RESEARCH ARTICLE

BORNEO SCIENCE
THE JOURNAL OF SCIENCE & TECHNOLOGY

Journal homepage: https://jurcon.ums.edu.my/ojums/index.php/borneo-science

SEMI-PARAMETRIC ANALYSIS OF CERVICAL CANCER DATA: A SINGLE-CENTRE STUDY

Nurliyana Juhan*[1], Yong Zulina Zubairi[2] and Agnes Ayang Kenyang[3,4]

[*1] Preparatory Centre for Science and Technology, Universiti Malaysia Sabah, Sabah, Malaysia.
[2] Institute for Advanced Studies, University of Malaya, Kuala Lumpur, Malaysia.
[3] Faculty of Science and Technology, Universiti Malaysia Sabah, Sabah, Malaysia.
[4] College of Computing, Informatics and Mathematics, Universiti Teknologi MARA, Samarahan 2 Campus, Sarawak, Malaysia.

Correspondence:
liyana87@ums.edu.my

Received: 15 March 2025
Revised: 14 April 2025
Accepted: 26 May 2025
Published online: 27 May 2025

DOI:
10.51200/bsj.v46i1.6226

Keywords:
Cervical cancer; Survival analysis; Proportional hazards assumption; Cox proportional hazards

ABSTRACT. *In women, cervical cancer ranks fourth in terms of mortality and is the third most prevalent disease. It remains a significant concern among clinicians in Malaysia; however, published works on the survival of cervical cancer patients are somewhat scarce. Thus, this study aims to identify the prognostic factors that significantly affect the risk of death in cervical cancer patients using semi-parametric Cox proportional hazards regression analysis. This study began with univariate and multivariate Cox proportional hazards regression analyses, followed by a proportional hazards assumption test for the preliminary final model. Data on cervical cancer patients treated at Hospital Universiti Sains Malaysia (HUSM) between 2013 and 2017 were utilized. In the univariate analysis, stage at diagnosis and primary treatment were found to be statistically significant at the 5% level. In the multivariate analysis, histologic type, stage at diagnosis, and distant metastasis were found to be statistically significant. The proportional hazards assumption for each variable in the preliminary final model is tested based on the scaled Schoenfeld residuals. Accordingly, this study showed that patients with stage III–IV adenocarcinoma-type cervical cancer treated at HUSM have the highest likelihood of death from the disease.*

INTRODUCTION

The disease's progression is often asymptomatic in its early stages, underscoring the importance of regular screenings for timely detection. A study carried out in the northern state of Malaysia revealed that 38.6% of women there had a Pap smear in the last five years, compared to 48.5% of women in the southern state who visited outpatient clinics and had a Pap smear in the previous three years (Chin *et al*., 2022; Mustafa *et al*., 2022; Norkhafizah & Norehan, 2019). The alarming statistics drive clinicians worldwide to develop and implement various preventive strategies through early screening and creating awareness about getting HPV vaccinations to improve women's overall health outcomes (Okunade, 2020; Small *et al*., 2017).

Survival analysis is one of the most common statistical techniques used to estimate the time until an interesting event, such as death, disease relapse, the onset of an adverse reaction, or the emergence of a new disease entity. This method examines time-based data, starting from a specific origin time and ending when a specific endpoint event occurs (Rai *et al.*, 2021). This approach to time-

to-event series analysis was initially developed in the medical field and has since been extended to a range of other fields, such as event history analysis in the social sciences, reliability theory in engineering, and duration analysis in economics (Lánczky & Győrffy, 2021).

Most survival analyses in cancer journals use some or all of the survival analysis methods, such as Kaplan-Meier (KM) plots, log-rank tests, and Cox (proportional hazards) regression. The Cox proportional hazards regression analysis is a method that researchers frequently use to investigate the impacts that several variables may have on the time it takes for a specified event to occur (Tshewang *et al*., 2021). Previous studies showed that the authors applied the Cox proportional hazards model to investigate the impact of COVID-19 on the Mexican population, survival prediction in women newly diagnosed with cervical cancer, and the prognosis of survival rates among several factors in inpatient breast cancer cases in Indonesia (Salinas *et al*., 2020; Matsuo *et al*., 2019; Nadjib *et al*., 2018).

In Malaysia, there has been a scarcity of nationally representative published works on the survival of cervical cancer. Therefore, this study aims to identify the prognostic factors that significantly affect the risk of death in cervical cancer patients receiving treatments and clinically diagnosed at Hospital Universiti Sains Malaysia (HUSM) in Kelantan, Malaysia, using the semi-parametric Cox proportional hazards regression analysis.

MATERIALS AND METHODS

Source of Data

The cervical cancer data were obtained from Hospital Universiti Sains Malaysia (HUSM), located in Kubang Kerian, Kelantan, Malaysia. This hospital serves as a referral facility for Malaysia's East Coast. This study includes all patients who underwent at least one cervical cancer treatment at HUSM and were histopathologically and clinically diagnosed with the disease between 2013 and 2017. Those who passed away for various reasons were excluded. There are 132 patients in total; 70 (53%) of them are deceased, and 62 (47%) are still alive. Ethical approval was obtained from the research ethics committee.

Definition of Key Medical Terms

Histologic type refers to the microscopic classification of cancer cells, such as squamous cell carcinoma or adenocarcinoma, while distant metastasis describes the spread of cancer cells from the original tumor site to distant organs or tissues in the body (Chin *et al*., 2022).

Statistical Analysis

Cox Proportional Hazard Regression Model

The prognostic model for cervical cancer patients is initiated using the Cox proportional hazards regression model. The analysis is conducted to identify significant factors associated with the risk of death in these patients. The proportional hazards model was first introduced by Cox (1972) and is known as the Cox proportional hazards regression model. This model is the most commonly used model in the analysis of survival data. It is also referred to as a semi-parametric model because no specific form of the probability distribution is assumed for the survival time. Similarly, no assumptions have been made regarding the actual form of the baseline hazard function h0(t), making the model more flexible and applicable. The hazard function of the Cox proportional hazards regression model is given by equation 1.

$$h(t) = h_0(t)exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right) \quad (1)$$

where the hazard function $h(t)$ is dependent on a set of p covariates $(x_1, x_2 \ldots, x_p)$, where the impact is measured by the size of the respective coefficients $(\beta_1, \beta_2, \ldots, \beta_p)$.

In the Cox proportional hazards regression model, β −parameters are estimated by maximizing the partial log-likelihood function. Of the n observed survival times, suppose that there are D uncensored times. Let $t_1 < t_2 < \cdots < t_D$ denote the ordered D distinct event times (no ties between the event times). The partial likelihood function of the Cox proportional hazards model may be written in equation 2.

$$L(\beta) = \prod_{i=1}^{D} \frac{exp\left[\sum_{j=1}^{p} \beta_j x_{(i)j}\right]}{\sum_{l \in R(t_i)} exp\left[\sum_{j=1}^{p} \beta_j x_{(i)j}\right]} \tag{2}$$

where $X_{(i)j}$ be the jth covariate associated with the individual whose failure time is t_i. Meanwhile, $R(t_i)$ consists of all individuals whose survival times are at least t_i.

Schoenfeld Residual Tests

Based on the scaled Schoenfeld residuals, the proportional hazards assumption for each variable is examined. In the meantime, the overall model is evaluated using the global Schoenfeld residuals test. If the tests show statistical significance (p-value < 0.05), the proportional hazards assumption is violated. From Collet (2023), the Schoenfeld residual denotes r_{Pji} may be obtained in equation 3 – 5:

$$r_{Pji} = \delta_i \{x_{ji} - \hat{a}_{ji}\} \tag{3}$$

where x_{ji} is the value of the jth explanatory variable, $j = 1, \ldots, p$ for the ith subject in the study,

$$\hat{a}_{ji} = \frac{\sum_{l \in R(t_i)} x_{ji} exp(\hat{\beta}\prime x_1)}{\sum_{l \in R(t_i)} exp(\hat{\beta}\prime x_1)} \tag{4}$$

and $R(t_i)$ is the set of all individuals at risk at the time t. Meanwhile, the scaled Schoenfeld residuals r^*_{Pji} are the components of the vector

$$r^*_{Pji} = r\ var(\hat{\beta}) r_{Pi} \tag{5}$$

where r is the number of deaths among the n individuals, $var(\hat{\beta})$ is the variance-covariance matrix of the parameter estimates, and $r_{Pi} = \left(r_{P_{1i}}, r_{P_{2i}}, \ldots, r_{P_{pi}}\right)'$ is the vector of Schoenfeld residuals for the ith subject.

RESULTS AND DISCUSSIONS

In univariate analysis, each factor is analyzed using the Cox proportional hazards regression model to identify the association between each covariate and the outcome individually. This analysis may also provide a preliminary idea of which variables have possible prognostic importance. Table 1 shows that the stage at diagnosis and primary treatment are statistically significant at the 5% level of significance.

All variables are further analyzed using the multivariate Cox proportional hazards regression analysis. Independent variables that are significantly associated with the hazards of death of the cervical cancer patients under study are selected based on the stepwise selection method, which involves systematically adding or removing variables based on specific criteria. Forward selection begins with no variables and adds them one at a time based on statistical significance, while backward elimination starts with all candidate variables and removes the least significant ones. This selection method consists of forward selection followed by the backward elimination process, with a p-value < 0.05 for variable entry and a p-value > 0.10 for variable removal (Collet, 2023). At this step, the preliminary main effects model is obtained. In the multivariate analysis, histologic type, stage at diagnosis, and distant metastasis are found to be statistically significant.

Table 1. The univariate Cox proportional hazards regression model**.**

Variables	Coefficient	Crude HR (95% CI)	LR (df)	*p*-value
Ethnicity				
Non-Malay				
Malay	-0.1653	0.8476 (0.4312-1.6660)	0.22(1)	0.6382
Lymph node involvement				
Negative				
Positive	-0.08687	0.9168 (0.5218-1.6113)	0.09(1)	0.7609
Histologic type				
Squamous cell carcinoma				
Adeno cell carcinoma	0.3219	1.3797 (0.8013-2.3760)	1.28(1)	0.2577
Age at diagnosis				
<40				
40 – 59	-0.4092	0.6642 (0.3194-1.3813)		
≥60	-0.6155	0.5404 (0.2277-1.2832)	2.97(2)	0.3962
Stage at diagnosis				
I– II				
III – IV	1.0176	2.7665 (1.4272-5.3621)	9.6(1)	0.0080
Primary Treatment				
Surgery				
Non-surgery	0.6360	1.889 (1.077-3.3123)	5.37(1)	0.0205
Distant Metastasis				
No				
Yes	0.3412	1.4143 (0.8001-2.3214)	1.18(1)	0.2769

Note: HR = Hazard Ratio, CI = Confidence Interval, SE = Standard Error, LR = Likelihood Ratio, df = degrees of freedom

The interaction between the covariates in the preliminary main effects model is checked by adding the interaction term to the model. The interaction is not statistically significant at the 5% level of significance. The preliminary final model is obtained after checking for the interaction. The results are presented in Table 2. It has been found that the stage at diagnosis, histologic type, and distant metastasis are significant prognostic factors that influence the risk of dying for these patients.

Table 2. The multivariate Cox proportional hazards regression model.

Variables	Coefficient	SE	Adjusted HR (95% CI)	*p*-value
Histologic type				
Squamous cell carcinoma				
Adeno cell carcinoma	0.5946	0.2897	1.8123(1.027-3.198)	0.0401
Stage at diagnosis				
I-II				
III-IV	0.8643	0.2700	2.3734(1.398-4.029)	0.0014
Distant Metasis				
No				
Yes	0.7211	0.2636	2.056(1.227-3.448)	0.0062

Note: HR = Hazard Ratio, CI = Confidence Interval, SE = Standard Error

This study found that the prognosis of cervical cancer depends significantly on the stage at diagnosis. Patients diagnosed with advanced stages (stage III-IV) of cancer have a higher risk of death than those with early stages (stage I-II). This finding is also concurred with findings from other research (Seifu *et al*., 2022; Mebratie *et al*., 2022; Manzour *et al*., 2022; World Health Organization, 2020; Juhan *et al*., 2013). This study also found that histologic type significantly affects survival, as patients

diagnosed with adenocarcinoma are identified to have a higher risk of dying compared to squamous cell carcinoma, which is supported by previous findings from other research. For example, Liu *et al*. (2022) found that adenocarcinoma had a significant negative impact on the prognosis of cervical cancer patients studied.

Furthermore, Li *et al*. (2022) claimed that adenocarcinoma was associated with a worse prognosis because those diagnosed with this histologic type were detected later and at more advanced stages than those with squamous cell carcinoma. As for distant metastasis, the results indicated that cervical cancer patients with distant metastasis were twice as likely to die compared to those without it. This is in line with other studies, where cervical cancer patients with metastasis have a poor survival rate (Singh *et al.*, 2023; Zhou & Peng, 2020; Zhang *et al*., 2020).

The proportional hazards assumption for each variable in the preliminary final model is tested based on the scaled Schoenfeld residuals. The global Schoenfeld residuals test is used to assess the assumption for the overall model. These tests are performed using the function cox.zph from the survival package (Therneau, 2014) in R software. The result is tabulated in Table 3. The global test indicates that the proportional hazards assumption for the overall model is violated. Meanwhile, the scaled Schoenfeld residuals test shows that the hazard for the distant metastasis variable is not proportional (p-value < 0.05). Therefore, the stratified Cox model might be adopted in future studies. The stage at diagnosis and histologic type remain significant prognostic factors in the model. Meanwhile, distant metastasis can be considered a stratification factor in the future stratified Cox model study.

Table 3. The proportional hazards assumption test results for the preliminary final model.

Variables	**ρ (rho)**	**2**	***p*-value**
Histologic type	0.0712	0.373	0.5416
Stage	-0.1434	1.336	0.2477
Distant metastasis	0.2934	6.591	0.0102
Global	NA	7.956	0.0419

Note: ρ (rho) = Correlation between scaled Schoenfeld residuals and time, χ^2 = Chi-square statistic

CONCLUSION

In this study, prognostic factors significantly affecting the risk of death in cervical cancer patients treated at HUSM are identified using Cox proportional hazards regression analysis. The findings indicate that cervical cancer patients treated at HUSM with stage III-IV adeno cell carcinoma are at the greatest risk of death from cervical cancer. Since the proportional hazard assumption was violated for distant metastasis in the Schoenfeld residuals tests, a stratified Cox regression model should be considered for future study. The results of the analysis will provide a better understanding of the survivorship and hazards facing cervical cancer patients in Malaysia.

ACKNOWLEDGEMENT

The authors are very thankful to the late Dr. Nuradhiathy Abd Razak and Universiti Sains Malaysia Hospital for the data. This research was funded by a grant from the Ministry of Higher Education of Malaysia (FRGS Grant Project Code: FRGS/1/2023/STG06/ UMS/02/2).

REFERENCES

Chin, S.S., Jamonek Jamhuri, N.A., Hussin, N., Md Zubir, N.L., Tan, J.R. & Chan, S.C.W. 2022. Factors influencing pap smear screening uptake among women visiting outpatient clinics in Johor. *Malaysian Family Physician*, 17(2): 46-55.

Collett, D. 2023. *Modelling survival data in medical research*. New York: CRC Press.

Cox, D.R. 1972. Regression models and life-tables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 34(2): 187-202.

Juhan, N., Abd Razak, N., Zubairi, Y.Z., Khattak, M. N. & Naing, N. N. 2013. Survey of patients with cervical cancer in Hospital Universiti Sains Malaysia: survival data analysis with time-dependent covariate. *Iranian Journal of Public Health*, 42(9): 980-987.

Lánczky, A. & Győrffy, B. 2021. Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. *Journal of Medical Internet Research*, 23(7): e27633.

Li, J., Liu, G., Luo, J., Yan, S., Ye, P., Wang, J. & Luo, M. 2022. Cervical cancer prognosis and related risk factors for patients with cervical cancer: a long-term retrospective cohort study. *Scientific Reports*, 12(1): 13994.

Liu, P., Ji, M., Kong, Y., Huo, Z., Lv, Q., Xie, Q. & Chen, C. 2022. Comparison of survival outcomes between squamous cell carcinoma and adenocarcinoma/adenosquamous carcinoma of the cervix after radical radiotherapy and chemotherapy. *BMC Cancer*, 22(1): 326.

Manzour, N., Chiva, L., Chacón, E., Martin-Calvo, N., Boria, F., Minguez, J. A. & Alcazar, J. L. 2022. SUCCOR risk: design and validation of a recurrence prediction index for early-stage cervical cancer. *Annals of Surgical Oncology*, 29(8): 4819-4829.

Matsuo, K., Purushotham, S., Jiang, B., Mandelbaum, R. S., Takiuchi, T., Liu, Y. & Roman, L. D. 2019. Survival outcome prediction in cervical cancer: cox models vs deep-learning model. *American Journal of Obstetrics and Gynecology*, 220(4): 381.e1-381.e14.

Mebratie, A. E., Moges, N. A., Meselu, B. T. & Melesse, M. F. 2022. Time to death from cervical cancer and predictors among cervical cancer patients in Felege Hiwot Comprehensive Specialized Hospital, North West Ethiopia: facility-based retrospective follow-up study. *PLoS One*, 17(6): e0269576.

Mustafa, W.A., Halim, A., Nasrudin, M.W. & Shakir, K. 2022. Cervical cancer situation in Malaysia: A systematic literature review. *BIOCELL*, 46(2): 367-381.

Nadjib Bustan, M., Arman, Aidid, M. K., Gobel, F. A. & Syamsidar. 2018. Cox proportional hazard survival analysis to inpatient breast cancer cases. *Journal of Physics: Conference Series*, 1028: 012230.

Norkhafizah Saddki, M. & Norehan Mokhtar, M. 2019. Cervical cancer and pap smear screening: knowledge, attitude and practice among working women in northern state of Malaysia. *The Medical Journal of Malaysia*, 74(1), 8-14.

Okunade, K.S. 2020. Human papillomavirus and cervical cancer. *Journal of Obstetrics and Gynaecology*, 40(5): 602-608.

Rai, S., Mishra, P. & Ghoshal, U.C. 2021. Survival analysis: A primer for the clinician scientists. *Indian Journal of Gastroenterol*, 40: 541-549.

Salinas-Escudero, G., Carrillo-Vega, M.F. & Granados-García, V. 2020. A survival analysis of COVID-19 in the Mexican population. *BMC Public Health*, 20: 1616.

Seifu, B., Fikru, C., Yilma, D. & Tessema, F. 2022. Predictors of time to death among cervical cancer patients at Tikur Anbesa specialized hospital from 2014 to 2019: A survival analysis. *PLoS One*, 17(2): e0264369.

Singh, D., Vignat, J., Lorenzoni, V., Eslahi, M., Ginsburg, O., Lauby-Secretan, B., Arbyn, M., Basu, P., Bray, F. & Vaccarella, S. 2023. Global estimates of incidence and mortality of cervical cancer in 2020: a baseline analysis of the WHO Global Cervical Cancer Elimination Initiative. *The Lancet. Global Health*, 11(2): e197–e206.

Small Jr., W., Bacon, M. A., Bajaj, A., Chuang, L. T., Fisher, B. J., Harkenrider, M. M. & Gaffney, D. K. 2017. Cervical cancer: a global health crisis. *Cancer*, 123(13): 2404-2412.

Therneau, T. M. 2014. Cox models and "type III" tests (Technical Report). Mayo Clinic. https://cran.r-project.org/web/packages/survival/vignettes/survival.pdf

Tshewang, U., Satiracoo, P. & Lenbury, Y. 2021. Survival analysis of cervical cancer patients: a case study of Bhutan. *Asian Pacific Journal of Cancer Prevention*, 22(9): 2987-2993.

World Health Organization. 2020. *Malaysia Factsheet*. IARC Global Cancer Observatory-GLOBOCAN database. http://gco.iarc.fr/today/data/factsheets/populations/458-malaysia-fact-sheets.pdf

Zhang, J., Qin, L., Chen, H.M., Hsu, H.C., Chuang, C.C., Chen, D. & Wu, S.Y. 2020. Overall survival, locoregional recurrence, and distant metastasis of definitive concurrent chemoradiotherapy for cervical squamous cell carcinoma and adenocarcinoma: before and after propensity score matching analysis of a cohort study. *American Journal of Cancer Research*, 10(6): 1808.

Zhou, S. & Peng, F. 2020. Patterns of metastases in cervical cancer: a population-based study. *International Journal of Clinical and Experimental Pathology*, 13(7): 1615.