

COMPARISON OF STRATIFIED WEIBULL MODEL AND WEIBULL ACCELERATED FAILURE TIME (AFT) MODEL IN THE ANALYSIS OF CERVICAL CANCER SURVIVAL

Nurliyana Juhan^{a*}, Nuradhiathy Abd Razak^b, Yong Zulina Zubairi^c, Nyi Nyi Naing^d, Che Haziqah Che Hussin^a, Muhamad Azlan Daud^a

^aPreparatory Centre for Science and Technology, Universiti Malaysia Sabah, 88400 Kota Kinabalu, Sabah, Malaysia

^bGraduate Studies, University of Malaya, 50603 Kuala Lumpur, Malaysia

^cCentre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia

^dUnit of Biostatistics and Research Methodology, School of Medical Sciences, Health Campus, Universiti Sains Malaysia, Kelantan, Malaysia

Article history

Received

27 July 2015

Received in revised form

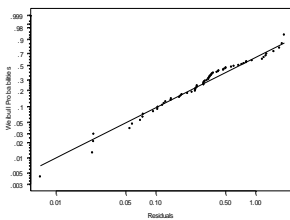
23 December 2015

Accepted

19 January 2016

*Corresponding author
liyana87@ums.edu.my

Graphical abstract



Abstract

Cervical cancer is the fourth most common cancer affecting women worldwide, after breast, colorectal, and lung cancers with 528 000 new cases every year. It is also the fourth most common cause of cancer death with 266 000 deaths in 2012 among women worldwide. In Malaysia it remains to be a great concern among clinicians; yet published works on survival of cervical cancer patients are somewhat limited. In this study, two survival regression models which are parametric Stratified Weibull model and Weibull Accelerated Failure Time (AFT) model are considered as the alternative and improvement of the well-known Cox proportional hazard model to evaluate the prognostic factor that effect on survival of patients with cervical cancer. Comparisons were made to find the best model. Data were taken from Hospital University Science Malaysia (HUSM) over a period of 12 years. From the analyses it was found that the AFT model was the most appropriate. The AFT model has shown that the median survival time for patient at stage III & IV (14 months) is about one third that of those at stages I & II (40 months) for the same distant metastasis group. While, the median survival time for patient with distant metastasis (17 months) is half that of those without distant metastasis (34 months) for the same stage group.

Keywords: Accelerated Failure Time (AFT), cervical cancer, prognostic factor, stratified weibull, survival

© 2016 Penerbit UTM Press. All rights reserved

1.0 INTRODUCTION

Cervical cancer is the fourth most common cancer affecting women worldwide, after breast, colorectal, and lung cancers with 528 000 new cases every year [1]. It is also the fourth most common cause of cancer death with 266 000 deaths in 2012 among women worldwide [1]. It remains to be one of the major cancer that burdens worldwide particularly in under-

developed and developing countries [2]. Cervical cancer develops from cells that changes caused by the virus called the Human Papilloma virus (HPV) which is predominantly transmitted through sexual intercourse [3].

In Malaysia, cervical cancer is the fifth most common cancer to occur, but ranked second among female-related cancers [4]. The age standardized incidence (ASR) for cancer of the cervix was 7.8 as

per 100,000 populations and there were about 847 cases registered with National Cancer Registry Malaysia in 2007 [4]. The burden of cervical cancer in Malaysia is comparable with other developing countries such as Thailand, Myanmar, Vietnam, India and some countries in South America [5].

The incidence rate increased with age after 30 years and has its peak at ages 65-69 years [4]. According to ethnicity in Malaysia, Indian women had the highest incidence for cervical cancer followed by Chinese and Malay [4]. In Malaysia, it remains to be a great concern among clinicians; yet published works on survival of cervical cancer patients are somewhat limited. Latest information on cancer survival is important as reference material for clinicians, oncologists, epidemiologists and scientists involved in clinical work, medical auditing or research [6].

In published studies, Cox proportional hazard regression model was frequently used when examining the relationship of the survival distribution to covariates [7]-[9]. This is perhaps due to the fact that although baseline hazard is not specified in Cox model, the parameter can still be estimated. When the hazard assumption is not satisfied, stratified Cox model which is a modification of the Cox proportional hazard model that allows for control by stratification of a predictor that does not satisfy the proportional hazard assumption is used instead [10].

But in some cases, parametric models are more informative than the Cox model such as the baseline hazard and survival estimates are known. Several examples of parametric models can be found in the following studies [11]-[15]. Previous study has compared the Cox and Weibull model in modeling the gastric cancer data and found that the Weibull model gave more a precise results in than the Cox model [16].

In this study, two survival regression models which are parametric Stratified Weibull model and Weibull Accelerated Failure Time (AFT) model are considered to evaluate the prognostic factor that affect on survival of patients with cervical cancer. Comparisons were made to find the best model. A stratified Weibull model was used since there was a time-dependent covariate that caused the proportional hazard assumption violated. While the accelerated failure time (AFT) model was presented as an alternative to the Cox regression model in the analysis of time to event data.

2.0 MATERIALS AND METHODS

The cervical cancer data was taken from Hospital University Science Malaysia (HUSM). The data comprise of 120 patients who have been diagnosed as cervical cancer and obtained treatment from the hospital from 1st of July 1995 to 30th of June 2007. Patients who were died due to other competing causes of death (not cervical cancer), or with incomplete data were excluded from this study.

The variable time (in months) is the survival time which was measured from the patient was diagnosed with cervical cancer up to the time of death. Therefore, time was considered as the variable of interest. In the analysis, other variables that were considered were ethnicity, lymph node involvement, distant metastasis, histology, primary treatment, stage and age at diagnosis. Statistical package TIBCO Spotfire S-Plus ver 8.1 was used to perform data analysis.

By using log-cumulative hazard plot, the suitability of the Weibull model for the data was assessed. Univariate analysis was conducted using the simple Weibull regression analysis to identify the significant prognostic factors individually. Significant factors from the univariate analysis were then further analyzed by the Weibull multivariate analysis to model the prognostic factors. Forward variable selection method was used where variables added to a model one at a time and selection ends when the next term for inclusion ceases to be significant at a pre-assigned level $\alpha=0.10$.

The test based on Schoenfeld or known as the cox.zph test in S-Plus was applied to assess the proportional hazards assumption [17]. As the proportional hazard assumption was not satisfied, a stratified Weibull model [10] and Weibull AFT model were used instead. For the stratified Weibull model, the strata divide the subjects into disjoint groups, each of which has a distinct baseline hazard function but common values for the coefficient vector β [17]. It is assumed that the effect of every covariate is constant across strata. Under AFT model the direct effect of the explanatory variables on the survival time are measure instead of hazard.

To assess the goodness of fit of the model, Martingale residual was performed for the stratified Weibull model and the standardized residual plots for the Weibull AFT model. Figure 1 shows a flow chart which summarizes the steps of the statistical analysis performed in this study.

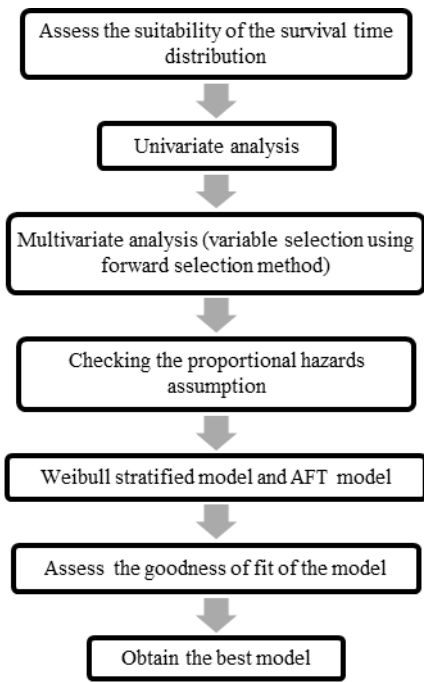


Figure 1 Flow chart of the study

3.0 RESULTS

The log-cumulative hazard plot in Figure 2 shows that the dots in the plot form a straight line. This indicates that Weibull model is suitable to fit the data.

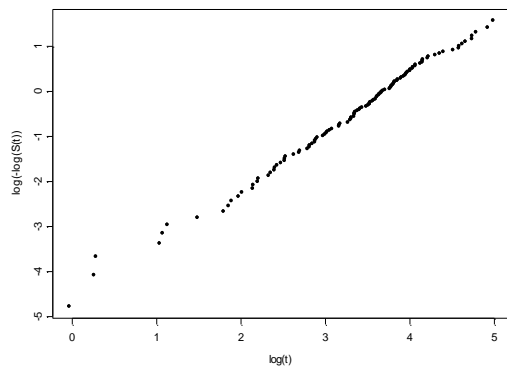


Figure 2 The log-cumulative hazard plot

The characteristics of cervical cancer patients treated in HUSM were shown in Table 1. From the univariate analysis, three factors were identified significant namely, stage ($p < 0.001$), distant metastasis ($p = 0.0005$) and primary treatment ($p = 0.009$). The three potential factors were then further analyzed by the forward selection procedure in multivariate analysis, two variables namely stage and distant metastasis were found to be significant factors ($p < 0.001$).

Table 1 Characteristics of cervical cancer patients treated in HUSM

Variables	Description	No. of patients	(%)
Ethnicity	Non malay	21	17.5
	Malay	99	82.5
Lymph node involve	Negative	89	74.2
	Positive	31	25.8
Distant metastasis	without distant metastasis	83	69.2
	with distant metastasis	37	30.8
Histologic type	Squamous cell carcinoma	93	77.5
	Adeno cell carcinoma	27	22.5
Stage	I & II	89	74.2
	III & IV	31	25.8
Primary Treatment	Surgery	40	33.3
	Radiotherapy & chemotherapy	80	66.7
Age at diagnosis	< 40	2	1.7
	40-59	97	80.8
	≥ 60	21	17.5

From the cox.zph test, it was noted that distant metastasis ($p = 0.032$) did not satisfy the proportional hazard assumption and a stratified model was then considered. The first stratum consists of patients without distant metastasis and the second stratum were patients who do have distant metastasis.

Parameter estimates of the stratified Weibull model were performed in order to obtain the hazard ratio for stage variable as shown in Table 2. Hazard of death for patient at stage III & IV without distant metastasis is 2.3 times greater than patient at stage I & II. While hazard of death for patient at stage III & IV with distant metastasis is 3.5 times greater than patient at stage I & II.

Table 2 Hazard ratio for variable stage stratified on distant metastasis

Stratum	Coefficient (β)	Hazard ratio (ψ)
1=without distant metastasis	0.832	2.30
2=with distant metastasis	1.261	3.53

Then Martingale residual was performed for the stratified Weibull model. Figure 3 shows that most of the residuals are between -50 to 50. According to [17], for stratified model, the best model need not have the smallest sum of square (ss) Martingale residual. No influential observation was seen in the plot. The Martingale residuals were skewed. Those near to -50 have long survival time and those near to 50 died too soon. The relationship looks reasonably linear.

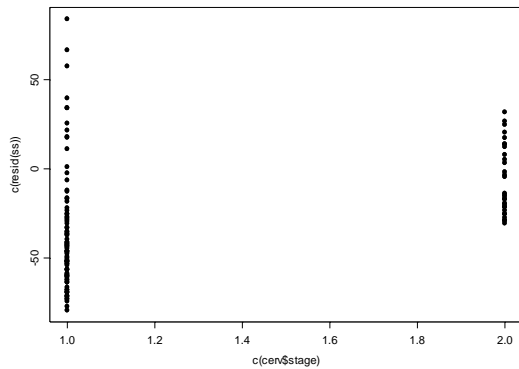


Figure 3 Martingale residual for the stratified Weibull model

The AFT model was applied to the previous model which consists of variable stage and distant metastasis. The data set was fitted using Weibull AFT model.

Table 3 Output of Weibull AFT model

Variable	Value
(Intercept)	4.656898
(Scale)	0.9437051
Stage	-0.8422007
Distant metastasis	-0.6478107

The hazard function and median survival time can be obtained from the output in Table 3. The first value appears in the output is the intercept, $\mu = 4.656898$ and scale, $\sigma = 0.9437051$. The survival time under Weibull AFT model has a $W(ae^{\eta_i b}, b)$ distribution. The parameter a, b, β_j in the Cox model can be expressed by σ, α_j .

$$a = \exp\left(-\frac{\mu}{\sigma}\right) = 0.00719 \tag{1}$$

$$b = \frac{1}{\sigma} = 1.059 \tag{2}$$

$$\beta_j = -\frac{\alpha_j}{\sigma} \tag{3}$$

Where a is the scale parameter and b is the shape parameter. The shape parameter, b is greater than one, thus the hazard is decreasing over time. Under the Weibull AFT model the hazard of death at time t for the i -th patient is

$$h_i(t) = e^{\beta x_i} h_0(e^{\beta x_i} t) \tag{4}$$

where x_i takes the value zero if the i -th patient is at stage I & II and unity if the patient is at stage III & IV. For the Weibull distribution, the baseline hazard function is

$$h_0(t) = abt^{b-1} \tag{5}$$

which is the hazard function for patient at stage I & II. Hence,

$$h_i(t) = (e^{\eta_i})^b abt^{b-1} \tag{6}$$

where $\eta_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_p x_{pi}$ is the linear component in which x_{ji} is the explanatory variables $j = 1, 2, \dots, p$ for i -th individual, $i = 1, 2, \dots, n$.

On fitting this model, for the variable stage, the estimated values of the parameters are given by $a = 0.00719$, $b = \frac{1}{\sigma} = 1.0597$ and $\beta = -\frac{\alpha}{\sigma} = \frac{-0.8422007}{0.9437051} = 0.8924$. The accelerated factor, ϕ of AFT model is estimated by $e^{\beta} = e^{0.892441} = 2.441$. The survival time of a patient at stage III & IV for the same distant metastasis group is therefore accelerated by a factor of 2.4 under this model. The median survival time under the Weibull AFT model is

$$t(50) = \left[\frac{\log 2}{\alpha e^{\beta} \delta x_i} \right]^{1/b} \tag{7}$$

The estimated median survival time for patient at stage I & II ($x_i = 0$) is 40 months, while that for patient at stage III & IV ($x_i = 1$) is 14 months. The median survival time for patient at stage III & IV is therefore about one third that of those at stage I & II for the same distant metastasis group.

Next, for the variable distant metastasis under the Weibull AFT model, the hazard of death at time t for the i -th patient is

$$h_i(t) = e^{\beta x_i} h_0(e^{\beta x_i} t) \tag{8}$$

where x_i takes the value zero if the i -th patient without distant metastasis and unity if the patient with distant metastasis. The baseline hazard function given as

$$h_0(t) = abt^{b-1} \tag{9}$$

which is the hazard function for patient without distant metastasis.

The estimated values of the parameters are given by $a = 0.00719$, $b = \frac{1}{\sigma} = 1.0597$ and $\beta = -\frac{\alpha}{\sigma} = 0.68645$. The accelerated factor, ϕ of AFT model is estimated by $e^{\beta} = e^{0.68645} = 1.987$. The survival time of a patient with distant metastasis for the same stage is therefore accelerated by a factor of about 1.98 under this model.

The estimated median survival time for patient without distant metastasis ($x_i = 0$) is 34 months, while that for patient with distant metastasis ($x_i = 1$) is 17 months. The median survival time for patient with distant metastasis is therefore half that of those without distant metastasis for the same stage group.

To assess the goodness of fit of the Weibull AFT model, standardized residual plot and probability plot using the Kaplan Meier against standardized residual were performed.

Plot of standard residual against the fitted model in Figure 4 does not indicate the presence of any outlier. The values of sqrt (residuals) are small.

Figure 5 shows a probability plot of Weibull AFT model using the Kaplan Meier against standardized residual which gives a straight line. This suggests that the Weibull AFT model fit the data well.

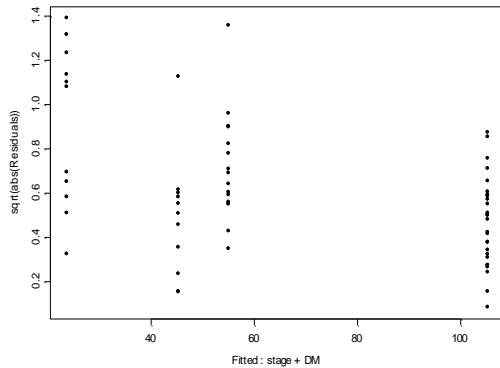


Figure 4 Residual plot for Weibull AFT model

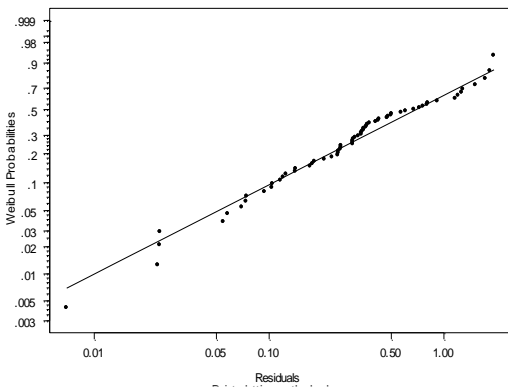


Figure 5 Probability plot for residual of Weibull AFT model

4.0 DISCUSSION

Published studies in the field of medical sciences are often interested in Cox proportional hazard model instead of parametric models. However, in a review of survival analyses in cancer journals, it was found that only 5 percent of all studies using the Cox proportional hazard model with respect to checking the underlying assumptions [11]. If this assumption does not hold, the Cox model can lead to the unreliable conclusions. Therefore, parametric stratified model became one of the solutions.

As mentioned earlier, the main objective of this study is to obtain a survival model for cervical cancer patients in a hospital. In particular, two methods are applied on the analysis namely parametric Stratified

Weibull model and Accelerated Failure Time (AFT) model.

In order to identify a set of explanatory covariates that have the potential for being included in the model, forward selection procedure has been carried out. It is found that the most feasible model consists two covariates namely stage and distant metastasis are significant.

However, the proportional hazard assumption for distant metastasis was not satisfied. Therefore, we apply the parametric stratified model. It is found that the baseline hazard is Weibull with shape parameter, $b=0.900901$ and scale parameter, $a=0.019229$ for stratum 1 (without distant metastasis) and shape parameter, $b=1.364256$ and scale parameter, $a=0.002520$ for stratum 2 (with distant metastasis). This study found that patients who were diagnosed at stage III & IV have greater risk of death compared to those who were diagnosed at early stage, stage I & II for both stratum, with and without metastasis.

This finding is similar to a study of 515 cervical cancer patients in [18] that showed a significant result for advance stage III & IV with adjusted hazard ratio of 1.54 (95% CI= 1.11- 2.14), indicating that patients with advanced stage of disease had a 54% higher risk of progression or death at any time than earlier stage patients. Similar findings were also obtained by other studies [19]-[22]. The advantage of stratified model is that it handles variables that do not satisfy the proportional hazard assumption and define the baseline hazard. However, the disadvantage is that we cannot obtain an estimated coefficient of the categorical variable effect [10]. Moreover, the sample size become smaller since it was divided into 2 strata and stratification do affects parameter estimates.

Finally, the accelerated failure time (AFT) model is presented as an alternative to the proportional hazard model. Unlike the proportional hazard model, the AFT model measured the direct effect of the explanatory variables on the survival time instead of hazard. It is found that the median survival time for patient at stage III & IV (14 months) is about one third that of those at stage I & II (40 months) for the same distant metastasis group. While, the median survival time for patient with distant metastasis (17 months) is half that of those without distant metastasis (34 months) for the same stage group. From the analyses and the results obtained, the AFT model is seen to be a more appropriate modeling framework and has the added advantage of being easier to interpret. In comparison to the stratified Weibull model, the Weibull AFT model is a better model.

Acknowledgement

The authors thank the Hospital Universiti Sains Malaysia (HUSM) for providing the data.

References

- [1] Ferlay, J., I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, F. Bray. 2013. *GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11*. Lyon, France: International Agency for Research on Cancer. [Online]. From: <http://globocan.iarc.fr>, accessed on day/month/year. [Accessed on 5 December 2014].
- [2] Bray, F., J. S. Ren, E. Masuyer, J. Ferlay. 2013. Global Estimates Of Cancer Prevalence For 27 Sites In The Adult Population In 2008. *Int J Cancer*. 132(5): 1133-1145.
- [3] Berrington de Gonzalez, A. 2007. Comparison Of Risk Factors For Invasive Squamous Cell Carcinoma And Adenocarcinoma Of The Cervix: Collaborative Reanalysis Of Individual Data On 8,097 Women With Squamous Cell Carcinoma And 1,374 Women With Adenocarcinoma From 12 Epidemiological Studies. *International Journal of Cancer*. 120(4): 885-889.
- [4] Omar, Z. A., N. Ibrahim. 2011. *National Cancer Registry Report 2007*. Malaysia Cancer Statistics-Data and Figure. Kuala Lumpur: National Cancer Registry, Ministry of Health Malaysia.
- [5] Domingo, E. J., R. Noviani, M. R. Noor, C. A. Ngelangel, K. K. Limpaphayom, T. V. Thuan. 2008. Epidemiology And Prevention Of Cervical Cancer In Indonesia, Malaysia, The Philippines, Thailand And Vietnam. *Vaccine*. 26(12): 71-79.
- [6] Talbäck, M., M. Stenbeck, M. Rosén. 2004. Up-To-Date Long-Term Survival Of Cancer Patients: An Evaluation Of Period Analysis On Swedish Cancer Registry Data. *Eur J Cancer*. 40: 1361-1372.
- [7] Pomros, P., S. Sriamporn, V. Tangvoraphonkchai, S. Kamsa-Ard, K. Poomphakwaen. 2007. Factors Affecting Survival Of Cervical Cancer Patients Treated At The Radiation Unit Of Srinagarind Hospital, Khon Kaen University, Thailand. *Asian Pac J Cancer Prev*. 8: 297-300.
- [8] Ho, K. C., C. C. Wang, J. T. Qiu, C. H. Lai, J. H. Hong, Y. T. Huang. 2011. Identification Of Prognostic Factors In Patients With Cervical Cancer And Supraclavicular Lymph Node Recurrence. *Gynecol Oncol*. 123(2): 253-256.
- [9] Kim, T. E., B. J. Park, H. S. Kwack, J. Y. Kwon, J. H. Kim, S. C. Yoon. 2012. Outcomes And Prognostic Factors Of Cervical Cancer After Concurrent Chemoradiation. *J Obstet Gynaecol Res*. 38(11): 1315-1320.
- [10] Kleinbaum, D. G. 1995. *Survival Analysis-A Self-Learning Text*. New York: Springer-Verlag.
- [11] Pourhoseingholi, M. A., E. Hajizadeh, B. M. Dehkordi, A. Safaee, A. Abadi, M. R. Zali. 2007. Comparing Cox Regression And Parametric Models For Survival Of Patients With Gastric Carcinoma. *Asian Pac J Cancer Prev*. 8: 412-416.
- [12] Moghimi-Dehkordi, B. A., Safaee, M. A. Pourhoseingholi, R. Fatemi, Z. Tabeie, M. R. Zali. 2008. Statistical Comparison Of Survival Models For Analysis Of Cancer Data. *Asian Pacific J Cancer Prev*. 9: 417-420.
- [13] Ding, S., S. J. Soong, H. Y. Lin, R. Desmond, C. M. Balch. 2009. Parametric Modeling Of Localized Melanoma Prognosis And Outcome. *J Biopharm Statistics*. 19: 732-747.
- [14] Nakhaee, F., M. Law. 2011. Parametric Modelling Of Survival Following HIV And AIDS In The Era Of Highly Active Antiretroviral Therapy: Data From Australia. *East Mediterr Health*. 17(3): 231-237.
- [15] Ravangard, R., M. Arab, A. Rashidian, A. Akbarisari, A. Zare, H. Zeraati. 2011. Comparison Of The Results Of Cox Proportional Hazards Model And Parametric Models In The Study Of Length Of Stay In A Tertiary Teaching Hospital In Tehran, Iran. *Acta Med Iran*. 49(10): 650-658.
- [16] Zhu, H. P., X. Xia, C. H. Yu, A. Adnan, S. F. Liu, Y. K. Du. 2011. Application Of Weibull Model For Survival Of Patients With Gastric Cancer. *BMC Gastroenterology*. 11: 1.
- [17] Therneau, T. M., P. M. Grambsch. 2000. *Modeling Survival Data. Statistics for Biology and Health*. New York: Springer-Verlag.
- [18] Dueñas-González, A., M. Orlando, Y. Zhou, M. Quinlivan, H. Barraclough. 2012. Efficacy Inhigh Burden Locally Advanced Cervical Cancer With Concurrent Gemcitabine And Cisplatin Chemoradiotherapy Plus Adjuvant Gemcitabine And Cisplatin: Prognostic And Predictive Factors And The Impact Of Disease Stage On Outcomes From A Prospective Randomized Phase III Trial. *Gynecol Oncol*. 126: 334-340.
- [19] Monk, B. J., C. Tian, P. G. Rose, R. Lanciano. 2007. Which Clinical/Pathologic Factors Matter In The Era Of Chemoradiation As Treatment For Locally Advanced Cervical Carcinoma? Analysis Of Two Gynecologic Oncology Group (GOG) Trials. *Gynecol Oncol*. 105: 427-433.
- [20] Juhan, N., N. Abd Razak, Y. Z. Zubairi, M. N. Khattak, N. N. Naing. 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: 980-987.
- [21] Chen, J., O. K. Macdonald, D. K. Gaffney. 2008. Incidence, Mortality, And Prognostic Factors Of Small Cell Carcinoma Of The Cervix. *Obstet Gynecol*. 111: 1394-1402.
- [22] Wang, K. L., T. C. Chang, S. M. Jung, C. H. Chen, Y. M. Cheng, H. H. Wu. 2012. Primary Treatment And Prognostic Factors Of Small Cell Neuroendocrine Carcinoma Of The Uterine Cervix: A Taiwanese Gynecologic Oncology Group Study. *Eur J Cancer*. 48: 1484-1494.