Identifying Prognosis Factors of Pathological Staging Among Colorectal Patients Using Ordinal Logistic Regression Model

NOR AZURA MD GHANI[1], KHAIRUL ASRI MOHD GHANI[2], ZAMALIA MAHMUD[1], NURHASNIZA IDHAM ABU HASSAN[3] & NORAZAN MOHAMED RAMLI[1]
[1]Center for Statistical Studies and Decision Sciences
Faculty of Computer & Mathematical Sciences
Universiti Teknologi MARA
40450 Shah Alam
MALAYSIA
[2]Department of Surgery
Faculty of Medicine & Health Sciences
Universiti Putra Malaysia
43400 Serdang, Selangor
MALAYSIA
[3]Center for Statistical Studies and Decision Science
Faculty of Computer & Mathematical Sciences
Universiti Teknologi MARA
32610 Bandar Baru Seri Iskandar, Perak
MALAYSIA
{azura, zamalia, norazan}@tmsk.uitm.edu.my, k_asri@medic.upm.edu.my, nurha098@perak.uim.edu.my

Abstract: - Colorectal cancer is known as one of the cancer disease that is often related to dietary habits, age, sex, and family history. Laparoscopic resection is one of the recent techniques used to treat colorectal cancer patients. The main objective of this paper is to model the success of laparoscopic resection of colorectal cancer patients at various operative stages using ordinal logistic regression. One hundred patients who underwent laparoscopic resection for colorectal cancer were analyzed. All patients were operated on by 3 surgeons at Hospital Kuala Lumpur tertiary referral center using standardized techniques and care plans assessed for operative indications. Results indicate that the prognosis factors that can explain the pathological staging of colorectal cancer were adjuvant therapy, metastasis recurrence and tumor thickness level. Pathologist may use these findings to propose guidelines for appropriate treatment plan for a particular patient according to their staging.

Key-Words: - Colorectal cancer, laparoscopic resection, prognosis factors, pathological staging.

1 Introduction

According to a recent study, colorectal cancer is known to be the cancer disease that is related with the changing of dietary habits and lifestyle factor [1]. As reported in the Second National Cancer Registry [2], colorectal cancer was among the top ten most common cancer diseases in Malaysia which comprises 14.2% males and 10.7% females. This makes it the commonest cancer among men and the third most common cancer among women. The male to female ratio for colon cancer was nearly equal (0.98 female: 1male), with the frequency in males rising more rapidly after the age of 60 years. In this study our main objective is to model pathological staging for 100 patients of colorectal cancer. Patients were follow-up on a monthly basis for one year where their conditions were examined regularly after the surgery. Significant prognosis factors are expected to be identified from the proposed model.

2 Ordinal Logistic Regression Model

Ordinal logistic regression is a method of estimating the effect of predictor variables on order categorical variables [3]. From literature review, no study had been done using ordinal logistic regression to identify prognosis factors of pathological staging for

patients of colorectal cancer. The ordinal logistic regression was used in this study to model the relationship between response variable, which represents the four different levels of pathological staging and the four major predictor variables namely as factor 1: demographic factors (gender, age group, race, BMI group); factor 2: pre-operative (neoadjuvant therapy, ASA score, abdominal history); factor 3: intra-operative (type of resection, duration of resection, anastomic bleeding, length of hospital stay, resumption of normal diet) and factor 4: post-operative (adhesive obstruction, deep vein thrombosis, relaparotomy, miscroscopic resection margin, distal margin, circumferential resection margin, positive lymph nodes, negative lymph nodes, metastasis recurrence, adjuvant therapy, tumour thickness level, conversion to open resection).

The response variable for pathological staging was measured on an ordered, categorical, based on four point scale namely 'cancer stage I', 'cancer stage II', 'cancer stage III', and 'cancer stage IV'. Definition of four different categories in pathological staging is as follows [4].

i) Stage I - Cancer has begun to spread, but is still in the inner lining.
ii) Stage II - Cancer has spread to other organs near the colon or rectum. It has not reached lymph nodes.
iii) Stage III - Cancer has spread to lymph nodes, but has not been carried to distant parts of the body.
iv) Stage IV- Cancer has been carried through the lymph system to distant parts of the body.

The ordinal logistic regression model used in this study is as shown in equation (1) below [5] [6] [7].

$$y^* = \alpha + \sum_{k=1}^{K} \beta_k x_k + \varepsilon \quad (1)$$

where y^* is unobserved and thus can be thought of as the underlying tendency of an observed phenomenon, ε is assumed to follow a certain symmetric distribution with zero mean such as standard normal distribution and a logistic distribution with the following conditions:

$$y = 1 \; if \; y^* \leq \mu_1 = 0,$$
$$y = 2 \; if \; \mu_1 \leq y^* \leq \mu_2$$
$$y = 3 \; if \; \mu_2 \leq y^* \leq \mu_3$$
$$\vdots$$
$$y = J \; if \; y^* > \mu_{J-1}$$

where y is observed in J number of ordered categories and the μs are unknown threshold parameters separating the adjacent categories.

In general,

$$\begin{aligned} &\mathrm{P}(y \leq j|\mathbf{x}) \\ &= \mathrm{P}(y^* \leq \mu_j) \\ &= \mathrm{P}(\alpha + \sum_{k=1}^{K} \beta_k x_k + \varepsilon \leq \mu_j \\ &= \mathrm{P}(\varepsilon \leq \mu_j - \left(\alpha + \sum_{k=1}^{K} \beta_k x_k\right) \\ &= \mathrm{F}\left[\mu_j - \left(\alpha + \textstyle\sum_{k=1}^{K} \beta_k x_k\right)\right] \end{aligned} \quad (2)$$

where F(.) is the cumulative distribution function of ε. If ε follows a standard normal distribution, then an ordinal probit regression model is obtained as in equation (3).

$$\mathrm{P}(y \leq j|\,\mathbf{x}) = \Phi\left[\mu_j - \left(\alpha + \textstyle\sum_{k=1}^{K} \beta_k x_k\right)\right] \quad (3)$$

where $\Phi(.)$ is the cumulative distribution function of standard normal distribution. If ε follows a logistic distribution, we have the ordinal logistic regression model:

$$\mathrm{P}(y \leq j|\,\mathbf{x}) = \frac{\exp[\mu_j - (\alpha + \sum_{k=1}^{K} \beta_k x_k)]}{1 + \exp[\mu_j - (\alpha + \sum_{k=1}^{K} \beta_k x_k)]} \quad (4)$$

3 Results and Discussion

Before running the ordinal logistic regression, it is required to choose link function to produce the models that provide good fit to the data [8]. To choose a link function, it is helpful to examine the distribution of values for the outcome variable by using bar chart. Based on Figure 1, probit link function will be considered as the appropriate function due to the distribution for pathological staging was normally distributed.

The completed model with the probit link in Table 1 for probit shows that, pathological staging are significantly associated with three clinical variables of tumour thickness level, adjuvant therapy, and metastasis recurrence. These significant explanatory variable exhibited positive regression coefficients, except tumour thickness at level 2 (T2).

Table 2 shows the model fitting statistics for the observed reduced model using probit link which indicates that, the -2LL of the model with only intercept was 206.277 while, -2LL of the model with intercept and three independent variables were 0.000. That is the difference (Chi-square statistics) was 206.277 (206.277-0.000) which is significant at 5%. It can be concluded that there was association between pathological staging and independent clinical variables of adjuvant therapy, tumour thickness level, and metastasis recurrence.

Table 3 displays the Goodness of Fit statistics for reduced model with probit link. The additional model fitting statistic, the Deviance= 11.559 (with degree freedom of 14 and p-value= 0.642) for the model with the probit link which indicate that, the observed data were consistent with the estimated values in the fitted model. It means that the model

with probit link fits the data well so goodness of fit statistics suggests that model-predicted cell proportions are acceptably close to the observed proportions.

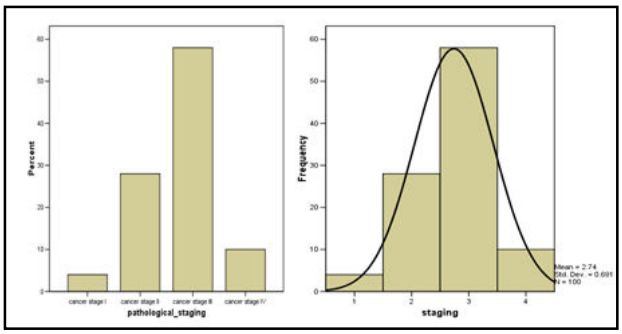

Fig. 1: The bar chart and distribution of pathological staging

Table 1: Explanatory Variables Based on Completed Model Using Probit Link

		Estimate	*p*-value
Threshold	[staging = 1]	-1.718	0.494
	[staging = 2]	1.680	0.504
	[staging = 3]	5.644	0.028
Location	Duration of resection	0.000	1.000
	Length of hospital stay	0.000	1.000
	Resumption of normal diet	0.000	1.000
	Distal margin	0.000	1.000
	CRM	0.000	1.000
	Positive lymph node	0.000	1.000
	Negative lymph node	0.000	1.000
	[Gender=Male]	0.000	1.000
	[Age group=<60 years old]	0.000	1.000
	[Abdominal history=Yes]	0.000	1.000
	[Anastomic bleading=Yes]	0.000	1.000
	[Type of operation=LAR]	0.000	1.000
	[Conversion to open =Yes]	0.000	1.000
	[MRM=Yes]	0.000	1.000
	[BMI group=Non obese]	0.000	1.000
	[Neoadjuvant therapy=Yes]	0.000	1.000
	[Adhesive obstruction=Yes]	0.000	1.000
	[Deep vein thrombosis= Yes]	0.000	1.000
	[Relapatomy=Yes]	0.000	1.000
	[Metastasis recurrence=Yes]	4.074	0.001*
	[Adjuvant therapy=Yes]	3.597	0.024*
	[ASA Score=Level I]	0.000	1.000
	[ASA Score=Level II]	0.000	1.000
	[Tumour thickness=T2]	-3.714	0.040*
	[Tumour thickness=T3]	0.000	1.000
	[Race=Malay]	0.000	1.000
	[Race=Chinese]	0.000	1.000

**Association is significant at the 0.05 significance level*

Table 2: Reduced Model Fitting Using Probit Link

Probit link				
	-2LL	χ^2	df	*p*-value
Intercept only	206.277			
Final	0.000	206.277	4	0.000

**significance at 0.05*

Table 3: Goodness-of-Fit for Reduced Using Probit Link

Probit link			
	χ^2	df	*p*-value.
Deviance	11.559	14	0.642

**significance at 0.05*

The model-fitting statistics, namely the Pseudo R square as shown in Table 4, measured the success of the model in explaining the variations in the data. The larger the Pseudo R square was, the better the model fitting was. The Pseudo R squares for Cox & Snell (0.873), McFadden (1.000), and Nagelkerke (1.000) in the model with probit link. It seems that the pathological staging explaines 87.3% of the variance in three clinical independent variables included in the reduced model according to Cox & Snell R square value, 100% according to both McFadden, and Nagelkerke value for probit link.

Table 4: Pseudo R-Square Using Probit Link

Pseudo R-square	*p*-value
Cox & Snell	0.873
Nagelkerke	1.000
McFadden	1.000

The classification table was used to categorize the classified and the actual response. Table 5 displays the accuracy of the classification results for the pathological staging response categories using probit link. The model with the probit link classified categories of “staging I" (4), "staging II" (28), "staging III" (58) and "staging IV" (10). The model demonstrated the perfect prediction of pathological staging with the accuracy of 100% when the three prognosis factor of adjuvant therapy, tumour thickness level, and metastasis recurrence included in the model.

The test for each of parameter estimates are displayed in Table 6. Using the model with the probit link, the pathological staging was found to be

ISBN: 978-1-61804-106-7

significantly associated at 5% with *p*-value of 0.013, 0.000 and 0.000 with three explanatory variables of tumour thickness level, metastasis recurrence and adjuvant therapy respectively. The sign of parameter estimate, which measures the relationship between the variables and the probability of having pathological staging stage I, II, III, and IV, are coherent for all the significant variables.

Table 5: Classification Result for Pathological Staging Based on Probit Link

		Predicted Group				
		stage I	stage II	stage III	stage IV	Total
Actual Group	stage I	**4 (100%)**	0 (0)	0 (0)	0 (0)	4
	stage II	0 (0)	**28 (100%)**	0 (0)	0 (0)	28
	stage III	0 (0)	0 (0)	**58 (100%)**	0 (0)	58
	stage IV	0 (0)	0 (0)	0 (0)	**10 (100%)**	10

Table 6: Parameter Estimates and Test Statistics

	Estimate	Standard error	Wald	df	*p*-value
Threshold (staging=1)	-1.718	0.581	8.743	1	0.003
Threshold (staging=2)	1.680	0.573	8.591	1	0.003
Threshold (staging=3)	5.644	0.761	54.957	1	0.000
Adjuvant therapy (yes)	3.597	0.547	43.279	1	0.000*
Adjuvant therapy (no)	0 (a)			0	
Tumour thickness (T2)	-3.714	1.494	6.182	1	0.013*
Tumour thickness (T3)	0.000	0.450	0.000	1	1.000*
Tumour thickness (T4)	0 (a)			0	
Metastasis recurrence (yes)	4.074	0.986	17.061	1	0.000*
Metastasis recurrence (no)	0 (a)			0	

(a)This parameter is set to be zero because redundant
**Significance at 0.05*

The observed positive signs of the estimated parameters for metastasis recurrence, tumour thickness level of T3 and adjuvant therapy and negative sign for tumour thickness at level 2 (T2). Patients who need of receiving adjuvant therapy were ($e^{3.597}$) =36 times more likely to be the odds of having serious problem of cancer (higher stage of pathological staging) to less serious problem of cancer (lowest stage of pathological staging). Besides that, tumour thickness level 2 (T2), have $e^{-3.714}$ =0.002 times less likely invasion among patients who had serious problem of cancer (higher stage of pathological staging) to less serious problem of cancer (lowest stage of pathological staging).

4 Conclusion

Staging is the process of finding out how far the cancer has spread. This is very important because the treatment and the outlook for recovery depend on the cancer stage. These finding clearly justify that ordinal logistic regression models are appropriate to find the prognosis factors that can affect the pathological staging of colorectal cancer. Adjuvant therapy, metastasis recurrence and tumour thickness level were found to be significant prognosis factors for determining the pathological staging. Pathologist may use these findings to propose guidelines and consequently propose appropriate treatment plan for a particular patient according to their cancer staging.

Acknowledgments

We would like to devote our appreciation and gratitude to Hospital Kuala Lumpur (HKL) for the cooperation and assistance given throughout this research work. Special thanks also go to the Ministry of Higher Education of Malaysia and Universiti Teknologi MARA Malaysia for supporting this research with the Research University Grant No. 600-RMI/ST/FRGS 5/3/Fst (191/2010).

References:

[1] J.J.Y. Sung, J.Y.W. Lau, K.L. Goh & W.K. Leung, Increasing Incidence of Colorectal Cancer in Asia: Implications for Screening, *Lancet Oncol,* Vol.6. No.11, 2005, pp. 871-876.

[2] G.L.C. Chye & H. Yahaya, Second Report of the National Cancer Registry Cancer Incidence in Malaysia, *National Cancer Registry*, Ministry of Health Malaysia, 2003.

[3] B. Peterson & F.E. Harrell, Partial Proportional Odds Model for Ordinal Response Variables, *Applied Statistics*, Vol. 39, 1990, pp. 205-217.

[4] Colorectal Cancer Staging http://www.healthcommunities.com/colon-cancer/staging.shtml [23 April 2012]
[5] M.H. Kutner, C.J. Nachtsheim & J. Neter, *Applied Linear Regression Models*,4th Edition, New York: McGraw-Hill, 2004.
[6] H.E. Zhen & W.U. Du, A Comparative Study of Ordinal Probit and Logistic Regression for Affective Product Design, *Advanced Materials Research*, Vol. 452-453, 2012, pp. 642-647.
[7] R. Bender & U. Grouven, Ordinal Logistic Regression in Medical Research, *Journal of the Royal College of Physicians of London*, Vol. 31, No. 5, 1997, pp. 546- 551.
[8] Y.H. Chan, Basic Statistics or Doctors: Multinomial Logistic Regression, *Singapore Med J*, Vol. 46. No 6, 2005, pp. 259- 268.