

RESEARCH ARTICLE

Application of the Weibull Distribution with a Non-Constant Shape Parameter for Identifying Risk Factors in Pharyngeal Cancer Patients

Haleh Aghamolaey¹, Ahmad Reza Baghestani^{2*}, Farid Zayeri¹

Abstract

Background: In its standard form, the parametric survival model assumes that the shape parameter is constant and the scaling parameter is not. This article focuses on how a model with a non-constant shape parameter could make differences in oncology studies and lead to more precise results. **Materials and Methods:** Online data for part of a large clinical trial conducted by the Radiation Oncology Group in the United States available online on UMass Amherst's website were employed. The full study included patients with squamous cell carcinoma from fifteen sites in the mouth and throat, although only data on three sites in the oropharynx reported by the six largest institutions were considered here. To identify clinical, pathological and biological characteristics of patients which might have had an effect on their survival, we compared Weibull distributions once with a constant shape parameter and again with a non-constant shape parameter. Analyses were performed using SAS university edition. The level of significance was set at $P \leq 0.05$. **Results:** Based on the model with a constant shape parameter only the patient status was identified as a risk factor and the AIC of this model was 2152.4, but based on the model with a non-constant shape parameter, sex, patient status, stage of the tumor and the institute at which the patient had been treated were significant, with an AIC of 2150.1. **Conclusion:** On the basis of the AIC, the second model with a non-constant shape parameter was suggested to be more accurate for identifying risk factors, leading to more precise results.

Keywords: Non-constant shape parameter- pharynx cancer- risk factors- survival analysis- Weibull distribution

Asian Pac J Cancer Prev, **18** (6), 1537-1542

Introduction

Medical researchers are greatly interested in studying the survival of patients with cancer and it is criticized for them to recognize the effective characteristics of patients that might have an effect on patient survival. For this reason, choosing and using the appropriate model to accurately determine and estimate effective characteristics is really important. Survival analysis is a set of statistical procedures for data analysis for which the result variable of interest is the time until an event occurs. By event we mean death, disease incidence, relapse of remission, recovery (eg, return to work) or any designated experience of interest that may occur to an individual (Klein and Moeschberger, 1977). One of the most important characteristics of survival data is the presence of censored observation, which are in fact either lost during follow-up or are cases that do not last in the study during the Period of study. In addition, for some results, there may be cases that are never confronted with the event (Lambert et al., 2007). Different methods are used in the analysis of survival and parametric survival model is one of them. It is shown that

in specific circumstances, parametric models especially with non-constant parameters have estimates that are more efficient than regular models (Oakes, 1977; Efron, 1977). Pharyngeal cancer is a disease that represents about a quarter of the cancer of the upper aero digestive tract (Shedd et al., 1968). It has been found to be a disease of the middle and older groups (Smith et al., 1963).

Materials and Methods

Our data and patient information are part of a large clinical trial conducted by the Radiation Therapy Oncology Group in the United States. The full study included patients with squamous cell carcinoma of fifteen sites in the mouth and throat, with sixteen participating institutions, although only data on three sites in the oropharynx reported by the six largest institutions are considered here. Patients included in the study were randomly assigned to one of two treatment groups, radiotherapy alone or radiotherapy with a chemotherapeutic agent. Deaths due to pharyngeal cancer were considered a failure and the survival time was calculated as the time interval between the date of cancer

¹Department of Biostatistics, ²Physiotherapy Research Center, Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Qods Square, Darband Street, Tehran, Iran. *For Correspondence: baghestani.ar@gmail.com

diagnosis and the date of death due to pharyngeal cancer. This study included measurements of many variables that would be expected relate to the survival experience. Six of these variables are given in the data (Sex, Stage T, Stage N, Age, General Condition, and Rank). The site of the primary tumor and possible differences between participants Institutions require an examination as well. The intermediate classification T, N gives a measure of the extent of the tumor in the primary site and at the regional lymph nodes. T = 1, refers to a small primary tumor, two centimeters or less in the largest diameter, while T = 4 is a solid tumor with extension to adjacent tissue. T = 2 and T = 3 refer to the intermediate cases. N = 0 refers to the existence of no clinical evidence of a lymph node metastases and N = 1, N = 2, N = 3 indicate, in increasing amplitude, the extent of the involvement of existing lymph nodes. Patients with classifications T = 1, N = 0; T = 1, N = 1; T = 2, N = 0; Or T = 2, N = 1, or with distant metastases are excluded of the study. The variable general condition gives a measure of the functional capacity of the patient at the time of diagnosis (1 refers to no disability while 4 denotes a confinement bed; 2 and 3 measure the intermediate levels). The variable quality is a measure of the degree of differentiation of the tumor (degree in which the tumor cell resembles the host cell) from 1 (well differentiated) to 3 (little differentiated). Weibull distribution has a lot of flexibility and that is why it is used routinely in survival analysis. In the parametric analysis for inferences on covariates and adapt precisely, we assume that the location parameter associated with covariates, but the shape parameter is not (The Department of Veterans Affairs Laryngeal Cancer Study Group, 1991). But for some data, the hypothesis of constant shape parameter is inappropriate, in some studies with fatigue materials, it is generally assumed that the shape parameter of the Weibull distribution depends on covariates, as we can see in Wang (2000), Meeker (1993), Pascual (1997), Meeter (1994), Hirose (1993), Chan (1991) and Smith (1991) (Meeker and Escobar, 1998). In this paper we use the model with constant shape parameter and also non-constant shape parameter and then we choose the best model according to AIC criteria which says that the less AIC the better. For data analysis, clinical, pathological and biological characteristics of patients were evaluated in the survival model. A Weibull distribution was proposed for survival time. The data were analyzed using SAS university edition. Quantitative results were expressed as mean \pm SD. The level of significance was set at $P \leq 0.05$.

Results

Patient characteristics are presented in Table 1. A total of 195 men and women with pharyngeal cancer were included in this analysis. The mean age at diagnosis was estimated to be 60.44 ± 11.21 years. The age of patients, ranged from 20 to 90 years. A total of 142 patients (72.8%) died due to pharyngeal cancer during the study. The Kaplan Meier curve for the whole population is shown in Figure 1. As the figure shows, the more time passes the chance of survival decreases. In order to check the appropriateness of the Weibull distribution, the log

[-log (survival probability)] against the log (time) are plotted which is available at Figure 2, and it shows that the Weibull distribution is the appropriate model to be choose. The prognostic variables included in the model were age at diagnosis, institutional involvement, type of treatment, grade, sex, patient status, tumor site, T STAGE

Table 1. Clinical, Pathological, and Biological Characteristics in Patients with Pharyngeal Cancer

Factors	No. Of patients	Percentage (%)
Participating institution		11.3
1	22	24.1
2	47	21
3	41	
4	26	13.3
5	30	15.4
6	29	14.9
Sex		
Male	149	76.4
Female	46	23.6
Treatment		
Radiotherapy	100	51.3
Radiotherapy with chimotherapy	95	8.7
Grade		
Well differentiated	49	25.1
Moderately differentiated	110	56.4
Poorley differentiated	36	18.5
Condition		
No disability	144	73.8
Restricted work	43	22.1
Requires assistance with self care	8	4.1
Site		
Faucial arch	65	33.3
Tonsillar fossa	64	32.8
Posterior pillar	0	0.0
Pharyngeal tongue	66	33.8
Posterior wall	0	0.0
T-Stage		
1	9	4.6
2	26	13.3
3	93	47.7
4	67	34.4
N-Stage		
1	39	20.0
2	28	14.4
3	37	19.0
4	91	46.7
Age		
Under 40 yrs	5	2.6
40 to 60 yrs.	94	48.2
Over 60 yrs.	96	49.2

Table 2. Identified Risk Factors Based on Weibull Distribution with Constant Shape Parameter

Factors	Estimation	SD	P-value
Constant			
b0	5.29	0.425	<.0001*
Participating institution			
1	0.08	0.288	0.775
2	-0.17	0.237	0.474
3	0.28	0.235	0.226
4	0.43	0.266	0.107
5	0.16	0.267	0.54
6**			
Sex			
Male	0.23	0.173	0.18
Female**			
Treatment			
Radiotherapy	-0.08	0.141	0.552
Radiotherapy with chemotherapy**			
Grade			
Well differentiated	-0.29	0.238	0.223
Moderately differentiated	-0.34	0.216	0.111
Poorly differentiated**			
Condition			
No disability	1.42	0.329	<.0001*
Restricted work	0.59	0.347	0.087
Requires assistance with self care**			
Site			
Faucial arch	-0.07	0.186	0.706
Tonsillar fossa	0.01	0.191	0.941
Posterior pillar	0	0	0
Posterior wall	0	0	0
Pharyngeal tongue **			
T-Stage			
1	0.11	0.395	0.775
2	0.37	0.272	0.172
3	0.25	0.17	0.134
4**			
N-Stage			
1	0.19	0.202	0.346
2	0.33	0.224	0.135
3	0.2	0.223	0.371
4**			
Age			
Under 40 yrs.	-0.13	0.439	0.761
40 to 60 yrs.	-0.1	0.144	0.479
Over 60 yrs.**			
Constant			
a0	0.24	0.067	0.0003*

* Significant at the 5% level; ** Stands for a control group.

and N STAGE. According to the results of the Weibull survival model with constant shape parameter, which

Table 3. Identified Risk Factors Based on Weibull Distribution with Non-Constant Shape Parameter

Factors	Estimation	SD	P-value
Constant			
b0	5.25	0.403	<.0001*
Participating institution			
2(b2)	-0.17	0.168	0.301
3(b3)	0.22	0.208	0.281
4(b4)	0.4	0.252	0.114
6**			
Sex			
Male (b6)	0.35	0.176	0.044*
Female**			
Grade			
Well differentiated (b8)	-0.24	0.235	0.308
Moderately differentiated(b9)	-0.13	0.2	0.513
Poorly differentiated**			
Condition			
No disability (b12)	1.11	0.405	0.006*
Restricted work (b13)	0.22	0.401	0.581
Requires assistance with self care**			
T-Stage			
2 (b15)	0.6	0.24	0.012*
3 (b16)	0.3	0.143	0.036*
4**			
N-Stage			
1 (b17)	0.23	0.168	0.161
2 (b18)	0.34	0.182	0.062
3 (b19)	0.22	0.251	0.375
4**			
Age			
40 to 60 yrs. (b21)	0.09	0.145	0.502
Over 60 yrs.**			
Constant			
a0	-0.03	0.274	0.009*
Participating institution			
2 (a2)	0.62	0.226	0.006*
3 (a3)	0.12	0.214	0.549
6**			
Sex			
Male (a6)	-0.19	0.198	0.324
Female**			
Treatment			
Radiotherapy (a7)	-0.08	0.167	0.595
Radiotherapy with chemotherapy**			
Grade			
Well differentiated (a8)	-0.34	0.27	0.202
Moderately differentiated (a9)	-0.02	0.228	0.906
Poorly differentiated **			

Table 3. Continued

Factors	Estimation	SD	P-value
Grade			
Well differentiated (a8)	-0.34	0.27	0.202
Moderately differentiated (a9)	-0.02	0.228	0.906
Poorly differentiated **			
Condition			
No disability (a10)	0.14	0.196	0.464
Requires assistance with self care **			
Site			
Faucial arch (a12)	0.22	0.233	0.339
Tonsillar fossa (a13)	0.2	0.21	0.339
Posterior pillar	0	0	0
Posterior wall	0	0	0
Pharyngeal tongue **			
T-Stage			
2 (a15)	0.33	0.321	0.301
3 (a16)	0.14	0.168	0.401
4**			
N-Stage			
1 (a17)	0.23	0.21	0.266
2 (a18)	0.12	0.242	0.604
4**			
Age			
Under 40 yrs.(a20)	-0.45	0.468	0.332
40 to 60 yrs.(a21)	-0.12	0.171	0.453
Over 60 yrs.**			

* Significant at the 5% level; ** Stands for a control group.

are presented in Table 2, only the patient’s status was significant and the AIC of this model became 2152.4. The results of the Weibull survival model with non-constant shape parameter are presented in Table 3. The letter “a” represents the non-constant shape parameter and in this model, if a characteristic of a patient becomes significant in one of the situation constant or non- constant shape (letter a “Or” b “), it counts as a significant characteristics. Therefore, according to this table; sex, patient status, tumor stage and the institute in which the patient participated were significant. The AIC of this model became 2150.1.

The fitted model according to Weibull distribution would be:

$$f(t | \hat{\alpha}, \hat{\lambda}) = \hat{\lambda} \hat{\alpha}^{\hat{\lambda}} (\hat{\alpha} \hat{\lambda} t)^{\hat{\lambda}-1} \exp(-\hat{\lambda} \hat{\alpha}^{\hat{\lambda}} t); t > 0, \hat{\alpha} > 0, \hat{\lambda} > 0$$

And the estimated parameters are:

$$\hat{\lambda} = \exp [5.25 - 0.17(\text{institute num2}) + 0.22(\text{institute num3}) + 0.4(\text{institute num4}) + 0.35(\text{Male}) - 0.24(\text{Well differentiated}) - 0.13(\text{Moderately differentiated}) + 1.11(\text{No disability}) + 0.22(\text{Restricted work}) + 0.6(\text{Stage 2 of the tumor}) + 0.3(\text{Stage 3 of the tumor}) + 0.23(\text{stage 1 of Nstage}) + 0.34(\text{stage 2 of Nstage}) + 0.22(\text{stage 3 of Nstage}) + 0.09(\text{age of 40 to 50 yrs})]$$

$$\hat{\alpha} = \exp [-0.03 + 0.62(\text{institute num2}) + 0.12(\text{institute num3}) - 0.19(\text{male}) - 0.08(\text{radiotherapy}) - 0.24(\text{well$$

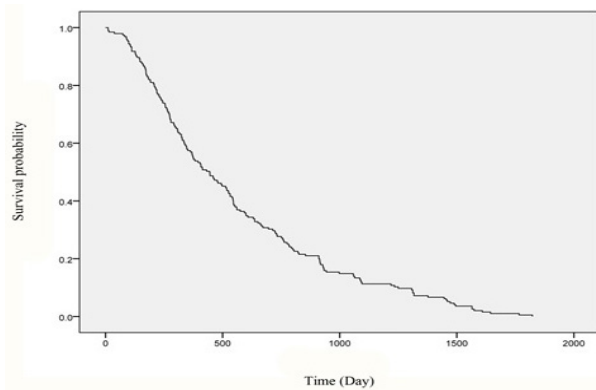


Figure 1. The Kaplan Meier Curve of the Total Population; It shows that the more time passes the more chance of survival decreases.

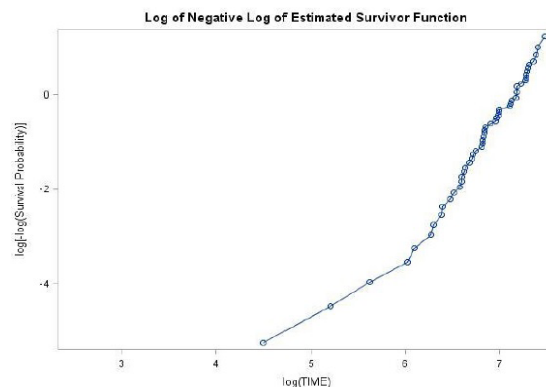


Figure 2. The Log [-Log (Survival Probability)] Against the Log(Time). This figure satisfies the Weibull distribution because all the points allies on the line.

differentiated)-0.02(moderately differentiated)+0.14 (No disability)+0.22(faucial arch)+0.2(tonsillar fossa)+0.33(Stage 2 of the tumor)+0.14(Stage 3 of the tumor)+0.23(stage 1 of Nstage)+0.12(stage 2 of Nstage)-0.45(age of under 40 yrs)-0.12(age of 40 to 60 yrs)].

Discussion

Cancer of the ear and pharynx, grouped together, is the sixth most common cancer in the world (Vigneswaran and Williams, 2014). One way that we could be prepared to investigate cancer patients is to monitor their survival patterns (Warnakulasuriya, 2009). There are two models that are generally used in survival studies, Cox regression as a semi parametric model and parametric models like Weibull distribution. There have been many clues that show parametric models are more flexible, a reason that shows this flexibility, is that scale and shape parameters may depend on covariates, so these models have fewer parameters than the Cox regression , so they lead to models that are easier to interpret (Abu Bakar et al., 2008). Baghestani (2010) showed by a simulation study that on a non-constant state of shape, the estimates in survival studies were accurate and impartial. In recent years or even more, there have been many studies on parametric models and many of them have been talked about non-constant state of parameters. In some statistical literature, it has

been shown that the non-constant shape parameter leads to a model with a lower AIC (Baghestani and Hajizadeh, 2010; Mazucheli et al., 2008). Luo (2013) used the Weibull model with non-constant shape parameter to demonstrate the reliability target with the specified reliability lifetime. The study by Seo (2009) showed that the use of ALTSP with a non-constant shape parameter in comparison to the usual model, a larger sample size was required and a larger number of test items in the sample must be allocated to low stress to satisfy producer and consumer requirements. The simulation study in LV's (2015) article showed that the model with random effects and non-constant form of parameters obtained better results than the other models in percentile estimation and it proposed to engineers to take into account random effects and non-constant shape parameters in ALTs to obtain a better estimate of model percentiles. In this study, we sought to find factors that could affect the survival of patients with pharyngeal cancer, including clinical, pathological and biological characteristics. In our study, a Weibull distribution with a non-constant shape parameter was proposed for survival time and according to our results: sex, patient status, tumor stage and institute to which the patient had participated in were significant. Due to the increasing incidence of this type of cancer, much research has been done to recognize risk factors. It has been shown in Llewellyn's (2004) study that tobacco and alcohol were the risk factors for young adults. In Elwood's (1984) study alcohol, smoking, low socio-economic status, unmarried status and poor dental care were identified. Burbone (1996) made a follow-up study and he concluded that smoking, occupancy and beta-carotene level are the common factors for primary and secondary metachronous primary tumors. A study by Jones (1998) had also shown that the age of patients had an effect on survival and other parameters in squamous cell carcinoma of this type of cancer. Yong Choi (1991) did a research to check the effect of smoking and alcohol on the etiology of this cancer.

We used a different method to analyze survival time and identify risk factors in this type of cancer so that we reached a model with lower AIC which indicated the risk factors among these patients in this sample more precisely. There are many parametric models that have the potential on their shape parameter to study on. So the research in this case is extended.

Acknowledgments

It's notable that the paper is extracted from MSc thesis of Ms. Haleh Aghamolaey in Biostatistics.

References

Abu Bakar MR, Salah KA, Ibrahim NA, et al (2008) . Cure fraction, modelling and estimating in a population-based cancer survival analysis. *Malaysian J Math Sci*, **2**, 113-34.
Baghestani AR, Hajizadeh E (2010). Parametric analysis for interval censored survival data with non-constant shape parameter. *JSLAU*.
Burbone F, Franceschi S, TulaminiR, et al (1996). A follow-up study of determinants of second tumor and metastasis among subjects with cancer of the oral cavity, pharynx, and larynx.

J Clin Epidemiol, **49**, 367-72.
Chan CK (1991). Temperature-dependent standard deviation of log (failure time) distributions. *IEEE Trans Reliabil*, **40**, 157-60.
Efron B (1977). The efficiency of Cox's likelihood function for censored data. *J Am Stat Assoc*, **72**, 557-65.
Elwood JM, Pearson JCG, Skipen DH, Jackson SM (1984). Alcohol, smoking, social and occupational factors in the aetiology of cancer of the oral cavity, pharynx and larynx. *Int J Cancer*, **34**, 603-12.
Hirose H (1993). Estimation of threshold stress in accelerated life-testing. *IEEE Trans Reliabil*, **42**, 650-7.
Jones AS, Beasley N, Houghton D, Husband DJ (1998). The effect of age on survival and other parameters in squamous cell carcinoma of the oral cavity, pharynx and larynx. *Clin Otolaryngol*, **23**, 51-6.
Klein JP, Moeschberger ML (1977). Survival analysis techniques for censored and truncated data. Springer-Verlag New York.
Lambert PC, Thompson JR, Weston CL, et al (2007). Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics*, **8**, 576-94.
Llewellyn CD, Linklater K, Bell J, Johnson NW, Warnakulasuriya S (2004). An analysis of risk factors for oral cancer in young people: a case-control study. *Oral Oncol*, **40**, 304-13.
Luo W, Zhang C, Chen X, Tan Y (2013). Accelerated reliability demonstration for Weibull distribution with a nonconstant shape parameter. In 2013 international conference on quality, reliability, risk, maintenance, and safety engineering (QR2MSE).
Lv S, Niu Zh, Qu L, He S , He Zh (2015). Reliability modeling of accelerated life tests with both random effects and nonconstant shape parameters. *Qual Eng*, **27**, 329-40.
Mazucheli J, Achcar JA, Coelho-Barros FA, Louzada-neto F (2008). A general survival regression model with nonconstant shape parameter. *Rev Bras Biom*, **26**, 27-40.
Meeker WQ, Escobar LA (1993). A review of recent research and current issues in accelerated testing. *Int Stat Rev*, **61**, 147-68.
Meeker WQ, Escobar LA (1998). Statistical methods for reliability data. Wiley-Interscience , pp 712.
Meeter CA, Meeker WQ (1994). Optimum accelerated life tests with a nonconstant scale parameter. *Technometrics*, **36**, 71-83.
Oakes D (1977). The asymptotic information in censored survival data. *Biometrika*, **64**, 441-80.
Pascual FG, Meeker WQ (1997). Regression analysis of fatigue data with runouts based on a model with nonconstant standard deviation and a fatigue limit parameter. *J Test Eval*, **25**, 292-301.
Radiation therapy oncology group in the United States (1980). A clinical trial in the Trt. of carcinoma of the oropharynx (PHARYNX.DAT). The statistical analysis of failure time data, by JD Kalbfleisch and RL Prentice , Published by John Wiley and Sons.
Seo JH, Jung M, Kim CM (2009). Design of accelerated life test sampling plans with a nonconstant shape parameter. *Eur J Oper Res*, **197**, 659-66.
Shedd DP, von Essen CF, Connelly RR, Eisenberg H (1968). Cancer of the pharynx in Connecticut , 1935-1959. *Cancer*, **21**, 706-17.
Smith RL (1991). Weibull regression models for reliability data. *Reliabil. Eng Syst Saf*, **34**, 55-76.
Smith RR, Frazel EL, Caulk R, Holinger PH, Russell WO (1963). The american joint committee's proposed method of stage classification and end-result reporting applied to 1,320 pharynx cancers. *Cancer*, **16**, 1505-20.
The department of veterans affairs laryngeal cancer study group (1991). Induction chemotherapy plus radiation compared

- with surgery plus radiation in patients with advanced laryngeal cancer. *N Engl J Med*, **324**, 1685-90.
- Vigneswaran N, Williams MD (2014); Epidemiologic trends in head and neck cancer and aids in diagnosis. *Oral Maxillofac Surg Clin North Am*, **26**, 123-41.
- Wang W, Kececioglu DB (2000). Fitting the Weibull log-linear model to accelerated life-test data. *IEEE Trans Reliabil*, **49**, 217-23.
- Warnakulasuriya S (2009). Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol*, **45**, 309-16.
- Yong Chois S, Kahyot H (1991). Effect of cigarette smoking and alcohol consumption in the aetiology of cancer of the oral cavity, pharynx and larynx. *Int J Epidemiol*, **20**.