International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Estimating the TNM Status for Squamous Cell Carcinomas Fusing Ordered Logistic Regression: A Seven Year Retrospective Case Study

Wan Muhamad Amir W Ahmad, Tang Liszen, Mohamad Arif Awang Nawi, Muhammad Azeem Yaqoob, Najwa Solehah Shamsul Bahrin, Nor Azlida Aleng

Abstract: ***Nowadays, many researchers discussed the importance of statistical analysis which was implemented in many areas of research which includes science and non-science field. For example, in medical or dental science research, there are many statistical tools had been used especially for the modeling and estimating purpose. One of the statistical tools which focus on the ordinal outcome is ordinal regression or ordered logistic regression. Generally, this study mainly focuses on ordinal logistics regression modeling. Through this ordered logistic regression we can estimate the TNM staging for those patients who attended Hospital University Sains Malaysia (USM) from 2011 to 2017 (based on secondary data). Besides that, the ordinal regression model also can be used as a tool for factor determination (having an association toward TNM staging) for an ordinal outcome. Through this study, we also address issues such as the global concept and interpretation of the ordinal logistic regression model. The significant result from this finding (output based) can be used to educate public people or stakeholder of how important this factor toward patients management.***

Index Terms**:** ***TNM staging, Ordinal regression, Maximum likelihood estimation, Oral squamous cell carcinomas.***

I. INTRODUCTION

The aim of ordinal regression is to model a dependent variable in terms of one or more independent variables. The method that used for the parameter estimation is maximum likelihood estimation (MLE). The dependent variable may consist of two or more than two categories (e.g., tumor less than 2 cm/between 2cm to 4 cm/more than 4cm). The ordinal regression model, also known as ordered logistic regression and sometimes referred to as the constrained cumulative logit model which originally proposed [1] and later called proportional odds model [2]–[4]. Ordered logistic regression estimates the probability of a certain event occurring and their goal is to gain the correctly predict the category of outcome for individual cases. Besides that, this ordered model can be used to determine the predictor factor associated with the focus studied parameter.

The best model can be created by considering all the significant predictor variables. There is a lot of research have been conducted on ordinal regression models. In this research paper, we are trying to determine the potential factor which contributes to the TNM staging of cancer.

Oral squamous cell carcinomas (OSCC) are the sixth most frequent malignant tumor and is a fatal disease of the oral cavity for up to 50% of mortality rate [5] – [6]. Globally, these cancers account for around 2 to 5 % total cancer cases with the highest prevalence being identified in Asia [7]–[8]. Although there is an advancement of therapeutic strategies in recent years, the overall survival rate remains the same over the past few decades [9]. Oral cancers have a multifactorial etiology which includes smoking, tobacco use, alcohol consumption, paan, betel quid, viral stimuli, and some genetic and epigenetic changes [10]–[12]. At present, one of the most effective prognostic tools for tumor survival is the tumor-node-metastasis (TNM) staging system [13]–[14]. Moreover, patient's socio-demographic and clinical characteristics with their age, gender, and smoking habits are also considered for choosing the appropriate therapeutic strategy, to determine the risk of complications and the prognostic value of numerous distinct types of cancer [15]–[16]. These several identification factors are associated with a poor prognosis thus has raised a critical issue in the treatment of OSCC. In previous studies, many different clinic-pathological parameters such as the age, smoking history, TNM staging [17], tumor spread in cervical lymph nodes, tumor size and microvascular invasion [18] have been studied as independent prognostic factors in patients with OSCC.t has been reported that 90% of the cancers of the oral cavity are OSCC [19]–[20].

Revised Manuscript Received on 30 May 2019.
* Correspondence Author
Wan Muhamad Amir W Ahmad*, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Malaysia.
Tang Liszen, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Malaysia.
Mohamad Arif Awang Nawi, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Malaysia.
Muhammad Azeem Yaqoob, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Malaysia.
Najwa Solehah Shamsul Bahrin, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Malaysia.
Nor Azlida Aleng School of Informatics and Applied Mathematics, University Malaysia Terengganu (UMT), Terengganu, Malaysia.

Oral cancer usually encompasses tumors derived from the lips, anterior two-thirds of the tongue, buccal mucosa, hard palate, the floor of mouth, upper and lower alveolar ridges, retromolar trigone, and sublingual area [21]–[22]. Squamous cell carcinoma is the most prevalent histological type, which is approximately 95% [23].

II. DATA AND METHODS

Data: Data from the medical unit record, Hospital USM were reviewed and related information was extracted from 2011 to 2017. The sampling frame was the list of patients, which diagnosed with oral cancer admitted to in Hospital USM. The data consists of several variables which are TNM staging, gender, salivary gland, and maxilla. The description of data in this study as shown in Table 1.1.

Table 1. 1 Description of Data

Variable	Code	Description
TNM Stage	Y	1 = Stage 1, 2 = Stage 2, 3 = Stage 3, 4 = Stage 4
Gender,	X_1	1 = Male, 2 = Female
Tumor at Salivary Gland	X_2	0 = No, 1 = Yes

Methodology: Logistic regression is useful when we are trying to model a categorical dependent variable (more than two categories) as a function of one or more independent variables. Ordinal logistic regression (OLR) is a type of logistic regression analysis when the response variable has more than two categories with having natural order or rank. In ordinal logistic regression, we need to have ordinal-scaled for the dependent variable. The model for ordinal is given by

$$y_i^* = x_i\beta + \varepsilon_i \qquad (1)$$

However, since the dependent variable is categorized, we must instead use:

$$C_x(x) = \ln\left[\frac{P(Y \le j \mid x)}{P(Y > j \mid x)}\right]$$

$$= \ln\left(\frac{\sum pr(\text{event})}{1 - \sum pr(\text{event})}\right) = \beta_0 + \beta_1 \text{Gender} + \beta_2 \text{Salivary Gland} \qquad (2)$$

where

α_j = called threshold or intercept

$\beta_0, \beta_1, \beta_2$ = Parameter in the model

X_{i1} = Set of factors or independent variables

Equation 2 above is an ordinal logistic model for k predictors with p-1 levels response variable [23].

III. RESULTS AND DISCUSSION

Table 1.2 shows the summary of the variable assessment.

Table 1.2 Parameter Estimate and the assessment of the variable significant

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[TNM_Stage = 1]	-4.077	1.540	7.006	1	0.008	-7.095	-1.058
	[TNM_Stage = 2]	-1.059	1.223	0.750	1	0.387	-3.455	1.338
	[TNM_Stage = 3]	-0.046	1.222	0.001	1	0.970	-2.440	2.349
Location	Gender	-1.147	0.430	7.127	1	0.008*	-1.989	-0.305
	Salivary_gland	1.401	0.562	6.210	1	0.013*	0.299	2.503

*Significant at 0.05

Table 1.3 Analysis of Maximum Likelihood Parameter Estimates

Parameter Estimates

		Estimate	Std. Error	Wald	df	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Threshold	[TNM_Stage = 1]	-4.586	1.033	19.709	1	0.000	-6.610	-2.561
	[TNM_Stage = 2]	-1.568	0.347	20.383	1	0.000	-2.248	-0.887
	[TNM_Stage = 3]	-0.555	0.302	3.386	1	0.066	-1.146	0.036
Location	[Gender=1]	1.147	0.430	7.127	1	0.008	0.305	1.989
	[Gender=2]	0[a]	.	.	0	.	.	.
	[Salivary_gland=1]	-1.401	0.562	6.210	1	0.013	-2.503	-.299
	[Salivary_gland=2]	0[a]	.	.	0	.	.	.

The association between TNM staging with gender and salivary gland variables was analyzed using an ordinal regression model (from Table 1.2). The result was considered statistically significant because the *p*-value is less than 0.05. All statistical analyses in this study were performed using SPSS software. Table 1.3 gives a detail explanation of TNM staging with the independent variables.

The estimated the coefficient for gender is 1.147. The value of OR (3.149) is obtained exp 1.147) = 3.149. The odds of achieving a higher stage of cancer is approximately three times in male compared to female. The estimated coefficient for salivary gland is -1.401.

The value of OR (0.246) is obtained exp (-1.401) = 0.246. This can be explained, about 75.4 % of those who were having cancer at a salivary gland diagnosed at an earlier stage of TNM compare to those who were not having cancer at a salivary gland.

Table 1.4 Model Fitting Information

Model	-2 Log Likelihood	Chi-Square	df	Sig.
Intercept Only	40.704			
Final	26.784	13.920	2	.001
Link function: Logit.				

According to Table 1.4 above, the statistically significant chi-square statistic ($p < 0.05$) indicates that the final model gives better predictions.

Table 1.5 Goodness-of-Fit

	Chi-Square	df	Sig.
Pearson	2.114	7	.953
Deviance	2.468	7	.930
Link function: Logit.			

Table 1.5 shows the output of Goodness of fit tables. This table indicates that the *p*-value of a person is not significant and it shows that the overall model is fit. This interpretation is going similar to deviance χ^2 in the second row of the same Table 1.5.

IV. SUMMARY AND CONCLUSION

This paper is focusing on the application based on ordinal regression. At first, the studied variable was check accordingly due to they're significant by adjusting the SPSS syntax (Table 1.2). If the factor is significant at the level of 0.05, we can include that this factor contributes to the TNM staging. By doing this, only significant factors will be entered into the model. The selected variable will be run once again using SPSS software (Table 1.3). The full output that gaining from the final analysis will be interpreted carefully. Ordered logistics regression model shows that there are two significant factors which are leading to the prediction of TNM Staging. The first factor is the gender factor and the second factor is the involvement of salivary gland.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Universiti Sains Malaysia (USM) for providing the research funding (Grant no.1001/PPSG/8012278, School of Dental Sciences).

REFERENCES

1. S. H. Walker, and D.B. Duncan, "Estimation of the probability of an event as a function of several independent variables," Biometrika, vol 54, 1967, pp. 167–179.
2. C. V. Ananth, and D.G. Kleinbaum, "Regression model for ordinal responses: A review of methods and applications," Int. J. Epidemiol., vol 26, 1997, 1323–1332.
3. P. McCullagh, "Regression models for ordinal data," Journal. R. Stat. Soc., vol 42, 1980, pp. 109–142.
4. A. Agresti. (2007). An Introduction to Categorical Data Analysis. 2nd Edition. Wiley, New York.
5. 5. V. Ajila, H. Shetty, S. Babu, V. Shetty, and S. Hegde, "Human Papilloma Virus Associated Squamous Cell Carcinoma of the Head and Neck," Journal of Sexually Transmitted Diseases, Sept. 2015, pp. 1–5.
6. 6. R. Mehrotra, and S. Yadav, "Oral squamous cell carcinoma: etiology, pathogenesis and prognostic value of genomic alterations," Indian Journal of Cancer, vol 43, no 2, 2006, 60–66.
7. 7. I. A. Siddiqui, M. U. Farooq, R. A. Siddiqui, and S. T. Rafi, "Role of toluidine blue in early detection of oral cancer," Pakistan Journal of Medical Sciences, vol 22, no 2, 2006, pp. 184–187.
8. 8. K. D. Shield, J. Ferlay, and A. Jemal, "The global incidence of lip, oral cavity, and pharyngeal cancers by subsite in 2012," Cancer Journal for Clinicians, vol 67, no 1, 2017, pp. 51–64.
9. 9. A. K. Chaturvedi, "Epidemiology and clinical aspects of HPV in head and neck Cancers," Head Neck Pathology, vol 6, 2012, pp. 16–24.
10. 10. F. Vargas-Ferreira, F. Nedel, A. Etges, A. P. N. Gomes, C. Furuse, and S. B. C. Tarquinio, "Etiologic factors associated with oral squamous cell carcinoma in non-smokers and non-alcoholic drinkers: a brief approach," Brazilian Dental Journal, vol 23, 2012, pp. 586–590.
11. 11. J. Ali, B. Sabiha, H. U. Jan, S. A. Haider, A. A. Khan, and S. S. Ali, S. S, "Genetic etiology of oral cancer," Oral Oncology, 2017, pp. 23 – 28.
12. 12. K. C. Ribeiro, L. P. Kowalski, and M. R. Latorre, "Impact of comorbidity, symptoms, and patients' characteristics on the prognosis of oral carcinomas," Arch Otolaryngol Head Neck Surg," vol 126, no 9, 2000, pp. 1079–1085.
13. 13. S. Fan, Q. L. Tang, Y. J. Lin, W. L. Chen, J. S. Li, Z. Q. Huang, Z. H. Yang, Y. Y. Wang, D. M. Zhang, H. J. Wang, E. Dias-Ribeiro, Q. Cai, and L. Wang, "A review of clinical and histological parameters associated with contralateral neck metastases in oral squamous cell carcinoma," Int. Journal Oral Sciences," vol 3(4), 2011, pp.180–191.
14. 14. V. Rossi, M. Tarozzi, G. Lodi, A. Sardella, F. Demarosi, and A. Carrassi, "Clinical aspect and survival rates in a subject with oral cancer: a retrospective cohort study," Minerva Stomatol, vol 56, 2007, pp. 591–601.
15. 15. D. Kawakita, S. Hosono, H. Ito, I. Oze, M. Watanabe, N. Hanai, Y. Hasegawa, K. Tajima, S. Murakami, H. Tanaka, and K. Matsuo, "Impact of smoking status on clinical outcome in oral cavity cancer patients," Oral Oncol, vol 48, no 2, 2012, pp. 186–191.
16. 16. M. Kreppel, T. Dreiseidler, D. Rothamel, H. T. Eich, U. Drebber, J. E. Zoller, and M. Scheer, "The role of clinical versus histopathological staging in patients with advanced oral squamous cell carcinoma treated with neoadjuvant radiochemotherapy followed by radical surgery," Journal Craniomaxillofac Surg, vol 41, no 1, 2013, pp. 22 – 27.
17. 17. M. Grimm, "Prognostic value of clinicopathological parameters and the outcome in 484 patients with oral squamous cell carcinoma: microvascular invasion (V+) is an independent prognostic factor for OSCC," Clinical and Translational Oncology, vol 14, no 11, 2012, pp. 870–880.
18. 18. N. W. Johnson, P. Jayasekara, and A. A. Amarasinghe, "Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and etiology," Periodontol 2000, vol 57, no 1, 2011, pp. 19–37.
19. 19. A. K. Markopoulos, "Current Aspects on Oral Squamous Cell Carcinoma," The Open Dentistry Journal," vol 6, 2012, pp. 126–130.
20. 20. L.Feller, and J. Lemmer, "Oral squamous cell carcinoma: epidemiology, clinical presentation, and treatment," Journal of Cancer Therapy, vol 3, 2012, pp. 263–268.
21. 21. B. V. Sundermann, L. Uhlmann, J. Hoffmann, K. Freier, and O. C. Thiele, "The localization and risk factors of squamous cell carcinoma in the oral cavity: A retrospective study of 1501 cases," Journal of Cranio-Maxillofacial Surgery, vol 46, Feb. 2018, pp. 177–182.
22. 22. B. Ayaz, K. Saleem, W. Azim, and A. Shaikh, "A clinicopathological study of oral cancers," Biomedical Journal, vol 27(1), 2011, pp. 29–32.
23. 23. K. A. Adeleke, and A. A. Adepoju, "Ordinal Logistic Regression Model: An Application to Pregnancy Outcomes," Journal Math Stat., vol 6, 2010, pp. 279–285.

AUTHORS PROFILE

Wan Muhamad Amir W Ahmad is a lecturer at the School of Dental Sciences, Universiti Sains Malaysia (USM). He is a statistician by profession with Ph.D. in Biostatistics, MSc. in Applied Statistics and BSc. in Applied Science (Applied Statistics). His core academic teaching includes Biostatistics, statistical software application, statistical computing, statistical consultant, epidemiological research designs in health sciences research, research methodology, elementary statistics, time series forecasting, operational research, advanced statistics and design of experiment (DOE). He has vast experience in teaching undergraduate and postgraduate students. He is active in many aspects of Biostatistics, having contributed to theoretical Biostatistics in the areas of building statistical methodology, as well as conducting SPSS workshops.

Tang Liszen is an Oral & Maxillofacial Surgeon and lecturer at the School of Dental Sciences, Universiti Sains Malaysia (USM). He obtained his bachelor degrees in DDS from USM and master degrees in MDS Oral & Maxillofacial Surgery (Hong Kong). His research interests are in oral cancer, orthogenetic surgery, and dental implantology.

Mohamad Arif Awang Nawi is a statistician by profession with Ph.D. in Applied Statistics, Master of Science (MSc.) in Mathematical Sciences Master by Research (Applied Statistics) and Bachelor of Science (Mathematics Computational). His publication focuses on statistics and publishes more than 15 papers in international and two academic books. His core academic teaching includes biostatistics, statistical software application, statistical computing, statistical consultation, research methodology and applied statistics in medical, education, financial, social science, and economic. He has experience in teaching undergraduate and postgraduate students. He also selected as a Reviewer for Advances in Science, Technology and Engineering Systems Journal and Science Alert.

Muhammad Azeem is currently doing his Ph.D. in Oral Medicine at the School of Dental Sciences, Universiti Sains Malaysia (USM). He obtained his Bachelor of Dental Surgery (2012) from Hamdard University and M.Sc. in Oral Medicine (2018) from USM. He has professional expertise in clinical, laboratory work and SPSS. His future interests are to continue working on oral cancer modeling and programming.

Najwa Solehah Shamsul Bahrin is currently pursuing her Bachelor Degree of Dental Surgery at School of Dental Sciences, Universiti Sains Malaysia (USM). She has an interest in community works. She involved in volunteerism programme at Gua Musang Kelantan, Yala Thailand and Acheh Indonesia. Her research interest is involving oral cancer cases.

Nor Azlida Aleng is currently pursuing her Ph.D. in Biostatistics. She is a lecturer at the School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu. She obtained her bachelor degree in Mathematics and a master degree in Management Mathematics. Her teaching experience for undergraduate students is in the field of biostatistics, neural network, mathematical modeling and Islamic financial in mathematics, and currently focusing on mathematical modeling in medical data using robust regression and multilayer feed forward neural network methods