

A Statistical Application on Determining the Prognostic Factors of Oral Squamous Cell Carcinomas (Osc) In Malaysia

Wan Muhamad Amir W Ahmad, Nurhayu Abdul Rahman, Muhammad Azeem Yaqoob, Nor Azlida Aleng, Nurfadhlina Abdul Halim, Mohamad Arif Awang Nawi

Abstract: Oral cancer is an important global health concern, representing the sixth most frequent malignant tumor. The oral squamous cell carcinoma (OSCC) is the most common malignant neoplasm of the oral cavity with up to 50% of mortality rate (highest prevalence being identified in Asia). In 2012, it has been reported that 14.1 million new cancer cases and 8.2 million cancer deaths. Numbers of studies have been performed to investigate the factors that have direct and indirect or both associated with the OSCC, including their survival time. In this paper, the potential clinicopathological prognostic factors will be determined in patients who attended Hospital Universiti Sains Malaysia from 2005 to 2015. For such prediction, the use of hazard regression is used previously, but here an attempt is made to propose a covariate-dependent prognostic model to identify the factors and the predictor importance according to the statistical significant point of view. The proposed model is very useful for the prediction and for the inferences of the patient's management time with the high-risk clinicopathological factors.

Index Terms: Clinicopathological, Covariate-Dependent Prognostic Model, Decision Tree, Prognostic Cancer Model

I. INTRODUCTION TO THE PROGNOSTIC SURVIVAL CANCER MODEL

Various kind of cancerous lesions from diverse cellular lineages originate in the oral cavity. Among these, oral squamous cell carcinoma (OSCC) is the most common, comprising 95% of head and neck cancers [1, 2]. Globally, the incidence of oral cancer holds the eighth position and show epidemiologic variability according to different geographic regions [3, 4]. Besides the proven causative agents like tobacco, alcohol, and human papillomavirus, there are certainly other factors that play a significant role in the selection of treatment strategies and determination of prognosis in OSCC patients. These range from simple demographic/general physical factors to the discovery of

Revised Manuscript Received on July 05, 2019.

Wan Muhamad Amir W. Ahmad, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia.

Nurhayu Abdul Rahman, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia.

Muhammad Azeem Yaqoob, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia.

Nor Azlida Aleng, School of Informatics and Applied Mathematics, Universiti Malaysia Terengganu, 21030, Kuala Terengganu, Malaysia.

Nurfadhlina Abdul Halim, Faculty Science and Technology, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800 Nilai, Negeri Sembilan, Malaysia.

Mohamad Arif Awang Nawi, School of Dental Sciences, Health Campus, Universiti Sains Malaysia (USM), 16150 Kubang Kerian, Kelantan, Malaysia.

newer molecular markers, comprising the clinical and histopathological factors [5]. However, none of the factors can alone influence the prognosis and it has been observed that a multitude of factors/parameters needs to be included to determine the prognosis and treatment outcome of a patient. In practice, the survival rate is often used to discuss the patient's prognosis. The survival rate refers to the percentage of people who live for a specific time after their cancer is diagnosed. Despite the significant advancement in the prevention and treatment protocols, the 5-year survival rate in patients with OSCC remains low due to recurrence and lack of suitable markers for early detection [6]. With the current growth of research in determining the prognosis of oral cancerous lesions, researchers have gained access to the considerable amount of gathering high-throughput data. However, data interpretation efforts seem lacking and obtained knowledge is not using in the right direction to translate into better diagnosis and treatment strategies. The key problem in the integration and interpretation of cancer data need to be addressed as a "network" which give the insight to further our understanding in dealing with the underlying mechanisms in oral cavity cancers. Statistical survival modeling offers an avenue for cancer network modeling [7]. In this paper, a prognostic survival cancer model is focused specifically on OSCC data.

II. MATERIAL AND METHODS

The archive of medical unit record of Hospital Universiti Sains Malaysia (USM) was reviewed and related information was extracted. A total of 57 eligible cases were selected from the list of a patient diagnosed with OSCC. The selected variables are shown in Table 2.1 as follows:

Table 2.1 Description of data

| No | Variables | Explanation of user variables |
|----|-----------------------|--|
| 1. | Time weeks | Time in weeks |
| 2. | Treatment | Treatment status : 0 =Yes, 1= Not |
| 3. | Tumour_classification | Tumor classification : 0 = Tumour less than 4 cm, 1 = Tumour more than 4 cm |
| 4. | Alcohol_Consumption | Alcohol_Consumption : 0 = Alcohol User, 1 = Non-alcohol User |
| 5. | Gender | Gender : 0 = Male, 1= Female |
| 6. | TNM staging | Code for Survival Analysis TNM staging : 0 = Early stage, 1= Last stage Code for Decision Analysis TNM_Staging : 1 = Stage 0, 2 = Stage 1, 3 = Stage 2, 4 = Stage 3, 5 = Stage 4a,4b,4c |
| 7. | HPV16 | HPV 16 status : 0 =Negative, 1= Positive, 2 = Not enough tissue for IHC |

A Statistical Application on Determining the Prognostic Factors of Oral Squamous Cell Carcinomas (Oscc) In Malaysia

- 8. **Perineural_Invasion** Perineural invasion : 1 = Not detected, 2 = Yes, 3 = Not specified
- 9. **Tumour location** Tumour location : 1 = Tongue, 2 = Buccal mucosa, 3 = Floor of mouth, 4 = Palate, 5 = Lips , 6 = Mandible/Retromolar 7 = Others
- 10. **TNM_Staging** TNM_Staging : 1= Stage 1, 2= Stage 2, 3= Stage 3, 4 = Stage 4a,4b,4c

Hazard Regression for Prognostics Model Of Oscc

To explore the underlying association between hazard regression with the selected explanatory variables, one set of equation model is fitted in this section. Equation (i) proposed the hazard regression model for our study.

$$HR = \text{Exp} \left(\beta_1 \times \text{Treatment} + \beta_2 \times \text{T Classification} + \beta_3 \times \text{Alcohol_Consumption} + \beta_4 \times \text{Gender} \right) \quad (i)$$

All the parameter in equation (i) will be fixed by using statistical software. All these independent variables will be tested at the early stage to make sure the proposed model has a high quality of inferences. Four variables were selected and listed in equation (i). Independent variables that have the p -value < 0.25 and clinically important is included in the model [8].

Decision Tree Of Tnm Staging Of Oscc

A decision tree is considered as a useful tool in statistical modeling in describing decisions. It is an efficient method for classification, prediction and for facilitating decision making in sequential decision problems [9]. Decision process which involves a lot of sequential decisions will lead to the decision problem, this is because the decision becomes difficult to visualize and to implement. Decision trees are indispensable graphical tools in such settings, which allow for an intuitive understanding of the problem and aid for the optimal decision making. The conceptual framework of our proposed decision tree is outlined as follows:

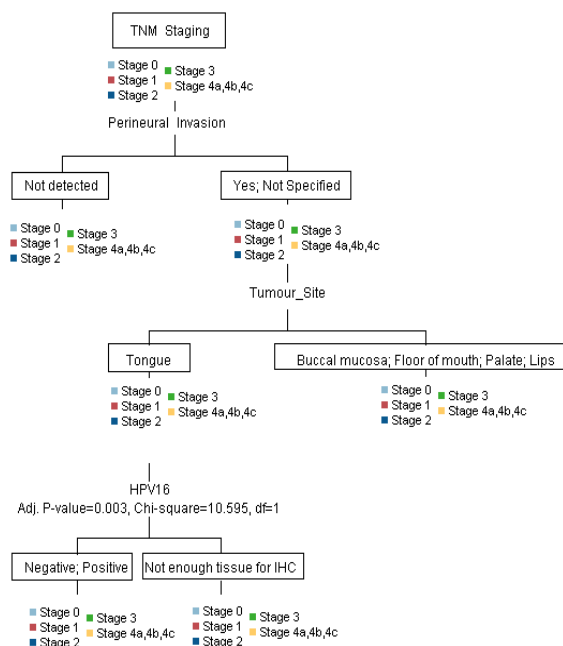


Figure 2.1. Conceptual Framework of TNM Staging for OSCC

III. RESULTS AND DISCUSSION

Hazard Regression for Prognostics Model Of Oscc

Table 3.1 shows the results for the hazard regression model. All four proposed factors were significant at $p < 0.25$, which indicates that the hazard regression is associated with four prognostics factors. The first prognostic factor is a treatment that patient received from the Hospital Universiti Sains Malaysia. Treatment received by patients had shown a significant prognostic factor for the OSCC ($\beta = 0.652$; $p < 0.25$). It can be concluded that patients with no treatment received have 1.920 times risk of dying than the patient with treatment received. The second prognostic factor is T classification ($\beta = 0.965$; $p < 0.25$). T classification is a significant prognostic factor for the OSCC. Patients with classification (between the stage of T3 and T4) have 2.625 times risk of dying than the patient with T classification (between the stage of T1 and T2). The third prognostic factor is an alcohol consumption with ($\beta = 1.664$; $p < 0.25$). Patients who used the alcohol had 5.279 times risk of dying than a patient without using alcohol. Alcohol seems to be a very large prognostic factor that contributes to hazard regression. The fourth factor is the gender factor ($\beta = -0.942$; $p < 0.25$). Female patients have 61% less risk of death than male patients.

Table 3.1 Variables in the equation

| | B | SE | Wald | D.f. | Sig. | Exp(B) |
|---------------------|--------|-------|-------|------|-------|--------|
| Treatment | 0.652 | 0.416 | 2.455 | 1 | 0.117 | 1.920 |
| T classification | 0.965 | 0.438 | 4.853 | 1 | 0.028 | 2.625 |
| Alcohol Consumption | 1.664 | 1.030 | 2.611 | 1 | 0.106 | 5.279 |
| Gender | -0.942 | 0.427 | 4.875 | 1 | 0.027 | 0.390 |

The results for the model (i) is displayed in Table 3.1. It can be written in an equation (ii) as follows:

$$HR = \text{Exp} \left(0.652 \times \text{Treatment} + 0.965 \times \text{T Classification} + 1.664 \times \text{Alcohol_Consumption} - 0.942 \times \text{Gender} \right) \quad (ii)$$

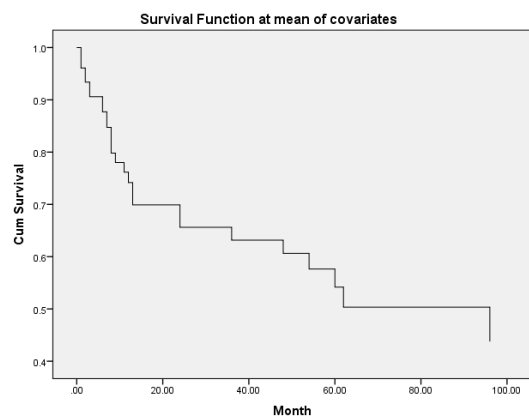


Figure 3.1 Overall Survival Function Estimates



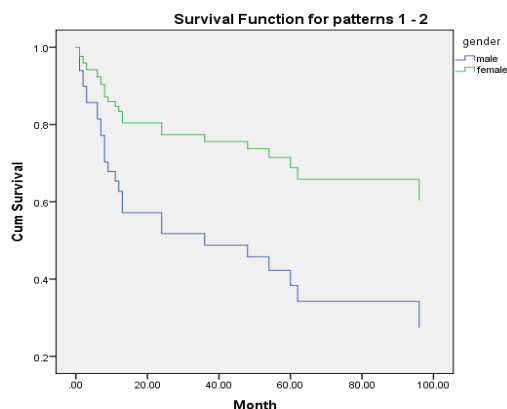


Figure 3.2 Survival Function Estimates According to Gender

Figure 3.1 shows the survival probabilities for OSCC. The plot shows that the survival probability is at all-time points to death among oral cancer patient which registered in Hospital University Sains Malaysia (USM). Figure 3.2 shows the survival function estimates according to gender. The survival probability is about lower for males compared to female at all-time points to death among the OSCC patients.

Decision Tree of Tnm Staging of Oscc

A decision tree model has been designed for the OSCC patient which consist of three predictors variable. The process will automatically include in its rule, only that attributes that really matter in making a decision. The attributes that do not contribute to the accuracy of the tree are ignored. This allows us to develop a classification system that predicts observation based on a set of decision rules. Figure 3.3 shows top three predictors, ranking to their contribution i.e. tumor site, perineural invasion, and HPV16 status. Using the CHAID method, nerve invasion and tumor size are the best predictors of TNM Staging. The highest predictor importance is on the tumor site.

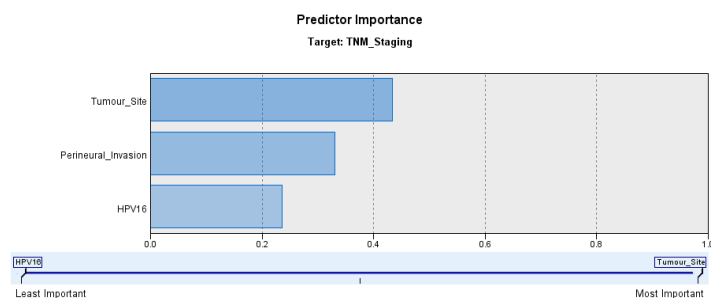


Figure 3.3: Predictor Important

The model was constructed based on the recommendation proposed by IBM SPSS Modeler 18.0 to ensure that this model fits the data. To determine the suitability of the model, the measurement of adj. p-value from the Chi-Square were evaluated for the decision tree. Decision tree analysis process will automatically include in its rule only that attributes that really matter in making a decision. Attributes that do not contribute to the accuracy of the tree are ignored. This provides very useful information about the data inferences. According to Figure 3.4, the top three predictors are shown that contribute in ranking i.e. tumor site, perineural invasion, and HPV16 status. Using the CHAID method, nerve invasion and tumor size are the best predictors of TNM Staging followed by perineural invasion and HPV16 status.

According to the decision tree analysis, the first split of TNM staging is based on perineural invasion (adj. p-value = 0.037, Chi-square = 6.243, d.f. = 1). In this case, we can interpret this as a smaller p-value indicates a better quality of the split. From the Figure 3.4, a perineural invasion with “Not Identified” status is assigned to Node 1, while perineural invasion with “Yes and Not Stated” is assigned to Node 2. Node 1 show that Stage 1 (8.7%), Stage 2 (13.0%), Stage 3 (30.4%), Stage 4 (4a, 4b and 4c) is about 47.8%. Node 2 show that Stage 1 (2.9%), Stage 2 (2.9%), Stage 3 (11.77%), Stage 4 (4a, 4b and 4c) is about 82.35% (n = 28). This is the highest of Stage 4 reported in our study. For the second split, the model has divided these patients into two sub-categories (Nodes 3 and 4), based on the reading of the tumor site. The split had divided into two groups Node 3 (tongue) and Node 4 (buccal mucosa, the floor of mouth, palate, and lips). Sixteen cases (28.07%) tumor site is reported at tongue site and eighteen cases (31.58%) at buccal mucosa, the floor of mouth, palate, and lips. For the third split, the model has divided these patients into two sub-categories (Nodes 5 and 6), based on HPV16. The split had divided into two groups Node 5 (Negative or Positive) and Node 6 (Not enough tissue for IHC). Thirteen cases (22.81%) with HPV16 reported (with Negative or Positive) and three cases (5.26%) reported (Not enough tissue for IHC).

A Statistical Application on Determining the Prognostic Factors of Oral Squamous Cell Carcinomas (Oscc) In Malaysia

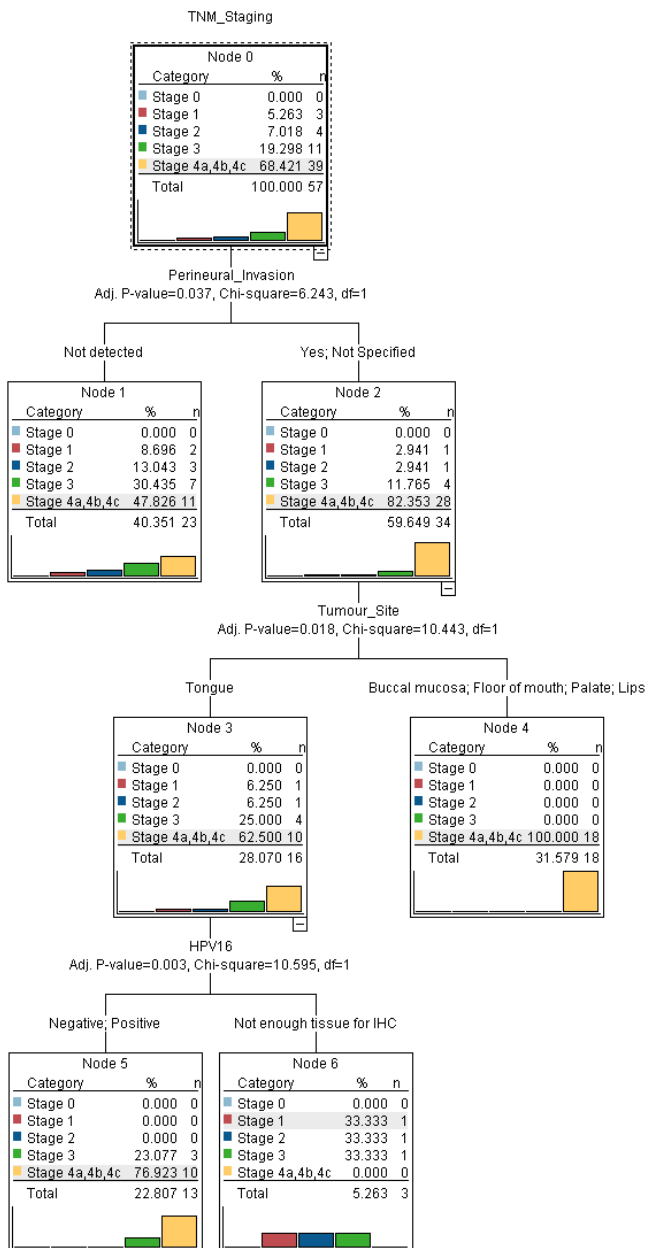


Figure 3.4: Decision Tree for Oral Cancer Classification

IV. SUMMARY AND DISCUSSION

The main purpose of this paper is to demonstrate such relationship and find the prognostic factor the inferences purpose. The first part of this paper examines the factors that lead to OSCC. The survival time of OSCC patients may depend on factors related to (i) the treatment which obtained from the hospital, (ii) T Classification, (iii) alcohol consumption and (iv) gender. This paper provides a preliminary overview of the associated factor for OSCC. According to Azeem (2018), T classification variable which represents the tumor size was statistically significant with a p -value of less than 0.05. Those who were in the late stage (T3, T4) were at almost five-time higher odds to deceased as compared to those who were in the early stage (T1, T2)(Azeem, 2018). Patients those who have a habit of

alcohol showed an increased risk of death as compared to non-alcoholic OSCC patients. However, the smoking status of the alcoholic user was not determining, therefore, the increased risk of death might not be attributable to alcohol consumption alone. Gender factor plays a significant role in the outcome as the p -value for this study was statistically significant at less than 0.05. The multivariable analysis shows that female with OSCC in Kelantan population has 8% lower risk to decrease as opposed to males (Azeem, 2018). The second part of the analysis is to develop a classification system that predicts observation based on a set of decision rules. From the decision tree analysis, we can determine the pattern and come out with inferences about the whole scenario. The decision tree model reveals the findings more explicitly due to use of diagram classification for OSCC patients. Using the CHAID method, nerve invasion and tumor size are the best predictors of TNM Staging. A major finding shows that the highest predictor importance is on tumor site and perineural invasion. Most of the OSCC cases were reported at stage four with the perineural invasion status of “Not Identified”, “Yes or Not Stated” status and the high number of a tumor comes from tongue site. From the both approached method, we can predict a better forecasting result in future for the decision making. This promising technique had lead to successful research and give the best results for the decision making especially for the decision maker.

V. ACKNOWLEDGMENT

The authors would like to express their gratitude to Universiti Sains Malaysia (USM) for providing the research funding RUI Grant no. 1001/PPSG/8012278 and no. Short Term 304/PPSG/61313195 School of Dental Sciences, Kampus Kesihatan).

REFERENCES

- Chimenes-Kustner, E., I. Font-Costa, and J. Lopez-Lopez, *Oral cancer risk and molecular markers*. Med Oral Patol Oral Cir Bucal, 2004. **9**(5): p. 381-4; 377-80.
- Vigneswaran, N. and M. D. Williams, *Epidemiological Trends in Head and Neck Cancer and Aids in Diagnosis*. Oral and maxillofacial surgery clinics of North America, 2014. **26**(2): p. 123-141.
- Johnson, N.W., Jayasekara P., and Amarasinghe A. A., *Squamous cell carcinoma and precursor lesions of the oral cavity: epidemiology and etiology*. Periodontol 2000, 2011. **57**(1): p. 19-37.
- Gupta, N., et al., *Changing Trends in oral cancer - a global scenario*. Nepal Journal of Epidemiology, 2016. **6**(4): p. 613-619.
- Majumdar, B., et al., *Clinico-pathological prognosticators in oral squamous cell carcinoma: An update*. Translational Research in Oral Oncology, 2017. **2**: p. 2057178X17738912.
- Taghavi, N. and I. Yazdi, *Prognostic factors of survival rate in oral squamous cell carcinoma: clinical, histologic, genetic and molecular concepts*. Arch Iran Med, 2015. **18**(5): p. 314-9.
- Blair, R. H., D.L. Trichler, and D. P. Gaille, *Mathematical and Statistical Modeling in Cancer Systems Biology*. Frontiers in Physiology, 2012. **3**: p. 227.
- Mickey, J. and Greenland, S., *A study of the impact of confounder-selection criteria on effect estimation*, American Journal of Epidemiology, 1989.129, 125-137.
- Mesarić J. and Šebalj D. *Decision trees for predicting the academic success of students*. Croatian Operational Research Review, 2016. **7**(2): 367-388.



10. Azeem, M., *Human Papillomavirus (HPV) Related Oral Squamous Cell Carcinoma In Kelantan Population*. Msc Thesis. Universiti Sains Malaysia, Malaysia.

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 BSc and MSc in Applied Statistics and PhD in Biostatistics from USM. His core academic teaching includes biostatistics, statistical software application, statistics computing, statistical consultant, epidemiology research designs in health sciences research, research methodology, elementary statistics, time series forecasting, operational research, advanced statistics and design of experiment.



Nurhayu Ab Rahman is a dental surgeon and lecturer at the School of Dental Sciences, Universiti Sains Malaysia (USM). She obtained her bachelor degrees in DDS from USM and DClintDent from University of Western Australia (UWA), Australia. Her research interests are in odontogenic tumours and non-odontogenic lesion affecting head and neck region, oral mucosal lesion and oral manifestation of systemic diseases. Her expertise are in oral medicine and oral pathology.



Muhammad Azeem Yaqoob 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.



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.



Nurfadhliha Abdul Halim is a lecturer at Faculty of Science and Technology, Universiti Sains Islam Malaysia. She obtained her BSc in Financial Mathematics from UMT, MSc in Management Mathematics and PhD in Mathematics from UKM. She has been conducting researches in various fields, namely financial mathematics, Islamic financial mathematics, mathematical modelling, statistics and optimization.



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.