

# THE USE OF LOGISTIC REGRESSION MODEL TO IDENTIFY THE RISK FACTOR OF H5N1 AVIAN INFLUENZA VIRUS OF NATIVE CHICKEN IN SUMATERA AND KALIMANTAN ISLAND, INDONESIA

<sup>1</sup>Etih Sudarnika, <sup>2</sup>Asep Saefuddin, <sup>3</sup>Abdul Zahid, and <sup>4</sup>Chaerul Basri

<sup>1,3,4</sup>Laboratory of Epidemiology, Faculty of Veterinary Medicine, IPB, 16680, Darmaga, Bogor, Indonesia.

<sup>2</sup>Departement of Statistics, Faculty of Mathematics and Science, IPB, 16680, Darmaga, Bogor, Indonesia.

e-mail: <sup>1</sup>etih@ipb.ac.id, <sup>2</sup>asaefuddin@gmail.com, <sup>3</sup>zahid@ipb.ac.id, <sup>4</sup>chaerul@ipb.ac.id

**Abstract.** The cross sectional study had been carried out in December 2005 at Kalimantan and Sumatera Island, Indonesia. The objective of this study was to apply the logistic regression model to identify the risk factor of H5N1 avian influenza virus of native chicken. 12,713 serum samples of chicken from 498 farmers was collected. The H5N1 virus was tested by Haemagglutination Inhibition (HI) test from serum samples and the information of risk factor was obtained from a questionnaire. The questionnaire included farmer's characteristic and farm management. Logistic regression model showed that an association with H5N1 virus infection risk at a 5% significance level was found for chicken house sanitation, feeding equipment sanitation, drinking equipment sanitation and chicken run sanitation. The multivariable regression analysis showed that an association with H5N1 virus infection risk at a 5% significance level was found for chicken house sanitation.

**Key word:** logistic model, generalized linear model, odds ratio, binary data

## 1 Introduction

Highly pathogenic avian influenza caused by H5N1 virus infection has spread out in many countries in the World, and continues to spread to new countries and to new areas in some countries (FAO 2007). The infection cause enormous economical losses in the agricultural sector, especially the poultry industry, due to high mortality and restriction of international trade as well as public health concern. The most serious threat of H5N1 is the possibilities to mutate into a new form that can infect human and spread rapidly from person to person leading to a pandemic situation. Since the disease first surfaced in 2003, H5N1 has killed at least 171 people worldwide, 66 in Indonesia, the highest human death toll of any country.

In Indonesia the disease was suspected to infect the fowls since July to August 2003. To date only 3 provinces from the 33 provinces in Indonesia are free from bird flu (FAO 2007). It is very difficult to control the avian influenza disease in Indonesia. Because Indonesia is a vast country which has 17,000 islands, weak national veterinary service and low financial resources to prevent and control the disease. Besides that, it is common for Indonesian to breed a few fowls in their backyard farm and this make it easier for the virus to spread and more difficult to control.

The spreading media and the infections of H5N1 virus can be caused through (a) fowl feces, (b) livestock transportation, (c) infected chicken house equipment, (d) infected feed and drinks, (e) farm workers, (f) wild birds (Agricultural Ministry of Indonesia 2007). This virus can survive on feces at least

for 35 days at 4 °C. One gram of the fowl infected feces can infect one million other fowls (Faculty of Veterinary Medicine, Bogor Agricultural University 2005).

Faculty of Veterinary Medicine, Bogor Agricultural University conducted a cross sectional study of virus H5N1 infection on fowls in Sumatera and Kalimantan Islands in 2005 funded by the Agricultural Ministry of Republic Indonesia. The objectives of the study is to predict seroprevalence H5N1 virus on fowls, especially the native chicken (Kampong chicken) and to identify the risk factors of H5N1 virus introduction.

Logistic regression model is used to identify the risk factor which influence the H5N1 virus infection and to measure the odds ratio for each risk factors. This information is used to design the avian influenza control program.

## 2 Data

Data is collected between November to December 2005 through cross-sectionals study. The survey had been carried out in seven provinces in Sumatera Island, i.e. North Sumatera, West Sumatera, South Sumatera, Jambi, Bangka Belitung, Bengkulu and Lampung, and also four provinces in Kalimantan, i.e. West Kalimantan, Center Kalimantan, South Kalimantan and East Kalimantan. The data were collected from 23 districts in Sumatera Island and 10 districts in Kalimantan Island. The sampling used is a multistage random sampling (probability proportional to size) (Dawson and Trapp 2004).

Sample types that had been collected were the blood serum and cloacal swab samples of native chicken. The other information were the characteristics of human resources (the farmers) and farming management which were filtered through structured questionnaires. The characteristics of farmers are categorized by age, formal educational level, knowledge, the business purpose, farming experience period, other activities besides farming and also extension services. The characteristic of farming management were categorized by farming system, number of chickens, chicken house type, feeding system, movement control, sanitation, health management, reporting system, quarantine action and handling feces management. The farmer's characteristic data was conducted through interview, and for the characteristic of farming management data was collected through observation and interview. Before the survey was carried out, all of the blood sample and cloacal swab collectors and enumerators were trained to minimize information bias.

The blood serum was tested by Haemagglutination Inhibition (HI) test and Enzyme-linked Immunosorbant Assay (ELISA) technique. The cloacal swab was tested by Reverse Transcriptase-Polymerase Chain Reaction (RT-PCR). Those tests were assembled to have a positive or negative result of avian influenza H5N1 virus infection.

Through the survey, 12,713 blood samples and cloacal swab were taken and 498 farmers were interviewed. The blood samples and cloacal swabs were obtained from 5 - 50 chickens from each farmer which is proportional to the number of chicken the farmer own.

On this research, the virus H5N1 infection data was based on HI test because HI test was used on every samples, but ELISA and PCR test were only used to verify the HI test. The H5N1 virus infection is positive if at least one chicken in the farm is tested positive HI, and it would be negative if none of the chicken in the sample is tested positive HI. The data was processed using SAS (Statistical Analysis Software, v9.1) (SAS Institute, Inc., Cary, NC)

## 3 The Model and Estimates

The logistic regression is a form of *Generalized Linear Model* for binary response variable. It is known that *Generalized Linear Model* consist of three components namely 1) random component, which is a distribution function for  $Y$  response variable. The distribution function of  $Y$  response variable has to be a

class of exponential family distribution. 2) systematic component, which is explanatory variables that are used as linear predictors in the model  $\eta = X\beta$  3) link function  $g$ ,  $E(y)=\mu=g^{-1}(\eta)$  which is the link function between the random component and the systematic component that shows the relation between the predictor  $X$  and the response  $Y$  (McCullagh and Nelder 1989, Dobson 2002).

Since the response variable follow Binomial distributions this model uses a logit function as the link function. Table 1 shows some common link functions for some exponential family distribution.

Table 1. Link function

Distribution	Name	Link Function	Mean Function
Normal	Identity	$X\beta = \mu$	$\mu = X\beta$
Exponential	Inverse	$X\beta = \mu^{-1}$	$\mu = (X\beta)^{-1}$
Gamma			
Poisson	Log	$X\beta = \ln(\mu)$	$\mu = \exp(X\beta)$
Binomial	Logit	$X\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(X\beta)}{1 + \exp(X\beta)}$
Multinomial			

Logistic regression is one of the regression models that are commonly used in medical research especially epidemiological science.

The Logistic Equation Function is (Kleinbaum 1994):

$$f(z) = \frac{1}{1 + e^{-z}}$$

$$\begin{aligned} \text{If } z &= \alpha + \beta_1 x_1 + \dots + \beta_k x_k \\ &= \alpha + \sum \beta_i x_i \end{aligned}$$

$$\text{Then } f(z) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i x_i)}}$$

Logistic regression model:

$$p_i = \Pr(Y_i = 1 | X_1, X_2, \dots, X_k) = \frac{1}{1 + e^{-(\alpha + \sum \beta_i X_i)}}$$

$$\text{logit}(p_i) = \ln\left(\frac{p_i}{1 - p_i}\right) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

The parameters  $\alpha, \beta_1, \beta_2, \dots, \beta_k$  are usually estimated by using the maximum likelihood method. Besides using the model to predict the probability of known individual event, logistic regression model is also used to predict the value of odds ratio (OR) on risk factors. OR is an association measure which can be used on case-control and cross sectional study. In logistic regression model, the odds ratio value for  $X_i$  factor is predicted by  $e^{\beta_i}$ , where  $\beta_i$  is the regression coefficient of  $X_i$  factor.

The problems that is commonly found on logistic regression model is **overdispersion**. Overdispersion is one condition where the observation's variance is higher than the theoretical model's variance. If this happen the results will effect the decision making, where the model tends to reject  $H_0$ .

Overdispersion may be due to a random variation in the outcome probability, correlated outcome, and cluster effect. Overdispersion may also happens because the data resulted of a clustered process each event randomly contributed to total data generally, in other words there is a mix on the distribution (McCullagh and Nelder 1989).

Overdispersion can be determined by checking the deviance value and Pearson  $\chi^2$  statistics divided by the degree of freedom, which means if the value is around one, indicate no overdispersion. If overdispersion is found then it can be solved by changing the distribution function of the model into beta binomial distribution for binomial distribution.

The observed response variable ( $Y$ ) is the positive or negative test result of H5N1 virus infection and its relation need to be found with  $X_1$ =treatment for sick chicken,  $X_2$ =quarantined of new chickens on the farm,  $X_3$ =farmer's knowledge about avian influenza,  $X_4$ =farming system,  $X_5$ =chicken house type,

$X_6$ =ventilation,  $X_7$ =population density,  $X_8$ =chicken house sanitation,  $X_9$ =feeding equipment sanitation,  $X_{10}$ =drinking equipment sanitation,  $X_{11}$ =chicken run sanitation,  $X_{12}$ =handling feces management and  $X_{13}$ =human and vehicle movement control. On each factor, the association with Y variable is tested using chi-square test for independence. Bivariate analysis result was used to conduct logistic regression model. Table 2 shows the bivariate analysis results.

Table 2. Odds ratio value for each risk factor X

Factor	HI Test +		HI Test -		Crude OR	95% CI
	n	%	n	%		
<b>Treatment for infected chicken</b>						
Not separated with other chicken	31	45	94	47	1.005	0.534 ; 1.894
Separated with other chicken	38	55	104	53	1.00	
<b>Quarantined new chicken on the farm</b>						
Yes	39	44	85	28	1.00	
No	50	56	132	72	0.873	0.492 ; 1.560
<b>Knowledge of Avian Influenza</b>						
Bad	16	18	40	14	1.452	0.510 ; 3.958
Average	47	52	151	52	1.199	0.591 ; 2.505
Good	27	30	97	34	1.00	
<b>Farming System</b>						
Indoor chicken	25	21.5	69	25	1.00	
Chicken house with fence	46	40	87	32	2.055	0.972 ; 4.483
Restricted ranged chicken (inside fence)	13	11	25	9	2.571	0.901 ; 7.272
Free ranged chicken	32	27.5	92	34	1.470	0.668 ; 3.299
<b>Chicken house Type</b>						
Stage	74	64	160	59	1.00	
Floor	42	36	113	41	1.344	0.750 ; 2.396
<b>Ventilation</b>						
Good	90	80	222	85	1.00	
Bad	22	20	40	15	1.226	0.599 ; 2.435
<b>Population density</b>						
< 8 Chicken/m <sup>2</sup>	38	34	110	42	1.00	
≥ 8 Chicken /m <sup>2</sup>	73	66	151	58	1.239	0.694 ; 2.243
<b>Chicken house Sanitation</b>						
Clean	17	15	79	30	1.00	
Moderate	70	60	143	54	3.297*	1.575 ; 7.466
Dirty	29	25	42	16	4.392*	1.815 ; 11.179
<b>Feeding Equipment Sanitation</b>						
Clean	23	20	84	33	1.00	
Moderate	70	60	143	54	3.297*	1.575 ; 7.466
Dirty	31	27	45	18	4.351*	1.908 ; 10.236
<b>Drinking Equipment Sanitation</b>						
Clean	22	19	88	35	1.00	
Moderate	70	60	143	54	3.297*	1.575 ; 7.466
Dirty	31	28	44	18	6.235*	2.705 ; 5.013
<b>Chicken run Sanitation</b>						
Clean	22	19	89	34	1.00	
Moderate	76	65.5	145	56	2.807*	1.403 ; 5.969
Dirty	18	15.5	27	10	3.647*	1.393 ; 9.706
<b>Handling Feces Management</b>						
No Treatment	51	45	116	45	1.693	0.895 ; 3.252
With Treatment (piled in a hole, placed in a sacked, etc)	63	55	140	55	1.00	
<b>Movement control</b>						
Exist	24	22	64	25	1.00	
None	87	78	196	75	1.211	0.642 ; 2.356

\*) Showing significant association at  $\alpha=0.05$

Table 2 shows the odds ratio value and the 95% confidence interval result for the bivariate analysis for each risk factor.

From Table 2 we can see that there are four factors that significant result which are chicken house sanitation (moderate: OR=3.297, 95%CI 1.575-7.466, dirty: OR=4.392, 95%CI 1.815-11.179 ), feeding

equipment sanitation (moderate: OR=2.292, 95%CI 1.150-4.763, dirty: OR=4.351, 95%CI 1.908-10.236), drinking equipment sanitation (moderate: OR=2.503, 95%CI 1.240-5.300, dirty: OR=6.235, 95%CI 2.705-5.013) and chicken run sanitation (moderate: OR=2.807, 95%CI 1.403-5.969, dirty: OR=3.647, 95%CI 1.393-9.706). Using those variables we conduct a multivariable logistic regression model where the parameter estimation was based on maximum likelihood method of stepwise logistic regression.

The stepwise logistic regression showed there is only one significant factor which is chicken house sanitation. Table 3 shows the OR measure for the factor.

Table 3. Relation between characteristic of farming management against H5N1 virus infection at native chicken husbandry

Factors	HI Test +		HI Test -		Adjusted OR	95% CI
	n	%	n	%		
Chicken house sanitation						
Clean	17	15	79	30		
Moderate	70	60	143	54	2.486	1.390 ; 4.643
Dirty	29	25	42	16	3.217	1.584 ; 6.699

Checking for overdispersion we found that the deviance values divided by degree of freedom is 1.4494 and Pearson  $\chi^2$  divided by degree of freedom is 1.3549. Both values were close to 1 which mean there was no overdispersion on the prediction model so it is quite a good model that can be used to describe the data.

## 4 Discussion

Univariate analysis result shows that the influence factors against H5N1 virus infection on native chickens generally are the sanitation factors namely feeding equipment, drinking equipment and chicken house's field. A good sanitation is one of the method to prevent the avian influenza to spread. H5N1 virus can be killed by disinfectant and sunlight. There was no previous information on the disease status of the survey area which determine whether it was a free avian infection area or not. This causes difficulties on measuring association of the quarantine treatment of new chickens and movement control against H5N1 virus infection. In a free infected area, there would be no virus infection as long as new chickens came from the same area even though no quarantine treatment were applied. Similar situation was found in the movement control, farm in the free area will have lower infection risk although no movement control procedure was applied. Most of native chicken in Indonesia are reared under traditional husbandry system with very weak movement control and therefore theoretically they will be easy to be infected with various infectious diseases. However, there was no adequate data to prove an association between the movement control and H5N1 virus Infection in this study.

The limitation of this study was the cross-sectional nature which is based only in a single observation. This can cause information bias during observation on farm situation. For example during the population density observation, enumerator may under estimate chicken population in a farm because chicken tend to move around to other places during the observation.

Selection bias might also happen during the blood and cloacal swab samples collection. The enumerators usually took samples from chicken in the house whereas other animals were difficult to catch. This may cause the samples taken were only from the chicken they were kept in the chicken house. The other chickens, which were running around outside the chicken house and thus have higher risk to be infected by the virus, were not selected to be sampled.

Another limitation of this study was no confounding factors and also the possibilities of a modifier effect have not been calculated. The modifier effect was not included because there is only one significant variable on the multivariable analysis.

Although some limitations are present in this study, results of this study provide useful basic information for designing avian influenza control program in Indonesia. Future research must anticipate

the possibilities of a selection bias and information bias and filtering important information which is related to the risk of the H5N1 virus infection.

## Acknowledgements

Thanks to Agricultural Ministry of Indonesia which funded this research and thanks to all of the analyst research team from Animal Medical Faculty of Bogor Agricultural University, Bogor Indonesia for their advice.

## References

- Dawson B and Trapp RG. 2004. Basic & Clinical Biostatistics 4<sup>th</sup> edition. Singapore: McGraw-Hill International
- Departemen Pertanian RI. 2007. Sekitar Flu Burung: Informasi untuk Peternak dan Konsumen. <http://www.litbang.deptan.go.id/berita/one/221/>. 11 April 2007.
- Dobson AJ. 2002. An Introduction to Generalized Linear Models 2<sup>nd</sup> ed. London: Chapman and Hall/CRC.
- FAO. 2007. Fewer bird flue outbreaks this year-sign of progress. <http://www.fao.org/newsroom/en/news/2007/1000534/index.html> 11 April 2007.
- Faculty of Veterinary Medicine Bogor Agricultural University. 2005. Laporan Kajian Seroepidemiologi Penyakit Avian Influenza serta Strategi Penanggulangan dan Pencegahannya di Sumatera dan Kalimantan. Tidak Dipublikasikan. Bogor: IPB.
- Kleinbaum DG. 1994. Logistic Regression -A Self -Learning Text. New York: Springer.
- McCullagh P and Nelder JA. 1989. Generalized Linear Models 2<sup>nd</sup> ed. London: Chapman and Hall.