Chapter 4

Further Analysis and Statistical Modeling

The associations between sickness with diarrhea disease and the determinant variables are analysed in this chapter. We use logistic regression to fit a model. This analysis is based on the risk factors related to diarrhea disease.

4.1 Logistic Regression Analysis Method

Logistic regression analysis is used to fit a model relating the probability of occurrence to each determinant. There are four groups of determinant variables in this study, comparing socio-demographic factors, environmental factors, knowledge factors and behaviour factors, with response categories for each determinant. Using a chi-squared statistic can check the fit of the model. We start with choosing the determinant variables of interest, which have p-values < 0.1 from the preliminary data analysis in Chapter 3. Some variables which have p-values > 0.1 that are considered important may also be included in the final model (Table 4.1).

The list of all determinant variables of interest with their p-values obtained in Chapter 3 is as follows.

Determinant	P-value	Determinant	P-value
Age	0.044	Wash after use toilet	0.029
Sex	0.087	Keep food overnight	0.008
Water quality	0.010	Clean dress	0.002
Defecation place	0.097	Clean fingernails	0.000
Pets	0.000	Religion	0.267
Keep rubbish	0.060	Education	0.305
Knowledge of symptoms	0.085	Filthy water	0.545
Knowledge of keep ORS	0.005	Knowledge of cause	0.255
Knowledge of transfer	0.091	Knowledge of cause of death	0.200

Table 4.1 Association between determinants and diarrhea disease

4.2 Results

Figure 4.1 gives the result of the logistic regression analysis with all these variables included. These variables give 18 categories containing four variables from socio-demographic factors, five variables from the environmental factors, five variables from the knowledge factors and four variables from behaviour factors.

From Figure 4.1, eight variables are statistically significant as risk factors related to acute diarrhea disease. There are six variables having high p-values including age, water quality, keep rubbish, defecate place, knowledge transfer and knowledge how long to keep ORS with p-value of 0.61, 0.46, 0.96, 0.41, 0.50 and 0.36 respectively. The deviance from this model is 210.41 and the number of degrees of freedom is 199.

Figure 4.2 shows the result of fitting a reduced model in which six of the least statistically significant risk factors are omitted. The factors omitted comprise age of carer, three of the hygiene determinants and two of the knowledge factors.

In Figure 4.2, after reducing the model by omitting these six variables, it is found that four variables have high p-values including sex (p-value = 0.06), knowledge cause (p-value = 0.11), knowledge symptom (p-value = 0.27), and washing hands after use of toilet (p-value = 0.06). The deviance and the number of degree of freedom have increased from 210.41 to 215.28 and from 199 to 206, respectively.

logistic regression analysis: Study of diarrhea disease with risk factors

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
sick						
sick / not sick	-5.1038	1.3509	0.0002	0.0061	0.0004	0.0858
Age	0.0117	0.0231	0.6126	1.0118	0.967	1.0586
Sex						
female	(0)					
male	-2.6669	1.3058	0.0411	0.0695	0.0054	0.8981
Religion						
buddhist	(0)					
muslim	0.7539	0.4571	0.099	2.1253	0.8677	5.2055
Education						
primary	(0)					
above primary	1.2922	0.4658	0.0055	3.6407	1.4612	9.0711
Water quality			0.4589			
boil	(0)					
filter	0.7774	0.8157	0.3406	2.1759	0.4398	10.7646
not improve	0.3568	0.415	0.3899	1.4288	0.6334	3.2229
pets						
no pets	(0)					
have pets	1.7747	0.4644	0.0001	5.8987	2.3738	14.6578
Filthy water						
have filthy water	(0)					
not filthy water	1.0803	0.5173	0.0368	2.9456	1.0686	8.1189
Keep rubbish						
keep rubbish	(0)					
not keep	0.0216	0.4128	0.9582	1.0219	0.455	2.2949
defecate place						
in the toilet	(0)					
out the toilet	0.3905	0.471	0.407	1.4777	0.5871	3.7196
knowledge cause						
incorrect	(0)					
correct	-0.5457	0.5824	0.3488	0.5795	0.185	1.8145
Knowledge symptom						
incorrect	(0)					
correct	-0.3586	0.3685	0.3305	0.6987	0.3393	1.4386
Transfer						
incorrect	(0)					
correct	-0.2824	0.42	0.5014	0.754	0.331	1.7174
Knowledge cause death						
incorrect	(0)					
correct	0.854	0.3834	0.0259	2.3489	1.1078	4.9804
How long to keep ORS						
incorrect	(0)					
correct	-0.4019	0.4365	0.3571	0.669	0.2844	1.574
Wash after use toilet						
every time	(0)					
not every time	0.79	0.5379	0.1419	2.2035	0.7677	6.3244
Keep food over night						
every time	(0)					
not every time	1.0992	0.394	0.0053	3.0019	1.3868	6.4979
clean dress						
boil and wash with soap	(0)					
wash and not boil	1.1855	0.5608	0.0345	3.2722	1.0901	9.8218
clean fingernails			0.0154			
twice a week	(0)					
once a week	1.3287	0.4623	0.004	3.7763	1.5261	9.3445
less than once a week	0.899	0.5588	0.1076	2.4573	0.8219	7.3469

df: 199 deviance: 210.412 number of iterations: 4

Figure 4.1 Logistic regression model for diarrhea disease with risk factors

logistic regression analysis: Study of diarrhea disease with risk factors

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
sick						
sick / not sick	-4.7711	1.0055	0	0.0085	0.0012	0.0608
Sex						
female	(0)					
male	-2.2973	1.204	0.0564	0.1005	0.0095	1.0646
Religion						
buddhist	(0)					
muslim	0.8204	0.4087	0.0447	2.2714	1.0195	5.0607
Education						
primary	(0)					
above primary	1.1998	0.436	0.0059	3.3194	1.4124	7.8009
pets						
no pets	(0)					
have pets	1.9372	0.44	0	6.9393	2.9296	16.4372
Filthy water						
have filthy water	(0)					
not filthy water	1.0359	0.4895	0.0343	2.8177	1.0795	7.3549
knowledge cause						
incorrect	(0)					
correct	-0.8354	0.528	0.1136	0.4337	0.1541	1.2207
Knowledge symptom						
incorrect	(0)					
correct	-0.3973	0.3579	0.2669	0.6721	0.3333	1.3554
Knowledge cause death						
incorrect	(0)					
correct	0.7011	0.3574	0.0498	2.016	1.0005	4.0619
Wash after use toilet						
every time	(0)					
not every time	0.9516	0.5043	0.0592	2.5898	0.9638	6.9591
Keep food over night						
every time	(0)					
not every time	1.1378	0.3756	0.0025	3.1198	1.4941	6.5146
clean dress						
boil and wash with soap	(0)					
wash and not boil	1.1708	0.5389	0.0298	3.2246	1.1213	9.2729
clean fingernails			0.005			
twice a week	(0)					
once a week	1.4776	0.4561	0.0012	4.3824	1.7925	10.7147
less than once a week	1.0167	0.5469	0.063	2.7642	0.9462	8.0747

df: 206 deviance: 215.276 number of iterations: 4

Figure 4.2 Logistic regression model for diarrhea disease after six factors omitted

Figure 4.3 shows the result after omitting sex and the knowledge factors.

logistic regression analysis: Study of diarrhea disease with risk factors

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
sick						
sick / not sick	-4.863	0.9779	0	0.0077	0.0011	0.0525
Religion						
buddhist	(0)					
muslim	0.8409	0.3928	0.0323	2.3185	1.0737	5.0064
Education						
primary	(0)					
above primary	1.2145	0.427	0.0045	3.3687	1.4588	7.7791
pets						
no pets	(0)					
have pets	1.9286	0.434	0	6.8797	2.9388	16.1056
Filthy water						
have filthy water	(0)					
not filthy water	0.9803	0.4824	0.0421	2.6652	1.0354	6.8608
knowledge cause						
incorrect	(0)					
correct	-0.7934	0.5165	0.1245	0.4523	0.1644	1.2447
Knowledge cause death						
incorrect	(0)					
correct	0.6863	0.3494	0.0495	1.9864	1.0014	3.9402
Wash after use toilet						
every time	(0)					
not every time	0.9894	0.4984	0.0471	2.6895	1.0126	7.1437
Keep food over night						
every time	(0)					
not every time	1.0665	0.3656	0.0035	2.9051	1.4189	5.948
clean dress						
boil and wash with soap	(0)					
wash and not boil	1.1507	0.5288	0.0296	3.1605	1.121	8.9103
clean fingernails			0.0076			
twice a week	(0)					
once a week	1.3902	0.4479	0.0019	4.0156	1.669	9.6616
less than once a week	0.9282	0.5341	0.0822	2.53	0.8882	7.2072

df: 208 deviance: 221.357 number of iterations: 4

Figure 4.3 Further reduced logistic regression model for diarrhea disease after two more factors are omitted

There is now only one factor, causes knowledge, that is not statistically significant. We omitted this determinant to arrive at the model with nine determinants, shown in Figure 4.4

logistic regression analysis: Study of diarrhea disease with risk factors

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
sick						
sick / not sick	-5.4483	0.91	0	0.0043	0.0007	0.0256
Religion						
buddhist	(0)					
muslim	0.896	0.3863	0.0204	2.4497	1.1488	5.2235
Education						
primary	(0)					
above primary	1.1452	0.4217	0.0066	3.1432	1.3753	7.1833
pets						
no pets	(0)					
have pets	1.937	0.431	0	6.9377	2.9809	16.1468
Filthy water						
have filthy water	(0)					
not filthy water	0.913	0.477	0.0556	2.4918	0.9783	6.3468
Knowledge cause death						
incorrect	(0)					
correct	0.6076	0.3417	0.0753	1.8361	0.9399	3.5868
Wash after use toilet						
every time	(0)					
not every time	0.8817	0.4927	0.0735	2.4151	0.9194	6.3439
Keep food over night						
every time	(0)					
not every time	0.99	0.3579	0.0057	2.6912	1.3345	5.4272
clean dress						
boil and wash with soap	(0)					
wash and not boil	1.2167	0.5238	0.0202	3.376	1.2093	9.4253
clean fingernails			0.0074			
twice a week	(0)					
once a week	1.372	0.4411	0.0019	3.9432	1.6611	9.3603
less than once a week	0.9038	0.5296	0.0879	2.469	0.8745	6.971

df: 209 deviance: 223.744 number of iterations: 4

Figure 4.4 Further reduced model for diarrhea disease

Figure 4.5 presents the result of fitting a reduced model for diarrhea disease with risk factors knowledge of cause of death and wash after use of toilet omitted. The deviance and the number of degree of freedom have increased from 223.74 to 229.81 and 209 to 211, respectively.

logistic regression analysis: Study of diarrhea disease with risk factors

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
sick						
sick / not sick	-4.7708	0.8211	0	0.0085	0.0017	0.0424
Religion						
buddhist	(0)					
muslim	0.7532	0.3705	0.0421	2.1238	1.0273	4.3906
Education						
primary	(0)					
above primary	1.1353	0.4126	0.0059	3.1121	1.3863	6.9865
pets						
no pets	(0)					
have pets	1.8943	0.4187	0	6.6479	2.9259	15.1045
Filthy water						
have filthy water	(0)					
not filthy water	0.7456	0.4578	0.1034	2.1078	0.8593	5.1703
Keep food over night						
every time	(0)					
not every time	0.9093	0.3455	0.0085	2.4825	1.2613	4.886
clean dress						
boil and wash with soap	(0)					
wash and not boil	1.2984	0.5165	0.0119	3.6633	1.3311	10.0815
clean fingernails			0.0052			
twice a week	(0)					
once a week	1.3657	0.4328	0.0016	3.9184	1.6777	9.1517
less than once a week	0.7385	0.5165	0.1528	2.0928	0.7604	5.7592

df: 211 deviance: 229.814 number of iterations: 4

Figure 4.5 Logistic regression model for diarrhea disease with risk factors after knowledge of cause of death and wash after use of toilet are omitted

After reducing the model by omitting filthy water, it was found that every variable in this model is significant except religion (Figure 4.6). The deviance and the number of degree of freedom have increased from 229.81 to 232.57 and 211 to 212, respectively.

logistic regression analysis: Study of diarrhea disease with risk factors

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
sick						
sick / not sick	-4.1611	0.7184	0	0.0156	0.0038	0.0637
Religion						
buddhist	(0)					
muslim	0.5994	0.3553	0.0916	1.821	0.9075	3.6541
Education						
primary	(0)					
above primary	1.1078	0.4095	0.0068	3.0276	1.3567	6.7561
pets						
no pets	(0)					
have pets	1.6547	0.3801	0	5.2317	2.4838	11.0193
Keep food over night						
every time	(0)					
not every time	0.9517	0.3414	0.0053	2.59	1.3265	5.0569
clean dress						
boil and wash with soap	(0)					
wash and not boil	1.4238	0.5221	0.0064	4.1528	1.4926	11.5541
clean fingernails			0.0025			
twice a week	(0)					
once a week	1.3992	0.4281	0.0011	4.0519	1.7509	9.3769
less than once a week	0.6182	0.5015	0.2177	1.8556	0.6944	4.959

df: 212 deviance: 232.575 number of iterations: 4

Figure 4.6 Logistic regression model for diarrhea disease with risk factors after filthy water is omitted

When religion is omitted all of variables in this model become significant (Figure 4.7). The deviance and the number of degree of freedom have increased from 232.57 to 235.48 and 212 to 213, respectively. After adjusting for various factors, it is

observed that education, pets, keep food over night, clean dress and clean fingernails exhibit significant associations, particularly pets, which has the smaller p-value.

When comparing the values of the deviance and degrees of freedom from Figure 4.7 and the model from Figure 4.3, it was found that difference between the deviance and degree of freedom of this model is close to that of the model in Figure 4.3. However we choose this model because all variables are statistically significant (in Figure 4.3 the knowledge of cause is not significant). It is clear that the reduced model is better.

logistic regression analysis: Study of diarrhea disease with risk factors

factor	coeff	St.Error	p-value	Odds ratio	95% CI	
sick						
sick / not sick	-3.7658	0.6559	0	0.0231	0.0064	0.0837
Education						
primary	(0)					
above primary	0.8542	0.376	0.0231	2.3496	1.1245	4.9095
pets						
no pets	(0)					
have pets	1.5812	0.3743	0	4.8609	2.3341	10.123
Keep food over night						
every time	(0)					
not every time	0.8699	0.3361	0.0096	2.3867	1.2352	4.6118
clean dress						
boil and wash with soap	(0)					
wash and not boil	1.4843	0.5119	0.0037	4.4117	1.6176	12.0319
clean fingernails			0.0018			
twice a week	(0)					
once a week	1.4147	0.4226	0.0008	4.1152	1.7976	9.4209
less than once a week	0.6038	0.496	0.2235	1.8291	0.6919	4.8353

df: 213 deviance: 235.483 number of iterations: 4

Figure 4.7 Logistic regression model for diarrhea disease with risk factors after religion is omitted

The final model a shown in Figure 4.7 shows that the 5 variables education, pets, keep food over night, clean dress and clean fingernails are all related to acute diarrhea disease in children aged under 5 years in Pattani Province. These factors are used to assess the probability of risk factors related to acute diarrhea disease in children aged under 5 years in terms of log odds of sickness with diarrhea disease as a linear function of these explanatory variables.

4.3 Model of log odds of sickness with diarrhea disease in children aged under 5 years.

The mathematical model from the logistic regression print out as shown in Figure 4.7, where Y is the log odds of sickness with diarrhea disease, and X_n are the explanatory variables, is as follows.

$$Y = \{ - 1.3147 + 0.8542X_1 + 1.5812X_2 + 0.8699X_3 + 1.4843X_4 + 1.4147X_5 + 0.6038X_6 \}$$

Where X_1 = (education above primary school)

X_2 = (have domestic animal)

X_3 = (reheat milk and food over night not every time)

X_4 = (cleaning the patient's dress by washing and not boiling)

X_5 = (clean fingernails once a week)

X_6 = (clean fingernails less than once a week)

This equation is modeled in terms of the log odds of sickness with diarrhea disease. Given that the study is cross-sectional, this equation may be use to give the probability of diarrhea sickness.

4.4 Model of probability of sickness with diarrhea disease in children aged under 5 years

The model estimates of derived from the regression coefficients as shown in Figure 4.7 are obtained by using equation (2.14) in Chapter 2 to give an expression for the probability of sickness with diarrhea disease as follows.

$$P = \frac{1}{1+\exp\left[1.3147 - 0.8542X_1 - 1.5812X_2 - 0.8699X_3 - 1.4843X_4 - 1.4147X_5 - 0.6038X_6\right]}$$

For example, suppose that mother and child carers is female, education above primary school, have domestic animal and reheat food over night every time.

$$P = \frac{1}{1+\exp\left[1.3147 - 0.8542X_1^{(1)} - 1.5812X_2^{(2)} - 0.8699X_3^{(1)}\right]}$$

$$= 0.879$$

The conclusion is that this child has a probability of being sick with diarrhea disease of about 88%. We have obtained a probability model to provide useful estimation of probabilities of being sick with diarrhea disease in children aged under 5 years in Pattani Province, giving a simplified interpretation.

The advantage of this model is that physicians, health workers and public health volunteers can use it as a screening test for prevention and control of sickness with diarrhea disease in children aged under 5 years in Pattani Province.