month delay step might potentially increase the number of current cases by 13.8%.

3. Applying the ARIMA models based on climatic factors to predict short term diarrhea in Ha Tinh

SARIMA Model (1,1,1) $(0,1,1)_{12}$ showed to have a predictive error of 7,6% before 1 month and of 10,1% before 12 months by Ljung-Box Q test with $p>0.05$, AIC value of 3889.4 and BIC value of 3907.7.

SARIMA-X model (1,1,1) $(0,1,1)_{12}$ -Tmax possed a predictive error of 6,1% before 1 month and of 9,7% before 12 months by Ljung-Box Q test with $p> 0.05$ and AIC value of 3735.0 and BIC value of 3756.7.

RECOMMENDATION

Regular planning and funding are neccesary for diarrheal disease prevention and control in the community, especially for children under 5 years old and school age, including vaccination against diarrhea.

Measures should be taken to increase community knowledge on diarrheal prevention measures such as improving clean water supply, sanitation and hand washing with soap.

Assign specialized staff to manage and update diarrhea data and collect available weather data for periodically forecast

Implement the diarrheal early prediction model using available weather data for early warning system of weather-related diseases

LIST OF PUBLISHED SCIENTIFIC ARTICLES RELATED TO THIS THESIS

1. Phan Dang Than, Nguyen Ngoc Sang, Le Thi Phuong Mai (2017), "Cumulative incidence of acute diarrhea syndrome and some related weather factors in a commune of Ha Tinh province, period 2014 - 2015", *Journal of Vietnam Preventive Medicine, No* 27(9): pp 177 - 194.
2. Phan Dang Than, Nguyen Ngoc Sang, Le Thi Phuong Mai (2018), "Use of ARIMAX model and some weather factors for short-term prediction of diarrhea in Ha Tinh province", *Journal of Vietnam Preventive Medicine*, No 28(12): pp 70 -78.
3. Phan Dang Than, Le Thi Phuong Mai, Nguyen Tu Quyet, Luu Phuong Dung, Tran Van Dinh, Tran Ngoc Phuong Mai, Nguyen Thi My Hanh, Hoang Hai (2019), "The impact of temperature factors on the number of diarrhea cases in Ha Tinh in the period 1992 – 2017", *Journal of Vietnam Preventive Medicine*, No.29 (2), pp 103 -108.