

# LOG-LINEAR RISK RATE MODEL FOR HOSPITAL SERVICES

**Krieng Kitbumrungrat\***

*Received: August 10, 2012; Revised: January 02, 2013; Accepted: January 03, 2013*

## Abstract

This paper suggests that log-linear analysis can be used to predict the expected number of patients in hospital services. A log-linear risk rate model has been developed to analyze the time series of count data. In finding a risk rate model, parameters are estimated and goodness-of-fit is utilized to carefully extract the best model to fit the count data. The marginal effect is the basis function which can be used in the risk Rate analysis for flexibility. This study attempted to analyze the actual operations of a hospital and proposed modifications in the system to reduce waiting times for the patients, which should lead to an improved view of the quality of services provided. To develop a risk Rate model for the above situation, it is necessary to define a model for the expected number of patients for hospital services cases. Here, 2 underlying variables are of interest, “waiting time” and “hospital services”. “Waiting time” has been categorized into 7 groups. The variable “hospital services” contains 4 categories: Inpatient (IN), Referred (REF), Recurring (RCR), and Observation (OBS). As a result, significant levels of causal variables are not expected to be identical for each model. It was found that the IN category has a higher rate of increase in the average waiting time. The marginal allows better predictions of hospital services and rehabilitation decision making.

**Keywords:** Hospital services, risk rate model, log-linear model

## Introduction

In the categorical data analysis literature (Heien *et al.*, 2004), the survival models and/or the risk rate models are treated differently from standard logit models. In general, these models are termed as rate models or risk rate models. In its simplest form, a rate is defined as the number of individuals or observations possessing a particular characteristic divided by the total amount of exposure to the risk of having such a characteristic. The risk rate models can easily be connected to the standard

Poisson models (Powers and Xie, 2000). Then the Poisson models are directly related to the exponential Models by making conversion of rates per unit interval with the waiting time until the first occurrence. Here a risk rate model is used to determine the likelihood of the demand for hospital services in a queuing system in order to identify factors associated with increased health care utilization, particularly those factors related to hospital services. This is a difficult task for several

---

*Information Technology Faculty, Rangsit University. E-mail: Kriengstat@yahoo.com*

\* *Corresponding author*

reasons.

The Admissions Department consists of 4 major areas: front desk, registration desk, waiting area, and financial consulting area (within the Business Department). When a patient enters the admissions area, they are asked by the front desk clerk to provide his or her name and the reason for the visit. The clerk also clarifies if the patient was pre-registered for this service or not. If the answer is yes, the clerk gets the patient’s documentation ready for the the admissions representative. Then the patient receives an assigned number and is asked to wait in the waiting area for the admissions representative to call the number. The admissions representative determines if the patient has ever received services at the hospital and, if so, pulls up the patient’s data from Meditech and verifies the patient’s personal information. If the patient is visiting the hospital for the first time, an admissions clerk creates the patient’s profile in the hospital’s information database system.

Law and Kelton (2001) proposed an algorithm of a successful computer simulation study. This algorithm includes the following key steps: 1) Problem formulation, 2) Data collection and the conceptual model design, 3) The validation of the model, 4) The constructions of the computer representation of the model, 5) The verification of the model, 6) The design of experiments needed to address theproblem, 7) Production runs using the computer model, 8) The statistical analysis of the data obtained from the production runs, and 9) The interpretation of the results.

First, even in the case of constant demand levels over the day, statistical fluctuations in individual patient waiting times and the variability in the time needed by a provider to service patients can create long delays even when overall average steady state capacity is greater than average demand. Second, the magnitude of delays is a log-linear function of the demand for hospital services and is thus impossible to predict without the use of a queueing model (Green and Nguyen, 2001).

## Methodology of Risk Rate Model

### Risk Rate Model Analysis

Let  $Y_i$  be the number of patients of an event of interest for the  $i^{th}$  subject and denote the independent variables by  $x_i, i = 1, \dots, n$ . We assume that  $Y_i$  follows a Poisson distribution,  $Y_i \sim \text{Poisson}(\lambda_i)$ , with density

$$f(y_i|x_i) = \frac{\lambda_i^{-y_i} e^{-\lambda_i}}{y_i!} \tag{1}$$

Let  $t_1, t_2, \dots, t_n$  be the waiting times of the  $n^{th}$  individual, and assume the distribution function to be  $F(t) = Pr(T < t)$  with the probability density function  $f(t)$ . The risk rate is denoted by  $\mu(t)$ , and can be viewed as the instantaneous probability of an event in the interval  $[t, t + \Delta t]$ , given the event has not occurred before time  $t$ . Formally, the risk rate (Manski and McFadden, 1981), is defined by the following limit:

$$\mu(t) = \lim_{\Delta t \rightarrow 0} \frac{1}{\Delta t} Pr[t \leq T < t + \Delta t | T \leq t] \tag{2}$$

The density of an exponential distribution with parameter  $\mu$  is given by

$$f(t) = \mu e^{-\mu t}, t > 0. \tag{3}$$

The distribution function equal

$$F(t) = 1 - \mu e^{-\mu t}, t \geq 0. \tag{4}$$

For this distribution, we have

$$E(x) = \frac{1}{\mu}, \sigma^2(x) = \frac{1}{\mu^2}, \tag{5}$$

The probability of an event not occurring up to time  $t$  is given by the function

$$P(t) = Pr[T > t] = 1 - F(t) = e^{-\mu t} \tag{6}$$

Assuming the waiting times are exponentially distributed, the equation (6) may be written as:

$$P(t) = e^{-\mu t} \tag{7}$$

The risk rate is defined by the ratio

$$\mu(t_i) = \frac{f(t)}{P(t)} = \frac{f(t)}{1-F(t)} = \frac{\mu e^{-\mu t}}{e^{-\mu t}} = \mu \tag{8}$$

The general risk rate model may be written as:

$$\mu(x_i^T \beta_i) = e^{(\beta_0 + \beta_1 x_{1i} + \dots + \beta_n x_{ni})} \tag{9}$$

where  $x^T = [1, x_1, x_2, \dots, x_n]$ , and  $\beta_0, \beta_1, \dots, \beta_n$  are unknown constants as the rate is determined by several regressors. This exponential risk rate model can be estimated using a risk rate models for counts. In a time interval of length  $t$ , the probability of  $y$  events is given by:

$$Pr(y | \mu, t) = \frac{(\mu t)^y e^{-\mu t}}{y!} \tag{10}$$

Because the mean number of events in the time interval  $t$  is  $\lambda = \mu t$ , for the  $i^{th}$  individual, the expected number of events in the time interval  $t_i$  is

$$\lambda_i = \mu_i t_i, \tag{11}$$

or

$$\lambda_i = t_i e^{(x_i^T \beta_i)} \tag{12}$$

Taking the log of the Poisson means results in the log-linear regression model:

$$\frac{\lambda_i}{t_i} = e^{(x_i^T \beta_i)} \tag{13}$$

or

$$\log\left(\frac{\lambda_i}{t_i}\right) = x_i^T \beta_i \tag{14}$$

namely,  $\log(\lambda_i) - \log(t_i) = x_i^T \beta_i$ . (15)

**Goodness-of-Fit**

The log-likelihood function of the process cannot be the only index of fit because the

likelihood ratio statistics are dependent on the size of the sample. Different values of the log-likelihood function result when competing models, namely models that differ in the number of parameters, are fitted to the same data. The number of parameters, in general, should be more than 1, and significantly less than the number of observations. To assess the model goodness-of-fit, we need to know how 1 model fits relative to another. An indicator of a model goodness-of-fit that measures the extent to which the current model deviates from a more generalized model is given by the likelihood ratio statistics:

$$G^2 = -2 \log \left( \frac{L_c}{L_f} \right) = -2(\log L_c - \log L_f), \tag{16}$$

where  $\log L_c$  is the log-likelihood of the current model, and  $L_f$  logis the log-likelihood of the more generalized model. The likelihood ratio statistics has a chi-square distribution with  $K_2 - K_1$  degrees of freedom, where  $K_2$  and  $K_1$  denote the number of parameters in the more generalized model and the current model, respectively (McCullagh and Nelder, 1989).

**Marginal Effects**

For risk rate models, the marginal effects can be thought of as the relative risk associated with a certain variable. The overall mean effect in (12) is

$$\lambda(x^T \beta) = e^{(x^T \beta)} \tag{17}$$

Then, the marginal effect due to the  $k^{th}$  factor can be considered as

$$\hat{\theta}_k = \bar{x}_k e^{(\bar{x}^T \hat{\beta})} \tag{18}$$

where  $\bar{x}_k$  is  $k^{th}$  the mean of the factor values in the sample and  $\bar{x}^T$  is the vector of the means of the factor values in the sample. An estimate of  $\theta_k$  can be computed as

$$\theta_k = e^{(\beta)} \tag{19}$$

## Model Results

### Estimation Results and Interpretation

The following data was obtained from Lop Buri Hospital in Thailand. The Admissions Department is one of the most highly congested hospital services, and faces a great deal of pressure, compared with other components of the health care system. An admissions clerk determines the patient's type (Inpatient (IN), Referred (REF), Recurring (RCR), and Observation (OBS)) and creates a new account using the hospital's informational system. Admissions serve most outpatient and inpatient types, with an exception for all REF and some RCR, OBS, and IN. It is essential to assess the relationship between hospital services and the average waiting time in a queuing system. Using SAS and MATLAB to perform the iterations necessary for the maximum likelihood method, the following results have been obtained.

To develop a risk rate model for the above situation (Horvath, 1999), it is necessary to define a model for the expected number of patients for hospital services cases,  $E(Y_{ij})$ , in terms of the variables of interest. Here, 2 underlying variables are of interest, "waiting time" and "hospital services". Since "waiting time" has been categorized into 7 groups, 6

dummy variables are used to index them. The variable "hospital services", which contains 4 categories, requires only 3 dummy variables. Thus, 1 possible model for the expected number of patients for hospital services cases in the  $(i,j)^{th}$  group can be written as:

$$E(Y_{ij}) = \mu_{ij} = \eta_{ij} \lambda_{ij}, \text{ where } \log \lambda_{ij} = \alpha + \beta E, i = 1, 2, \dots, n, j = 1, 2, \dots, m.$$

Using this model, the risks  $\lambda_{ij}$  in terms of the parameters  $\alpha_i$  and  $\beta$  can be written to obtain

$$\begin{aligned} \log \lambda_{i0} &= \alpha + \alpha_i \text{ and } \log \lambda_{i1} = \alpha + \alpha_i + \beta, \\ \text{since } \log \lambda_{i1} - \log \lambda_{i0} &= (\alpha + \alpha_i + \beta - \alpha - \alpha_i) \\ \log \lambda_{i1} - \log \lambda_{i0} &= \theta_k, \text{ so } \theta_k = e^{(\beta)} \end{aligned}$$

The data set in Table 1 was used to estimate the average waiting time and hospital services for medical and health services on the risk rate analysis with the hospital services variables characterized by the marginal effect ( $\theta_k = e^{(\beta)}$ ). Two separate models were specified and estimated for each state since the various causal factors vary over time; as a result, significant levels of causal variables are not expected to be identical for each model. The model is to compare the average waiting time ( $t$ ) and hospital services ( $h$ ). A chi-square test

**Table 1. Contingency table of the hospital services**

Average waiting time		Hospital Services				Total
		REF ( $h_1$ )	RCR ( $h_2$ )	OBS ( $h_3$ )	IN ( $h_4$ )	
1-20	( $t_1$ )			1	3	4
21-40	( $t_2$ )		6	7	33	46
41-60	( $t_3$ )	10	12	9	31	62
61-80	( $t_4$ )		8		1	9
81-100	( $t_5$ )	33	7	5		45
101-120	( $t_6$ )				30	30
> 120	( $t_7$ )	4				4
<b>Total</b>		<b>47</b>	<b>33</b>	<b>22</b>	<b>98</b>	<b>200</b>

relation log-linear model and risk rate model were also computed:

Saturated log-linear model:

$$\log \lambda = \mu + \mu_i t_i + \mu_j h_j + \mu_{ij} t_i h_j \quad (20)$$

Risk rate model:

$$\log \lambda_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3ij} \quad i=1,2,\dots,n; j=1,2,\dots,m \quad (21)$$

where  $x_{1i}$ ,  $x_{2i}$ ,  $x_{3ij}$  are dummy variables and the interaction variable  $x_{3ij} = x_{1i} * x_{2i}$  is and the variable  $x_1$  corresponds to the REF cases,  $x_2$  corresponds to the RCR cases,  $x_3$  corresponds to the OBS cases,  $x_4$  corresponds to the IN cases,  $x_5$  corresponds to the group with an average waiting time of 1-20 min,  $x_6$  corresponds to the group with an average waiting time of 21-40 min,  $x_7$  corresponds to the group with an average waiting time of 41-60 min,  $x_8$  corresponds to the group with an average waiting time of 61-80 min,  $x_9$  corresponds to the group of with an average waiting time of 81-100 min,  $x_{10}$  corresponds to the group with an average waiting time of 101-120 min,  $x_{11}$  corresponds to the group with an average waiting time > 120 min, and  $x_{12} = x_1 x_5$ ,  $x_{13} = x_2 x_5$ ,  $x_{14} = x_3 x_5, \dots, x_{39} = x_4 x_{11}$ .

A saturated log-linear model rate was defined. The deviations between the maximized log-likelihood from each model were used to perform a series of chi-square tests in order to ascertain which model gives the best fit.

So, the saturated log-linear model is

$$\log(\lambda) = \mu + \sum_{i=1}^n \mu_i t_i + \sum_{j=1}^m \mu_j h_j + \sum_{i=1}^n \sum_{j=1}^m \mu_{ij} t_i h_j \quad (22)$$

The main effects model of factors t and h is

$$\log(\lambda) = \mu + \sum_{i=1}^n \mu_i t_i + \sum_{j=1}^m \mu_j h_j \quad (23)$$

The main effects model of factor t is

$$\log(\lambda) = \mu + \sum_{i=1}^n \mu_i t_i \quad (24)$$

The main effects model of factor h is

$$\log(\lambda) = \mu + \sum_{j=1}^m \mu_j h_j \quad (25)$$

Risk rate model:

So, the saturated log-linear model is

$$\log(\lambda) = \beta_0 + \sum_{i=1}^{28} \beta_i x_i \quad (26)$$

The main effects model of factors t and h is

$$\log(\lambda) = \beta_0 + \sum_{i=1}^{11} \beta_i x_i \quad (27)$$

The main effects model of factor t is

$$\log(\lambda) = \beta_0 + \sum_{i=1}^7 \beta_i x_i \quad (28)$$

The main effects model of factor h is

$$\log(\lambda) = \beta_0 + \sum_{i=1}^4 \beta_i x_i \quad (29)$$

The following results were obtained.

The saturated log-linear model yields

$$\begin{aligned} \log(\lambda) = & -14.8949 + 16.2841h_1 - 0.6423h_2 - \\ & 1.1326h_3 + 16.1394t_1 + 18.406t_2 + \\ & 18.8536t_3 + 17.6167t_4 + 18.10t_5 + \\ & 18.2963t_6 - 32h_1t_1 - 16h_2t_1 - \\ & 0.0365h_3t_1 - 32h_1t_2 - 0.0718h_2t_2 - \\ & 0.282h_3t_2 - 15.9804h_1t_3 - 0.5691h_2t_3 - \\ & 0.4642h_3t_3 - 32h_1t_4 - 16h_3t_4 - 16h_1t_5 + \\ & 32h_1t_6 - 16h_2t_6 - 16h_3t_6, \end{aligned} \quad (30)$$

$$\log\text{-likelihood} = 383.3958, \quad df = 26$$

The main effects model of factors h and t yields

$$\begin{aligned} \log(\lambda) = & 0.9728 + 0.04h_1 - 0.962h_2 - 1.559h_3 - \\ & 0.0957t_1 + 2.588t_2 + 2.9627t_3 + \\ & 1.6526t_4 + 2.5138t_5 + 2.4172t_5, \end{aligned} \quad (31)$$

$$\log\text{-likelihood} = 375.2952, \quad \text{chi-square} = 10.857, \quad df = 17$$

The main effects model of factor  $t$  yields

$$\log(\lambda) = 0.05539 - 0.1605t_1 + 2.481t_2 + 2.7188t_3 + 1.2t_4 + 12.4825t_5 + 2.8036t_6, \quad (32)$$

log-likelihood = 341.413, chi-square = 80.413, df = 20

The main effects model of factor  $h$  yields

$$\log(\lambda) = 3.2908 - 0.4153h_1 - 1.1057h_3 - 1.6614h_3, \quad (33)$$

log-likelihood = 307.565, chi-square = 125.923, df = 23

**Goodness-of-Fit**

Using these results, the competing models using the likelihood-ratio statistics as described in section 2.2 were tested in order to determine the goodness-of-fit. To perform the tests, the saturated model in (30) was tested against the main factors model in (31), and then the main factors model was tested against its nested counterparts. The results of the chi-square tests, performed with  $\alpha = 0.05$ , are as follows:

The main factors model in (31) compared with the saturated model (with all the interactions) in (30) is an adequate fit model. The model in (31) has an adequate fit compared with all the models. The main factors of the welfare model in (33) compared with all the models in (31). The model in (33) has an adequate fit compared with the welfare model. Thus, it was decided to choose the main factors model in (33) as the adequate model for this data set.

The main effects model is

$$\log(\lambda) = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_4$$

$$\log(\lambda) = 3.2908 - 0.4153h_1 - 1.105h_2 - 1.6614h_3 \quad (34)$$

Tables 2 and 3 show that the parameter estimates of hospital services are significant at the 5% level. The results indicate that better

predictions of health services can be made. As for the hospital services class, the categories are REF, RCR, OBS, and IN. The marginal effects are computed as described in Section 2.3. The marginal effect for the first factor, the REF group, is calculated as  $\hat{\theta}_1 = \exp(\hat{\beta}_1) = 0.66014$ . This means, the target population of the REF group is 0.66014 times. The marginal effect for the second factor, the RCR group, is calculated as  $\hat{\theta}_2 = \exp(\hat{\beta}_2) = 0.33098$ . This means the target population of the RCR group is 0.33098 times. The marginal effect for the third factor, the OBS, is calculated as  $\hat{\theta}_3 = \exp(\hat{\beta}_3) = 0.18987$ . This means the target population of the OBS group is 0.18987 times. The marginal effect for the fourth factor, the IN group, is calculated as  $\hat{\theta}_4 = \exp(\hat{\beta}_4) = 1$ . This means the target population of the IN group is 1 times.

**Conclusions**

This study attempted to analyze the actual operations of a hospital and proposed modifications in the system to reduce waiting times for the patients, which should lead to an improved view of the quality of service provided.

As a result, significant levels of causal variables are not expected to be identical for each model. The model compares the average waiting time ( $t$ ) and hospital services ( $h$ ). A chi-square test relation log-linear model and risk rate model were also computed.

**Saturated log-linear model:**

The saturated log-linear model yields

$$\log(\lambda) = -14.8949 + 16.2841h_1 - 0.6423h_2 - 1.1326h_3 + 16.1394t_1 + 18.406t_2 + 18.8536t_3 + 17.6167t_4 + 18.10t_5 + 18.2963t_6 - 32h_1t_1 - 16h_2t_1 - 0.0365h_3t_1 - 32h_1t_2 - 0.0718h_2t_2 - 0.282h_3t_2 - 15.9804h_1t_3 - 0.5691h_2t_3 - 0.4642h_3t_3 - 32h_1t_4 - 16h_3t_4 - 16h_1t_5 + 32h_1t_6 - 16h_2t_6 - 16h_3t_6, \quad (1)$$

log-likelihood = 383.3958, df = 26

The main effects model of factors of the hospital services (*h*) and waiting time (*t*) yields

$$\log(\lambda) = 0.9728+0.04h_1-0.962h_2-1.559h_3-0.0957t_1+ 2.588t_2+2.9627t_3+1.6526t_4+2.5138t_5+2.4172t_6, \quad (2)$$

log-likelihood = 375.2952, chi-square = 10.857, df = 17

The main effects model of factor of waiting time (*t*) yields

$$\log(\lambda) = 0.05539-0.1605t_1+2.481t_2+2.7188t_3+1.2t_4 +12.4825t_5+2.8036t_6, \quad (3)$$

log-likelihood = 341.413, Chi-square = 80.413, df = 20

The main effects model of the factor of hospital services (*h*) yields

$$\log(\lambda) = 3.2908-0.4153h_1-1.1057h_2-1.6614h_3, \quad (4)$$

log-likelihood = 307.565, chi-square = 125.923, df = 23

Using these results, the competing models were tested using the likelihood ratio statistics as described in order to determine the goodness-of-fit. The results of the chi-square tests, performed with  $\alpha = 0.05$ , are as follows:

The main effects model of hospital services (*h*):

$$\log(\lambda) = 3.2908-0.4153h_1-1.105h_2-1.6614h_3 \quad (5)$$

It was found that the IN category has a higher rate of increase in the average waiting time. The marginal effect ( $\hat{\theta}_k$ ) is a basis function that can be used in the risk rate analysis for flexibility. In this paper, an easily implemented estimation procedure for the coefficients of hospital services in the queuing system has been described. The estimation approach is based on using a series of prediction probabilities. Furthermore, the methodology for determining the prediction probabilities from the waiting time model is developed. Finally, testing for statistical significance was carried out and the risk rate function was found to be increasing.

**Table 2. Test goodness-of-fit model of hospital services**

Test	G <sup>2</sup>	df
(31) vs (30)	-2[(375.2952)-(383.3958)]=16.2012	9
(32) vs (31)	-2[(341.4130)-(375.2952)]=67.7644	3
(33) vs (31)	-2[(307.5654)-(375.2952)]=135.4596	3

**Table 3. Mean marginal effect of hospital services**

Variable	Marginal effect
Referred (REF) ( <i>x</i> <sub>1</sub> )	e <sup>-0.4153</sup> = 0.66014
Recurring (RCR) ( <i>x</i> <sub>2</sub> )	e <sup>-1.1057</sup> = 0.33098
Observation (OBS) ( <i>x</i> <sub>3</sub> )	e <sup>-1.6614</sup> = 0.18987
Inpatient (IN) ( <i>x</i> <sub>4</sub> )	e <sup>0</sup> = 1

**References**

- Green, L.V. and Nguyen, V. (2001). Strategies for cutting hospital beds: the impact on patient service. *Health Serv.Rep.*, 36:421-442.
- Heien, H.C., Baumann, W.A., and Rahman, M. (2004). Inference in Log-Rate Models. *Journal of Undergraduate Research*, Minnesota State University, Mankato, MN, USA, Vol. 3.
- Horvath, G. (1999). Approximate waiting time analysis of priority queues. In: *Quality Management in Social Welfare and Health Care for the 21<sup>st</sup> Century; National Recommendation*. Outinen, M and Lindqvist, T., (eds). STAKES National Research and Development Centre for Welfare and Health, Helsinki, Finland, 45p.
- Law, A.M. and Kelton, W.D. (2001). *Simulation Modeling and Analysis*, 3<sup>rd</sup> ed. McGraw-Hill, New York, NY, USA, 759p.
- McCullagh, P. and Nelder, J.P. (1989). *Generalized Linear Models*. 2<sup>nd</sup> ed. Chapman and Hall/CRC, Boca Raton, FA, USA, 532p.
- Manski, C. and McFadden, D. (eds) (1981). *Structural Analysis of Discrete Data with Econometric Applications*. The MIT Press, Cambridge, Mass, USA, 504p.
- Powers, D. and Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis*. Academic Press, San Diego, CA, USA, 308p.