Comparison between Probit Link and Cloglog Link Functions for the Identification of Pathological Staging in Colorectal Cancer Data

Nor Azura Md.Ghani, Khairul Asri Mohd Ghani, Zamalia Mahmud, Nurhasniza I.A. Hasan and Norazan Mohamed Ramli

***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 study is to effectively model the success of pathological staging of colorectal cancer patients using two ordinal regression link functions, i.e. probit link and cloglog link. Medical records of 100 patients who underwent laparoscopic resection for colorectal cancer were collected and analyzed. All patients were operated on by three surgeons at General Hospital Kuala Lumpur tertiary referral center using standardized techniques and care plans assessed for operative indications. Results indicate that probit link function has effectively explained the prognosis factors that lead to identification of pathological staging of colorectal cancer. The factors 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.

***Keywords*—** Probit link, cloglog link, colorectal cancer, laparoscopic resection, pathological staging

I. Introduction

IN today's global and competitive economic climate, the importance of both good health and healthcare services especially among workforce has become essential for continuation of socio-economic development.

Dr Nor Azura Md Ghani. is a lecturer at the Department of Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA. (Tel:+60355435371; fax:+60355455501; e-mail:azura@tmsk.uitm.edu.my)

Dr Khairul Asri Mohd Ghani is a urology surgeon at General Hospital. Kuala Lumpur. He is also attached with the Department of Surgery, Faculty of Medicine & Health Sciences, Universiti Putra Malaysia, 43400 Serdang, Selangor, MALAYSIA. (e-mail: k_asri@medic.upm.edu.my)

Dr Zamalia Mahmud is an Assoc. Prof. at the Department of Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA (e-mail:zamalia@tmsk.uitm.edu.my)

Nurhasniza I.A. Hassan is a research assistant at Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA (e-mail: haz_iza@yahoo.com.my)

Dr Norazan Mohamed Ramli a senior lecturer at the Department of Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor, MALAYSIA (e-mail: norazan@tmsk.uitm.edu.my)

As society becoming more advance, it changes the people lifestyles and cultural habits. In a negative point of view, unhealthy lifestyle is increasing amongst people who adopt western diets on daily basis which lead to the rising rate of cancer incidence. 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 old.

For colorectal cancer the preponderance of males was more noticeable with ratio of 1.26 to 1, where a steeper rise in age specific incidence of males occurring at age of 50 years onwards. Ethnic variation is also observed where the disease being most common among the Chinese population. Chinese community recorded the highest incidence rate of rectal cancers according to gender which is 2.8 times higher among the males, and 3.4 times higher among the females. Therefore, the purpose of the study is to investigate the prevalence of colorectal cancer among patients of different ethnic groups. This shall be done through description of patients' profile at various stages of cancer and who have also underwent laparoscopic colorectal surgery

TABLE 1
POPULATION IN MALAYSIA 2010 ACCORDING TO ETHNICITY

Ethnic group	Estimated Population of Malaysia (%)	Patients Diagnosed with colorectal cancer (%)	Difference (%)
Malay	53.3	44.0	-9.3
Chinese	26.0	38.0	12.0
Indian	11.8	18.0	10.3
Indigenous	7.7	nil	nil
Other	1.2	nil	nil

Table 1 shows the composition of population and incidence of colorectal cancer according to ethnic groups. Out of the 100 patients who have been diagnosed with colorectal cancer, 44.0% were Malay followed by 38% Chinese and 18.0% Indian. In this sample, there was no record of colorectal cancer incidence among the indigenous group and other ethnic

groups. Based on the estimated population of Malaysia, colorectal cancer incidence among the Malay population is still considerably lower by 9.3%. However, for the Chinese and Indian populations the incidence of colorectal cancer is higher at 12.0% and 10.3%, respectively.

In this study, the main objective is to effectively model the success of laparoscopic resection of colorectal cancer patients at various operative stages using two ordinal regression link functions, i.e. probit link and cloglog link which lead to identification of pathological staging of 100 patients with 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.

A. Risk Factors of Colorectal Cancer

B. Everyone cannot be avoided from being infected by colorectal cancer. Although 20% to 25% cases occur among individuals with a family history of colorectal cancer and about 75% of cases occur in people without these risk factors [3]. There are several factors that were expected to lead to the increment of this type of cancer. Thees factors include age, sex, overweight and obesity, diet, physical inactivity, smoking, alcohol intake, medications or dietary supplements, and family history.

- Age: 90% cases of colorectal cancer occur after the age of 50 for which the risk increases significantly with the advanced age. The incidence of colorectal cancer is 14 times higher in the older (> 50 years) than for those who are younger (<50years) [6].
- Sex: the incidence and mortality rates are 35% higher in men than women due to higher frequency of abdominal obesity, smoking, and drinking in men as well as hormone differences.
- Overweight and obesity: overweight and obesity increase the risk of colorectal cancer even though when the physical activity is accounted [4],[5] with stronger associations in men than women [7].
- Diet: diet with high amount of red, and meat [7],[8] low in calcium or [9],[10], and low intake for fruit and vegetable [11],[12] is associated with the increase incidence of colorectal cancer.
- Physical inactivates: regular physical activity is associated with lower risk of colon cancer [13],[14].
- Smoking: smoking cigarettes lead to increases the risk of developing adenomatous polyps [15] more cigarettes smoked increases the risk of having rectal cancer than colon cancer [16].
- Alcohol intake: high alcohol consumption is associated with an increased risk of colorectal cancer, in particular beer consumption [17].
- Family history: For those people, whose family members have colorectal cancer especially at a young age has higher risk of developing this cancer [18],[19]. On the other hand, for those people who have a first-degree relative (parent, sibling, or offspring) where about twice the risk of developing the disease compared to individuals with no family history [20],[21].

C. Colorectal Cancer Staging

The wall of colon and rectum is made up of several layers with starts in the innermost layer and can grow through some or all of the other layers. The stage of a colorectal cancer depends on how deeply it invades these layers. T stage is defined as layer penetration [22]. Assessment of depth of cancer invasion (T stage) remains the primary and most importance feature in treatment of patients with colorectal cancer.

The presence of lymph node involvement is in two circumstances: if local excision in the absence of lymphadenopathy is performed and if lymph node metastasis is shown outside the endopelvic envelope; in this case the tumor is considered locally advanced [22]. T classification measures how deeply the cancer cell invades into layers that form the wall of the colon and rectum. These layers, from the inner to the outer, include

- The inner lining (mucosa)
- A thin muscle layer (muscularis mucosa)
- The fibrous tissue beneath this muscle layer (submucosa)
- A thick muscle layer (muscularis propria) that contracts to force the contents of the intestines along
- The thin, outermost layers of connective tissue (subserosa and serosa) that cover most of the colon but not the rectum. The stage of a colorectal cancer depends on how deeply it invades these layers as seem in Fig. 1.

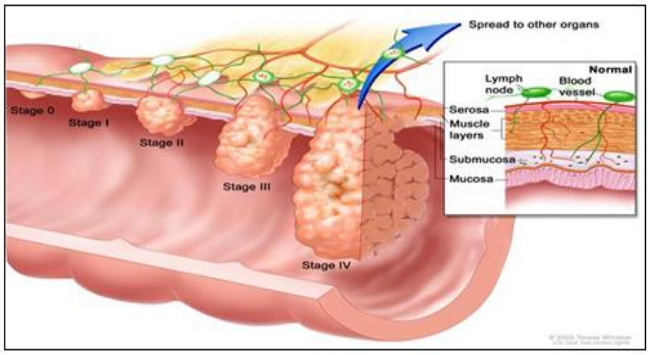

Fig. 1 Tumour staging

Fig. 2 shows the image of colon cancer with the help of colonoscopy. The most left image shows a normal colon lining; the middle image shows a polyp, and the most right image shows a malignant tumor.

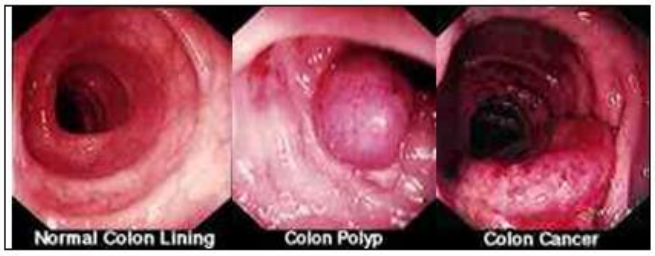

Fig. 2 The image of colon cancer

Low-grade cancers tend to grow and spread more slowly than high-grade cancers. Most of the time, the outlook is better for low-grade cancers than it is for high-grade cancers [23]. Staging needed among pathologist to determine the amount of cancer in the body and its location. This is important in identifying plan a treatment that may be suitable for a particular patient for which as cancer advanced in stage, more aggressive treatment may be necessary [24]. Cancers are classified at different stages, with depending on how advanced tumors progressed.

Staging also describes the severity of an individual's cancer based on the extent of the original (primary) tumor and the extent of spread in the body. For most cancers, the stage is based on three main factors:

- Location of the primary (original) tumor
- Tumor size and number of tumors
- Lymph node involvement (whether or not the cancer has spread to the nearby lymph nodes)
- Presence or absence of metastasis (whether or not the cancer has spread to distant areas of the body)

Each cancer type has its own classification system, so letters and numbers do not always mean the same thing for every kind of cancer. Stage I cancers are the least advanced and often have a better prognosis (outlook for survival). Higher stage cancers are often more advanced, but in many cases can still be treated successfully [25].

The stage of cancer is expressed in Roman numerals from stage I (the least advanced) to stage IV (the most advanced). The grades stage II and stage III fall somewhere in between. The grade is often simplified as either "low-grade" (stage I or stage II) or "high-grade" (stage III or stage IV). Low-grade cancers tend to grow and spread more slowly than high-grade cancers. In the early stages, this cancer is one of the most curable cancers whereas in the later stage it is the second most deadly [23].

II. Methodology

A. Data on Pathological Staging among Colorectal Patients

A total of 100 patients medical records who underwent laparoscopic resection for colorectal cancer between May 2005 and October 2009 were collected and analyzed. Patients were follow-up on a monthly basis for one year where their conditions were examined regularly after the surgery. Medical records have been regularly examined over the period of the study to prospectively update the database. Data collected cover the clinical information about colorectal patients at presentation, before (pre-operative), during (intra-operative) and after (post-operative) the operation. The clinical data and follow-up information that were collected from each colorectal patient include demographic factors, at presentation, pre-operative, intra-operative, and post-operative outcomes.

The results display the compositions of patients according to race with 44% Malay, 38% Chinese and 18% Indian. Patients who underwent emergency surgery were excluded. Laparoscopic resection that was excluded in presence of preoperative features at computed tomography (CT) scan suggest bulky tumors.

All patients were operated on by three surgeons at a single institution at General Hospital (tertiary referral center) using standardized techniques and care plans. Patients were assessed for operative indications, type of resection, operative time, conversion, intraoperative and postoperative outcome/complications. Post operative duration of stay, resumption of normal diet and surgical margins status, number of lymph node harvested were also reviewed.

TABLE 2
SUMMARY OF INDEPENDENT VARIABLES USED IN ORDINAL LOGISTIC REGRESSION

Independent variables		
Gender	Duration of resection	Conversion to open resection
Age group	Anastomotic Bleeding	Distal margin
Race	Length of hospital stay	CRM
Abdominal history	Resumption of normal diet	Metastatic recurrence
BMI group	Adhesive obstruction	Adjuvant theraphy
Neoadjuvant therapy	Deep vein thrombosis	Tumor thickness
ASA score	Relaparotomy	Positive lymph nodes
Type of resection	Miscroscopic resection margin	Negative lymph nodes

B. Ordinal Logistic Regression

Ordinal logistic regression link function was recognized as an appropriate technique for this study. It is used to estimate the effect of predictor variables on ordered categorical variables [26],[27],[28],[29],[30]. From literature review, no study had been done to identify prognosis factors of pathological staging for patients of colorectal cancer using ordinal logistic regression link functions. Probit link and cloglog link functions were used to model the relationship between the response variable, which represents four different levels of pathological staging and four major predictor

variables namely, 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 category 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 [31].

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 regression link function used in this study is as shown in equation (1) below [32], [33], [34], [35]

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

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:

$$y = 1 \text{ if } y^* \leq \mu_1 = 0,$$

$$y = 2 \text{ if } \mu_1 \leq y^* \leq \mu_2$$

$$y = 3 \text{ if } \mu_2 \leq y^* \leq \mu_3$$

$$\vdots$$

$$y = J \text{ if } y^* > \mu_{J-1}$$

where y is observed in J number of ordered categories and the μ s are unknown threshold parameters separating the adjacent categories. In general,

$$\mathrm{P}(y \leq j|\mathbf{x})$$

$$= \mathrm{P}(y^* \leq \mu_j)$$

$$= \mathrm{P}(\alpha + \sum_{k=1}^{K} \beta_k x_k + \varepsilon \leq \mu_j$$

$$= \mathrm{P}(\varepsilon \leq \mu_j - \left(\alpha + \sum_{k=1}^{K} \beta_k x_k\right)$$

$$= \mathrm{F}[\mu_j - (\alpha + \textstyle\sum_{k=1}^{K} \beta_k x_k)] \qquad (2)$$

where F(.) is the cumulative distribution function of

ε. If ε follows a logistic distribution, we have the general ordinal regression model:

$$\mathrm{P}(y \leq j|\,\mathbf{x}) = \frac{\exp[\mu_j - (\alpha + \sum_{k=1}^{K} \beta_k x_k)]}{1 + \exp[\mu_j - (\alpha + \sum_{k=1}^{K} \beta_k x_k)]} \qquad (3)$$

The following link function, probit link and cloglog link considered in this study were shown in Table 3.

Function	Form	Typical application
Probit	$\varphi^{-1}(y)$	Latent categories more probable
Cloglog	log(-log (y))	Higher categories more probable

III. Results and Discussion

Two link function namely, probit link and cloglog link function were analyzed and compared. The analysis showed the completed models for both link functions containing 26 clinical variables resulted in Table 4 and Table 5.

The completed model with the probit link function in Table 4 shows that pathological staging was significantly associated with three clinical variables of tumour thickness level, adjuvant therapy, and metastasis recurrence. These significant explanatory variables exhibited positive regression coefficients, except tumour thickness at level 2 (T2). On the other hand, based on completed model with cloglog link, Table 5 shows that the pathological staging was significantly associated with two explanatory variables, i.e., adjuvant therapy and metastasis recurrence. These significant explanatory variables exhibited positive regression coefficients. Only variables that were significantly associated to pathological staging were reported in this study.

Table 6 show 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. The difference in the chi-square statistics was 206.277 (206.277-0.000) which is significant at 0.05 level. It can be concluded that there was association between pathological staging and independent clinical variables - adjuvant therapy, tumor thickness, and metastasis recurrence.

For cloglog link, the model fitting statistics for observed reduced model indicates that the -2LL of the model with only intercept was 185.297 while, -2LL of the model with intercept and three independent variables were 0.000. That is the difference of Chi-square statistics was 185.297 (185.297-0.000) which is significant at 0.05 level. It can be concluded that there was association between pathological staging and two independent clinical variables - adjuvant therapy, and metastasis recurrence.

As a result of comparing two reduced models, it can be seen that the chi-square statistics from the cloglog link was smaller than the probit link. This indicates that cloglog link was preferred in this comparison. However, the model with probit link function was still good since the model did not violate the assumption of model adequacy.

TABLE 4
EXPLANATORY VARIABLES BASED ON COMPLETER 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 5
EXPLANATORY VARIABLES BASED ON COMPLETER MODEL USING CLOGLOG LINK

		Estimate	*p*-value
Threshold	[staging = 1]	-1.900	0.300
	[staging = 2]	0.736	0.681
	[staging = 3]	3.948	0.031
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]	4.450	0.012*
	[Adjuvant therapy=Yes]	2.784	0.020*
	[ASA Score=Level I]	0.000	1.000
	[ASA Score=Level II]	0.000	1.000
	[Tumour thickness=T2]	-0.977	0.402
	[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 7 display the Goodness of Fit statistics for reduced model with both probit and cloglog link. The additional model fitting statistic, the deviance equal 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. Whereas, the deviance equal 28.174 (with degree freedom of 4 and *p*-value=0.000) for the reduced model with the cloglog link indicate that the observed data were not consistent with the estimated values in the fitted model. Hence, the model with probit link was a more suitable and preferred model compared to cloglog link model based upon the chi-square test results. This means that the model with probit link fits the data well and the goodness of fit statistics suggest that model-predicted

cell proportions are acceptably close to the observed proportions.

TABLE 6
REDUCED MODEL FITTING USING PROBIT AND CLOGLOG LINK

Probit link				
	-2LL	χ^2	df	Sig.
Intercept only	206.277			
Final	0.000	206.277	4	0.000
Cloglog link				
	-2LL	χ^2	df	Sig.
Intercept only	185.297			
Final	0.000	185.297	2	0.000

**significance at 0.05*

TABLE 7
GOODNESS-OF FIT FOR REDUCED MODEL USING PROBIT AND CLOGLOG LINK

Probit link			
	χ^2	df	Sig.
Deviance	11.559	14	0.642
Cloglog link			
	χ^2	df	Sig.
Deviance	28.174	4	0.000*

**significance at 0.05*

The model-fitting statistics, namely the Pseudo R square as shown in Table 8 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 can be seen that the pathological staging explains 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.

However, the Pseudo R squares for model with cloglog link was Cox & Snell (0.873), McFadden (0.966), and Nagelkerke (0.898). It indicates that the pathological staging explained 87.3% of the variance in two clinical independent variables included in the reduced model according to Cox & Snell R square value, 96.6%, and 89.8% for McFadden, and Nagelkerke respectively with cloglog link. In comparison with both link functions, the reduced model with probit link had larger value of Nagelkerke and McFadden than the reduced model with the cloglog link. Thus, the reduced model with probit link was a better choice in this criterion.

TABLE 8
PSEUDO R-SQUARE USING PROBIT AND CLOGLOG LINK

Probit link		Cloglog link	
Pseudo R-square	Value	Pseudo R-square	Value
Cox & Snell	0.873	Cox & Snell	0.873
Nagelkerke	1.000	Nagelkerke	0.966
McFadden	1.000	McFadden	0.898

Therefore, the reduced model using probit link function was chosen over the cloglog link since the model fit the data well, met the assumption of parallel lines and has larger Pseudo R square value for Cox & Snell, McFadden, and Nagelkerke in the model.

The crosstabulation method was used in order to know how well the model obtain would be used to predict the prognosis factors for colorectal cancer patients staging. The classification table was used to categorize the classified and the actual response. Table 9 display the accuracy of the classification results for the pathological staging response categories using probit link. Table 9 and Table 10 show that classification rates were higher than 70%. From Table 9, the model demonstrated the perfect prediction of pathological staging with accuracy of 100% when the three prognosis factors of adjuvant therapy, tumor thickness, and metastasis recurrence were included in the model by using probit link function.

TABLE 9
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

However, the classification rate for pathological staging using clog-log link that shown in Table 10 is 72% [(4+0+58+10)/100%], which can be considered as good as many studies which found that the classification rates above 70% as acceptable [36].

TABLE 10
CLASSIFICATION RESULT FOR PATHOLOGICAL STAGING BASED ON CLOGLOG

		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	28 (28%)	**0 (0)**	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

The test for each of parameter estimates are displayed in Table 11. Using the model with probit link, the pathological staging was found to be significantly associated at 0.05 level of significance with *p*-value of 0.013, 0.000 and 0.000 for 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 11
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 positive signs were observed for the estimated parameters of metastasis recurrence, tumour thickness level of T3 and adjuvant therapy and negative sign for tumour thickness at level 2 (T2). Patients who received adjuvant therapy were ($e^{3.597}$) which is 36 times more likely to be the odds of having serious problem of cancer (higher stage of pathological staging) than less serious problem of cancer (lowest stage of pathological staging). Besides that, tumour thickness level 2 (T2) was $e^{-3.714}$ or 0.002 times less likely of invasion among patients who had serious problem of cancer (higher stage of pathological staging) than less serious problem of cancer (lowest stage of pathological staging).

IV. CONCLUSION

This study has identified the most suitable link function that can model the success of laparoscopic resection of colorectal cancer patients at various operative stages which lead to identification of pathological staging of colorectal cancer patients [34]. 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 stages. The findings in this study clearly justify that probit link function has effectively explained the prognosis factors that lead to the identification of pathological staging of colorectal cancer patients compared to cloglog link function. Adjuvant therapy, metastasis recurrence and tumour thickness level were found to be significant prognosis factors for determining the pathological staging. It is recommended that pathologist may use these findings to propose guidelines and consequently propose appropriate treatment plan for a particular patient according to their cancer staging. In conclusion, probit link function was considered to be a more suitable model for this study.

ACKNOWLEDGMENT

We would like to extend our deepest appreciation and gratitude to General Hospital Kuala Lumpur for their cooperation and assistance given throughout this research work. Our thanks also go to the Ministry of Higher Learning of Malaysia and Universiti Teknologi MARA Malaysia for supporting this research with the Fundamental Research Grant Scheme No. 600-RMI/ST/FRGS 5/3/Fst (191/2010).

REFERENCES

[1] J.J.Y. Sung, J.Y.W. Lau, K.L. Goh and W.K. Leung, "Increasing incidence of colorectal cancer in Asia: Implications for screening," *Lancet Oncol*, vol. 6, no. 11, pp. 871-876, 2005.
[2] G.L.C. Chye and H. Yahaya, "National cancer registry cancer incidence in Malaysia," *National Cancer Registry*, Ministry of Health Malaysia, 2003, vol. 2.
[3] H.T. Lynch and D.L.A, "Hereditary colorectal cancer," *N Engl J Med*, no. 384, pp. 919-93, 2003.
[4] E.E. Calle, C. Rodriguez, K. Walker-Thurmond and M.J. Thun, "Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults*," N Eng J Med*, vol. 17, no. 348, pp. 1625-1638, 2003.

[5] F.B. Hu, W.C. Willett, T. Li, M.J. Stampfer, G.A., Colditz and J.E., Manson, "Adiposity as compared with physical activity in predicting mortality among women*," N Engl J Med*, vol. 26, no. 351, pp. 2694-2703, 2004.

[6] R. Alteri, P. Bandi, D. Brooks, V. Cokkinides, M. Doroshenk, T. Gansler, K. Graves; E. Jacobs, D. Kirkland; J. Kramer, B. Levin, A. Magro, M. McCullough, D. Naishadham, B. McNeal; M. Shah, S. Simpson; R. Smith, K. Sullivan and D. Wagner, "Colorectal cancer facts & figures: Colorectal cancer screening*prevalence (%) among adults 50 years and older by state , 2006-2008," *American Cancer Society*, Atlanta, Geogria.

[7] S.C. Larsson and A. Wolk,"Meat consumption and risk of colorectal cancer: A meta-analysis of prospective studies*" Int J Cancer*, vol.11, no.119, pp.2657-2664, 2006.

[8] A. Chao, M.J. Thun, C.J. Connell,M.L., McCullough, E.J. Jacobs, W.D. Flanders, et al, "Meat consumption and risk of colorectal cancer", *JAMA*, vol. 2, no. 293, pp. 172-182, 2005.

[9] L.H., Kushi, T. Byers, C. Doyle, E.V. Bandera, M. McCullough, A. McTiernan, et al,"American Cancer Society Guidelines on nutrition and physical activity for cancer prevention: Reducing the risk of cancer with healthy food choices and physical activity", *CA Cancer J Clin*, vol. 5, no. 56, pp. 254-281, 2006.

[10] A. Chao, C.J. Connell, E.J. Jacobs, M.L. McCullough, A.V. Patel, E.E. Calle, et al,"Amount, type, and timing of recreational physical activity in relation to colon and rerctal cancer in order adults: The cancer prevention study II nutrion cohort," *Cancer Epidemiol Biomarkers Prev*, vol. 12, no. 13, pp.2187-2195, 2004.

[11] M.L. McCullough, A.S. Robertson, A. Chao, E.J. Jacobs, M.J. Stampfer, D.R. Jacobs, et al," A prospective study of whole grains, fruits, vegetables and colon risk," *Cancer Causes Control*, vol.10, no. 14, pp. 959-970, 2003.

[12] P. Terry, E. Giovannucci, K.B. Michels, L. Bergkvist, H. Hansen, L. Holmberg, et al,"Fruith, vegetables, dietry fiber, and risk of colorectal cancer*," J Natl Cancer Inst*, vol. 7, no. 93, pp. 525-533, 2001.

[13] A.K. Samad, R.S. Taylo, T. Marshall and M.A. Chapman, "A meta-analysis of the association of physical activity with reduced risk of colorectal cancer," *Colorectal Dis*, vol.3, no. 7, pp. 204-213, 2005.

[14] C.A. Tomeo, G.A. Colditz, W.C. Willett, E. Giovannucci, E. Platz, B. Rockhill, et al,"Harvard report on cancer prevention. Volume 3: Prevention of colon cancer in the United States," *Cancer Causes Control*, vol 3, no. 10, pp. 167-180, 1999.

[15] E. Botteri, S. Lodice, S. Raimondi, P. Maisonneuve, A.B. Lowenfels," Cigarette smoking and anenomatous polyps: A meta-analysis," *Gastroenterology*, vol. 2, no. 134, pp. 388-395, 2008.

[16] WHO, World health statistics annual. Geneva: WHO databank (2012, April 23). [Online], Available: http://www.dep.iarc.fr/

[17] P. Ferrari, M. Jenab, T. Norat, A. Moskal, N. Slimani, A. Olsen, et al," Lifetime and baseline alcohol intake and risk of colon and rectal cancers in the European prospective investigation into cancer and nutrion (epic)," *Int J Cancer*, vol.9, no. 121, pp. 2065-2072, 2007.

[18] A. Schatzkin, L.S. Freedman, S.M. Dawsey and E. Lanza, " Interpreting precursor studies: What polyp trials tell us above large-bowel cancer," *J Natl Cancer Inst*, vol.14, no. 86, pp.1053-1057, 1994.

[19] C.N. Bemstein, J.F., Blanchard, E. Kliewer and A. Wajda, "Cancer risk in patients with inflammantory bowel diese: A population-based study," *Cancer*, vol. 4, no. 91, pp. 854-862, 2001.

[20] L.E. John and R.S. Houlston,"A systematic review and meta-analysis of familial colorectal cancer risk," *Am J Gastroenterol*, vol. 10, no. 96, pp. 2992-3003, 2001.

[21] A.S. Butterworth, J.P. Higgins, and P. Pharoah, " Relative and absolute risk of colorectal ancer for individuals with a family history: A meta-analyis," *Eur J Cancer*, vol. 2, no. 42, pp. 216-227, 2006.

[22] How is colorectal cancer staged? ((2012, April 23). [Online], Available: http://www.cancer.org/Cancer/ColonandRectumCancer/DetailedGuide/colorectal-cancer-staged.

[23] American Cancer Society, "Cancer facts and figures 2008," *Atlanta, GA: American Cancer Society*, 2008.

[24] Malaysian Oncology Society (2012, April 23). [Online], Available: http://malaysiaoncology.org/article.php?aid=34

[25] AJCC, "Approved statement: Laparoscopic colectomy for curable cancer surgery or endoscopy," vol. 8, no. 18, pp. 1-9, 2002.

[26] B. Peterson and F.E. Harrell, "Partial proportional odds model for ordinal response variables," *Applied Statistics*, vol. 39, pp. 205-217, 1990.

[27] C. Beneki and A. Papastathopoulos, "The impact of structured, unstructured and integrated decision support systems on SME economic performance. An empirical study," in *Procedings of the 3rd International Conference on Communications and Information Technology*, Greece, 2009.

[28] N. Loukeris and C.C.F.E.A., "Comparative evaluation of multi layer perceptrons, to hybrid multi layer percetrons, with multicriteria hierarchical discrimination and logistic regression in corporate financial analysis," in *Procedings of 11th International Conference On Computers, Agios Nikalaos, Crete Island*, Greece, 2007.

[29] S.Y. Sohn and J.H. Lee, "Strategies for Advancing the Status of Women Scientists and Engineer in Kore," presented at the 7th WSEAS international Conference on Artificial Intelligence, Knowledge Engineering and Databases (AIKED'08), University of Cambridge, UK, 2008.

[30] Y.H. Chan, "Basic statistics or doctors: Multinomial logistic regression," *Singapore Med J*, vol. 46. no 6, pp. 259- 268, 2005

[31] Colorectal Cancer Staging (2012, April 23). [Online]. Available: http://www.healthcommunities.com/colon-cancer/staging.shtml

[32] M.H. Kutner, C.J. Nachtsheim and J. Neter, *Applied Linear Regression Models*, 4th Edition, New York: McGraw-Hill, 2004.

[33] H.E. Zhen and W.U. Du, "A comparative study of ordinal probit and logistic regression for affective product design," *Advanced Materials Research*, vol. 452-453, pp. 642-647, 2012.

[34] R. Bender and U. Grouven, "Ordinal logistic regression in medical research," *Journal of the Royal College of Physicians of London*, vol. 31, no. 5, pp. 546- 551, 1997.

[35] Stanislave Labatova, "Linear Multiple Regression Model of High Performance Liquid Chromatography." WSEAS Transactions on Infromation Science and Applications, Issue 3, vol.7, 2010.

[36] M.A. Hall, "Correlation based feature selection for machine learning," PhD. thesis, Department of Computer Science, University of Waikato, Hamilton, New Zealand, 1999.