DIGITALCOMMONS @WAYNESTATE

Journal of Modern Applied Statistical Methods

Volume 13 | Issue 1 Article 9

5-1-2014

Predicting Survival Time of Localized Melanoma Patients Using Discrete Survival Time Method

Taysseer Sharaf
University of South Florida, tsharaf@umich.edu

Chris P. Tsokos
University of South Florida, chris.tsokos@gmail.com

Recommended Citation
Sharaf, Taysseer and Tsokos, Chris P. (2014) "Predicting Survival Time of Localized Melanoma Patients Using Discrete Survival Time Method," *Journal of Modern Applied Statistical Methods*: Vol. 13 : Iss. 1 , Article 9.
DOI: 10.22237/jmasm/1398917280
Available at: http://digitalcommons.wayne.edu/jmasm/vol13/iss1/9

Journal of Modern Applied Statistical Methods
May 2014, Vol. 13, No. 1, 140-156.

ISSN 1538 − 9472

Predicting Survival Time of Localized Melanoma Patients Using Discrete Survival Time Method

Taysseer Sharaf
University of South Florida
Tampa, FL

Chris P. Tsokos
University of South Florida
Tampa, FL

Melanoma is the most fatal type of skin cancer. It is ranked first in death of skin cancer diseases. This study establishes a statistical model that can predict the survival time of localized melanoma patients, as a function of age at diagnosis, tumor thickness, and extension of the tumor (tumor invasion). The discrete time survival method was used to build the statistical model. The patients involved in the current study were observed from the SEER database. Patients were divided into nine groups according to age at diagnosis. Variation in survival time was found to be significant among some of the age groups.

Keywords: melanoma, survival time, discrete survival time, skin cancer, SEER, localized melanoma

Introduction

Melanoma is a malignant tumor associated with skin cancer. If melanoma is detected at a late stage, it can spread to other parts of the body and that's what makes it a lethal form of cancer. More general information about melanoma can be found in (www.melanoma.org), (Markovic, et al., 2007) and (Mackle, et al., 2009). Over the last decades, the incidence of melanoma has been rapidly increasing in the United States. It appears more in white populations than other races. According to clinical studies, risk factors of melanoma are but not limited to, ultraviolet light exposure, moles, light hair, freckling and family history of melanoma. Some of the statistical analyses done on the risk factors are shown in (Gandini, et al., 2005c; Naldi, et al., 2000; Cho, et al., 2005).

Mr. Sharaf is a Graduate Teaching Assistant in the Department of Mathematics and Statistics. Email him at mailto:tsharaf@mail.usf.edu. Dr. Tsokos is Distinguished University Professor in the Department of Mathematics and Statistics. Email him at chris.tsokos@gmail.com.

Consider the survival time for melanoma patients. The primary objective is the time between when the patient is diagnosed with melanoma and when death occurs. The study includes the effect of three risk factors that drives the survival time for a patient. Those risk factors are Age at diagnosis, tumor thickness and extension of the tumor (invasion of tumor through the body). Other factors include gender and sequence number (a number that indicates how many tumors the patient had prior to being diagnosed with melanoma). The main concern will be in estimating the survival time of melanoma patients diagnosed at stage one (localized Melanoma). More information regarding staging is discussed in the Methodology; for updates on the staging of melanoma visit www.cancer.gov. Soong, et al. (2010) developed an electronic prediction tool based on the American Joint Committee on Cancer (AJCC) melanoma staging database, to predict survival outcome of localized melanoma. Other predictive models of survival for localized melanoma have been developed in the United States and other countries (Clark, et al., 1989; MacKie, et al., 1995; Barnhill, et al., 1996; Schuchter, et al., 1996; Sahin, et al., 1997; Soong, et al., 2003). Soong, et al. (2010) used the Cox survival function model, which considers the survival time as a continuous random variable, where most survival times are recorded in discrete form as a number of months or years. They used same three risk factors in their analysis beside the primary melanoma site and primary tumor ulceration. As shown in Table 3, there exist 10 primary melanoma sites, and in order to reduce the variation in the model (biological variation between humans is a lurking variable), only one site of the ten was studied. Allison (1982) mentioned that in continuous survival time, maximum likelihood method ignores the discrete character of the data.

Xie, Mchugo, Drake, and Sengupta (2003), summarized the advantages of using discrete-time survival analysis. These were initially suggested by Singer & Willett (1993) as primarily useful for many longitudinal studies in clinical settings where data are often collected at discrete time periods. Secondly, the analysis facilitates the examination of the shape of the hazard function. Third, the analysis is simple and convenient to use, because it is a modification of the logistic regression model. Lastly and most important, time-varying covariates can easily be included in the model. After Cox presented the discrete time survival model, two basic versions of logistic models were introduced: the ordinal version and the dichotomous version. The dichotomous version (Allison, 1982; Singer & Willett, 1993; Xie, et al., 2003) represents each survival time as a set of indicators of whether or not an individual failed at each time point, until a person either experiences the event or is censored.

Methodology

Allison (1982) and Singer and Willett (1993) proposed a discrete survival time method. The method starts by dividing the continuous time into an infinite sequence of contiguous time $(0,t_1),(t_1,t_2),\ldots,(t_{k-1},t_k)\ldots$, and so on. Let k represent the number of time intervals. In this case, time is recorded in months, where time is divided into 20 intervals each consists of 12 months: (1,12),(12,24),…,(228,240). If a patient's survival time is 7 months, then this patient's event is classified as happening during the 1st time interval; if another patient's survival time is 50 months, then this is classified as happening during the 5th time interval.

To estimate the survival function, start with the discrete-time hazard model (Allison, 1982)

$$h_{ik} = \frac{1}{1+EXP\left\{-\left(\alpha_1 T_{1ik}+\alpha_2 T_{2ik}+\ldots+\alpha_J T_{Jik}\right)-\left(\beta_1 X_{1ik}+\beta_2 X_{2ik}+\ldots+\beta_p X_{pik}\right)\right\}} \quad (1)$$

where $\left[T_{1ik},T_{2ik},\ldots,\alpha_J T_{Jik}\right]$ are a sequence of dummy variables, with values $\left[t_{1ik},t_{2ik},\ldots,t_{Jik}\right]$ indexing time periods, where J refers to the last time period observed for any individual in the sample. If individual i was observed (experienced the event or censored) in the fourth period, then $J = 4$, and the time period's dummy variables are defined identically for each individual; $t_{1ik}=1$ when $j = 1$ and 0 when j takes any other value.

The coefficients $\left(\alpha_1\alpha_2\ldots\alpha_J\right)$ act as the intercept parameters for the baseline hazard in each time period, and the coefficients $\left(\beta_1\beta_2\ldots\beta_p\right)$ describe the effect of the predictors on the baseline hazard in the logit scale. Singer and Willett (1993) discussed briefly the procedures to construct the likelihood function (in terms of the discrete hazard function) used to estimates the latter intercepts and slope parameters.

The likelihood function presented by Singer and Willett (1993) is given by

$$L=\prod_{i=1}^{n}\prod_{k=1}^{j_i} h_{ik}^{y_{ik}}\left(1-h_{ik}\right)^{(1-y_{ik})} \quad (2)$$

where y_{ik} is a sequence of dummy variables that records the event history for patient i, whose values are defined as:

$$y_{ik} = \begin{cases} 1 \text{ if the } i^{\text{th}} \text{ patient experienced the event in period } k \\ 0 \text{ if the } i^{\text{th}} \text{ patient did not experience the event in period } k \end{cases}$$

The likelihood function is identical to the likelihood function to a sequence of $N = (k_1 + k_2 + \ldots + k_n)$ independent Bernoulli trials with parameters h_{ik}.

Using results by Allison (1982), the y_{ik} values can be considered as the outcome variable in a logistic regression analysis, which provides a simple model to obtain the maximum likelihood estimate rather than finding the solution by maximizing equation (2).

For discrete event history data, each record consists of the information for one patient like survival time, Age and whether or not the patient time is censored or not. In order to apply the logistic model discussed previously, the data need to be converted into new person-period data, in which each patient will have multiple records, one per time period of observation. As shown by Singer and Willett (1993), the new person-period data will contain the information about the k^{th} time period as follows

- ***The time indicators.*** The set of dummy variables $[T_{1ik}, T_{2ik}, \ldots, \alpha_J T_{Jik}]$.
- ***The predictors.*** Covariates under study, where the ability exists to use the time-varying covariates that have values differs from time period to time period.
- ***The event indicator*** *(response variable in the logistic model).* This variable records whether the event of interest occurred in period j or not. The variable takes value 1 if the event occurred, takes 0 if did not.

In this study, the survival time of melanoma patients diagnosed in the period of 1988 to 2008 was considered. The survival time was recorded up to the nearest month. The time period of the study was divided into 20 intervals, one year each. Besides the covariates age at diagnosis, tumor size and extension of the tumor, there were 20 dummy variables representing the 20 time periods as shown in Tables 1 and 2. In Table 1, there is a record of 3 patients as extracted from Surveillance Epidemiology and End Results database (SEER) database. In Table

2, there is a representation of the conversion of the data to fit the new person-period data to be used for the logistic model.

Table 1. A record of 3 patients from SEER database.

Patient ID	Survival Time (mos)	Age Diagnosis	Tumor Size	Extension of Tumor
7013574	49	84	120	10
8862253	3	66	230	30
8869492	86	61	134	30

Table 1 shows that the first patient survival time is 49 months. The first patient lived for four years and died in the first month of the fifth year, which means that the event took place during the fifth time period. In the new data setting the first patient will have five records, one record corresponding to every time period (from the first to the fifth). The event indicator variable will take 0 for the first four records and 1 in the fifth record where the event took place. Table 2 shows this conversion for the three patients in table one.

Table 2. A sample of three patients from the new person-period data.

Indc.	ID	ST	A	TS	D_1	D_2	D_3	D_4	D_5	D_6	D_7	D_8	…	D_{20}
0	1	49	84	120	1	0	0	0	0	0	0	0	…	0
0	1	49	84	120	0	1	0	0	0	0	0	0	…	0
0	1	49	84	120	0	0	1	0	0	0	0	0	…	0
0	1	49	84	120	0	0	0	1	0	0	0	0	…	0
1	1	49	84	120	0	0	0	0	1	0	0	0	…	0
1	2	3	66	230	1	0	0	0	0	0	0	0	…	0
0	3	86	61	134	1	0	0	0	0	0	0	0	…	0
0	3	86	61	134	0	1	0	0	0	0	0	0	…	0
0	3	86	61	134	0	0	1	0	0	0	0	0	…	0
0	3	86	61	134	0	0	0	1	0	0	0	0	…	0
0	3	86	61	134	0	0	0	0	1	0	0	0	…	0
0	3	86	61	134	0	0	0	0	0	1	0	0	…	0
0	3	86	61	134	0	0	0	0	0	0	1	0	…	0
1	3	86	61	134	0	0	0	0	0	0	0	1	…	0

The variables $D_1, D_2, \ldots, D_{20}$ represent the 20 dummy variables for the time intervals. Each row shows patient i information during each time interval until the event occurs or he/she is censored. The first column in Table 2 (Indc.) represents the indicator variable, which takes 0 if the event did not take place during the

current time period interval, or 1 if the event occurs during the time period interval. The patient's ID was changed to ID (1, 2 and 3) to optimize the table. The setting of the data shown in Table 2 can allow us to use time varying covariates easily, but because the data used does not support this information the covariates were repeated. Only the tumor size of the patient is known at the time of diagnosis; no follow up information was supported. Age at diagnosis can be changed, but no big difference will appear (as will be shown); ages were grouped into 9 intervals, each interval covering 10 years.

The data set used was collected from the Surveillance Epidemiology and End Results database (SEER) 1973-2008. 208,143 patients were diagnosed in the United States from 1973 through 2008. Taking into account the patients that were confirmed dead because of melanoma cancer, and removing all the missing records in the covariates shown in Table 4, results in studying the patients diagnosed from 1988 through 2008. Melanoma cancer is classified into 4 stages as shown in Table 3, and there exist 10 sites of the skin where the cancer appears. In order to reduce the variation (biological difference between humans) and to get less prediction error in the statistical model, only 'skin of trunk' patients in stage 1 (localized melanoma) are considered in this study. The sample size used in this study is 1,240. Table 3 shows the description of the coding used for the primary site and the staging of the cancer.

The risk factors (affecting the survival time of patients) that were involved in the study are age of the patient, the tumor size and the tumor extension (how far the tumor spread). Around 99 percent of the 1,240 patients were white (due to the fact that melanoma is rare in people with dark skin), which is why race was not considered in the modeling aspects.

Age of patients at Diagnosis was classified into 9 groups, 10 years each, starting from 11 to 20 years in the first group through 91 to 100 years old. Tumor thickness (instead of size as known in other cancer types, the thickness is measured, but size will be referred to for the remainder of this article) was classified into 3 groups, the first group from 1mm to 50mm, the second group from 51mm to 300mm and the last group from 301mm to 992mm. In the current sample the extension of the tumor contains 4 levels which are as stated in SEER EOD-88 3rd edition; (10) for papillary dermis (the middle layer of skin) invaded, (20) for papillary-reticular dermal interface invaded, (30) for reticular dermis invaded and finally (40) localized. A summary of the number of patients lying in the groups stated previously is shown in Table 4.

Table 3. SEER coding for the stage and primary site for Melanoma.

Variable	Code	Description
SEER Historic stage	0	In situ : A tumor which has not penetrated the basement membrane nor extended beyond the epithelial issue
	1	Localized : An invasive neoplasm confined entirely to the organ of origin
	2	Regional: A neoplasm that has gone beyond the bounds of the organ of origin or into regional lymph nodes.
	4	Distant : A neoplasm that has spread to parts of the body distant from the primary tumor
Primary Site	C440	Skin of lip
	C441	Eyelid
	C442	External Ear
	C443	Skin of other and unspecified parts of face
	C444	Skin of Scalp and neck
	C445	Skin of Trunk
	C446	Skin of upper limb and shoulder
	C447	Skin of lower limp and hip
	C448	Overlapping lesion of skin
	C449	Skin NOS

Table 4 illustrates that around 46.5% of the patients are diagnosed at the third level of tumor extension where the tumor invaded into the reticular dermis, indicating that there is a delay from patients until they figured out that they needed medical attention. It must be stressed that during the current study no treatment effects were added to the statistical model, so study results are considered as if patients did not get any treatment. The different treatment effects and histology effects will be studied in further publications.

Descriptive statistics of the survival time of melanoma patients recorded in months from time of diagnosis till death for each age group are recorded in Table 5. Because the survival time distribution is skewed, it is important to estimate the median survival time for the melanoma patients. The median will be more informative than the mean in this case. The large variance of survival time inside each group can be seen. This assures the presence of independent variables affecting the survival time.

Table 4. Distribution of the 1240 patients on the various groups.

Factor		Extension of the Tumor					
Age	**Tumor Size**	**10**	**20**	**30**	**40**	**Total**	**%**
	1						
11-20	2	1	4			5	0.4838710
	3				1	1	
	1	3	10	2	1	16	
21-30	2	3	11	12	4	30	4.2741935
	3			6	1	7	
	1	8	3	11	3	25	
31-40	2	11	52	50	8	121	13.3870970
	3		5	15		20	
	1	14	9	13	2	38	
41-50	2	10	87	73	12	182	20.3225810
	3		8	20	4	32	
	1	17	9	19	7	52	
51-60	2	17	61	73	14	165	20.3225810
	3		3	27	5	35	
	1	14	3	11	5	33	
61-70	2	14	59	87	18	178	20.4032260
	3		4	33	5	42	
	1	18	7	12	4	41	
71-80	2	8	36	60	15	119	15.2419350
	3		7	19	3	29	
	1	6	2	11	3	22	
81-90	2	3	9	14	5	31	5.0806452
	3		3	5	2	10	
	1	1		1		2	
91-100	2		1	2		3	0.4838710
	3			1		1	
%		11.94000	31.69000	46.53000	9.83871	1240	100%

Table 5. Descriptive statistics of survival time.

	Survival Time				
Age	**Count**	**Mean**	**Median**	**St. deviation**	**Skewness**
11-20	6	74.50	57.00	42.43	1.39
21-30	53	61.40	46.00	44.10	1.11
31-40	166	66.95	52.50	48.48	1.21
41-50	252	67.83	55.50	45.53	1.15
51-60	252	56.90	43.00	41.17	1.32
61-70	253	53.96	41.00	42.11	1.58
71-80	189	50.38	45.00	32.86	0.89
81-90	63	38.44	34.00	28.05	1.26
91-100	6	32.17	26.00	31.92	1.91

Results

Three out of the four models proposed models are discussed. The first model is the baseline model, estimating the survival function using the time periods. The second model introduces age at diagnosis as the first covariate with the time periods. The third model uses age at diagnosis and tumor size as covariates with the time periods. The fourth model introduces all three covariates with the time periods.

Model 1

The baseline model is the starting point of the proposed modeling procedure. The simplest hazard model from equation (1), considering only the 20 dummy variables that represent the time effect, is considered. The baseline hazard model for this case is represented as:

$$logit(h_i) = (\alpha_1 t_1 + \alpha_2 t_2 + \ldots + \alpha_{20} t_{20}) \tag{3}$$

Equation (3) represents the log transform of equation (1), where

$$logit(h_{ik}) = \log\left(\frac{h_{ik}}{1-h_{ik}}\right) \tag{4}$$

This model will answer the basic question 'what is the probability of obtaining the event (melanoma patient dies due to the cancer) in each time period?' In other words what is the probability that a melanoma patient will survive for one, or two years, etc.

The parameters in equation (3) can be converted by exponentiation the right hand side. For example if it is desired to know the estimate of the probability of event occurrence in the fifth interval will be equal to

$$\hat{h}_5 = \frac{1}{(1+e^{-\alpha_5})} \tag{5}$$

The estimates of the baseline hazard are presented in Table 6. After estimating the baseline hazard one can calculate the corresponding survival function using

$$\hat{S}_k = \prod_{j=1}^{k} (1 - \hat{h}_j) \tag{6}$$

Table 6. Estimates of the baseline hazard parameters.

Param.	$\hat{\alpha}$	**Param.**	$\hat{\alpha}$
D1	-2.687	D11	-0.743
D2	-1.642	D12	-1.326
D3	-1.372	D13	-1.017
D4	-1.242	D14	-0.758
D5	-1.275	D15	-1.273
D6	-1.125	D16	-0.405
D7	-1.233	D17	-0.693
D8	-1.031	D18	0.000
D9	-1.349	D19	1.386
D10	-1.185	D20	21.203

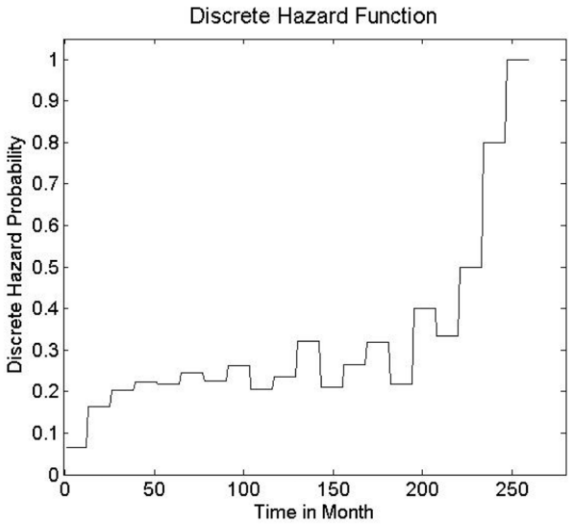

Figure 1. Scatter plot for the estimated baseline Hazard function.

Thus to illustrate the output, the hazard probability of first time period is calculated and by substituting the values in equation (3) $t_2, t_3, \ldots, t_{20} = 0$ while

$t_1 = 1, \hat{h}_1 = \frac{1}{(1+e^{2.687})} = 0.06$, and to get the hazard probability of the second interval $\hat{h}_2 = \frac{1}{(1+e^{1.642})} = 0.16$, and so on. The coefficient of the last time period had a high negative value, but was actually insignificant in the modeling process. This was due to the occurrence of only one event, which had held out till the final period in the sample. Also, removing the last interval from the modeling process did not induce any significant change to the −2loglikelihood, which was used to pick the best model from the four models that were tested.

Once calculated for all 20 time periods the baseline survival function can be calculated using equation (6). Graphical representation of the estimated discrete hazard function is shown in Figure 1. And a comparison of the estimated base line survival function by the model and the sample survival function is shown in Figure 2.

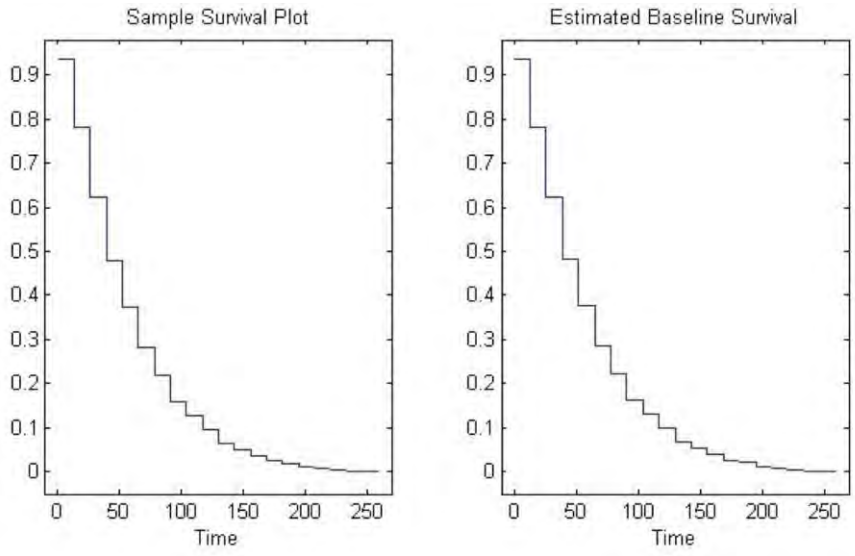

Figure 2. Sample Survival Function and Estimated Baseline Survival.

In Figure 2 it is shown that the estimated base line model fits the data well. This is also supported by the residual analysis from the logistic model used to estimate the baseline hazard function. The model residuals came to be uncorrelated and with constant variance. Because the distribution of the survival

time is skewed, the median survival time is of great interest, as shown in the first graph in Figure 2: the estimated median survival time when $S(t) = 0.5$, which is equal to $\hat{t}_{0.5} = 48$ months. It is customary to see the discrete survival function as a step down function in graphs, but for technical purposes and comparison issues, connected lines are used in these Figures rather than a step down function.

Model 2

The second model is to see the effect of the covariates on the survival time. The first covariate will be age at diagnosis. As discussed in the previous section the age of patient at diagnosis is grouped into 9 groups, 10 year interval each. The nine groups will be represented by 8 dummy variables with the first age group as the base. The parameter estimates are represented in Table 7.

Table 7. Parameter estimates for discrete hazard function with Age as covariate

Param.	α	Param.	α	Param.	β
D1	-3.13	D11	-1.02	age_1	0.28
D2	-2.04	D12	-1.60	age_2	0.11
D3	-1.75	D13	-1.28	age_3	0.09
D4	-1.60	D14	-1.00	age_4	0.41
D5	-1.62	D15	-1.51	age_5	0.39
D6	-1.46	D16	-0.63	age_6	0.73
D7	-1.55	D17	-0.90	age_7	1.00
D8	-1.33	D18	-0.22	age_8	3.11
D9	-1.64	D19	1.18		
D10	-1.48	D20	21.11		

The estimates for the alpha parameters correspond to the time periods. The first beta estimate 0.28 corresponds to the second age group (Age 21-30); recall that first age group is at base level. For example, the estimated discrete hazard probability for the first age group (Age 11-20) of the first time period is $\hat{h}_1 = \frac{1}{(1+e^{-(-3.13)})} = 0.041887,$ the estimate of the survival probability for the same age group in the first time period by equation (4) $\hat{S}_1 = (1-\hat{h}_1) = 0.958.$ Similar calculations were followed to get the survival for the 20 time periods for each Age

group. Figure 3 shows the graph illustration for those survival functions, for each age group.

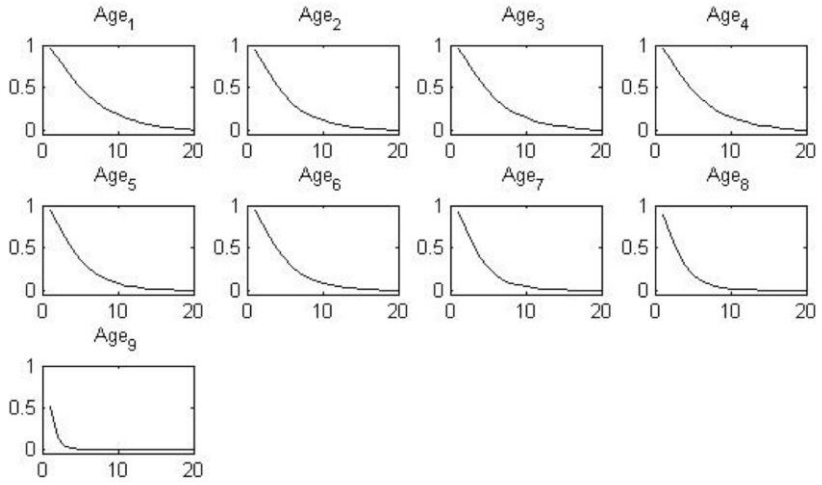

Figure 3. Estimated Survival function plot from Model 2, for different Age groups.

Looking at the survival plots in Figure 3, the duration time of the melanoma cancer is same for patients in the age group 5 and 6. The duration time for the second age group is lower than that for Age group 3 and 4, which is closer to the duration time of the first age group. For the last age group (Ages 91-100) the estimated median survival time corresponds to the first time period, which means it is between 1 and 12 months.

Model 3

Adding more covariates made the model more significant. Table 8 shows the different models that were applied to the data along with the corresponding −2loglikelihood and Cox & Snell R-square. The model that best fits the data is the last one with the three covariates: Age at diagnosis, Tumor Size and Extension of the tumor.

This model with 0.383 Cox and Snell R-square is considered to be significant for the analysis of binary data. This model gave more informative estimates about the behavior of the survival time of melanoma cancer patients, across the 9 age groups. A plot showing the estimated survival time using the last

model based on patients diagnosed with tumor size of group 3 and extension of tumor of group 4 (refer to Table 4) is given in Figure 4.

Table 8. −2loglikelihood for various models

Model	−2loglikehood	Cox & Snell R-square
Base	6088.32	0.369
Base + Age	6016.71	0.376
Base + Age + Tumor Size	5981.91	0.379
Base + Age + Tumor Size + Extension of Tumor	5939.26	0.383

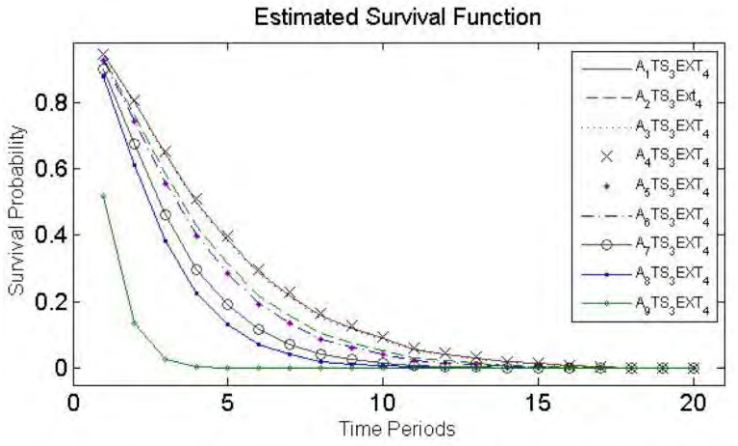

Figure 4. Estimated Survival Function for the different age groups using Model 4.

Figure 4 represents the estimated survival probability for the 9 age groups: A1 for first age group, A2 for second age group, … , A9 for the ninth group. According to the model (Model 4), the following results are found for patients diagnosed with tumor size between (301mm to 992mm) and at the fourth level of tumor extension (No treatment was involved in this model):

- A patient diagnosed at ages 11-20 and 31-50 have the same estimate of median survival time less than 5 years.
- A patient diagnosed at ages 21-30, have an estimate of median survival time less than 4 years.

- Age group 5 (51-60) and 6 (61-70) have the same estimate of median survival time between 3 to 4 years.
- Figure 4 shows that the maximum estimated survival time for all age groups is around 15 years from the time of diagnosis.

Conclusion

A statistical model was developed to predict the survival time of localized melanoma patients using the discrete survival time method. The discrete survival time method gives better results when applied on follow-up data sets. If the information about the progress a patient's tumor thickness and the time of treatment patients took is available, the results become more accurate and show less prediction error.

Four different statistical models were developed with a recommended model (fourth one) to be the best model for predicting the survival time of a given localized melanoma patient. This model is the one that takes into consideration the patient's age at diagnosis, tumor thickness and extension of the tumor.

In comparison with research by Soong, et al. (2010), the primary melanoma site was not considered as one of covariates in the model. Patients were divided into 10 groups according to the primary site and model each group separately. Results from this study show less error compared with Soong, et al. (2010) because the variation due to the difference in the primary site was removed.

References

Allison, P. D. (1982). Discrete-Time Methods for the Analysis of Event Histories. *Sociological Methodology 13*: 61-98.

Barnhill, R. L., Fine, J. A., Roush, G. C., & Berwick, M. (1996). Predicting five-year outcome for patients with cutaneous melanoma in population-based study. *Cancer 78*(3): 427-432.

Bataille, V. & de Vries, E. (2008). Melanoma – Part 1: epidemiology, risk factors, and prevention. *BMJ 337*: 1287-1291.

Cho, E., Rosner, B. A., & Colditz, G. A. (2005). Risk Factors of Melanoma by Body Site, *Cancer Epidemiology, Biomarkers & Prevention 14*: 1241-44.

Clark, W. H. Jr., Elder, D. E., Guerry, D. IV, Braitman, L. E., Trock, B. J., Schultz, D., Synnestvedt, M., & Halpern, A. C. (1989). Model predicting survival

in stage I melanoma based upon tumor progression. *J Natl Cancer Inst. 81*(24): 1893-1904.

Gandini, S., Sera, F., Cattaruzza, M. S., Pasquini, P., Abeni, D., Boyle, P., & Melchi, C. F. (2005a). Meta-analysis of risk factors for cutaneous melanoma: I. Common and atypical naevi, *European Journal of Cancer, 41*(1): 28-44.

Gandini, S., Sera, F., Cattaruzza, M. S., Pasquini, P., Picconi, O., Boyle, P., & Melchi, C. F. (2005b). Meta-analysis of risk factors for cutaneous melanoma: II. Sun exposure, *European Journal of Cancer, 41*(1): 45-60.

Gandini, S., Sera, F., Cattaruzza, M. S., Pasquini, P., Zanetti, R., Masini, C., Boyle, P., & Melchi, C. F. (2005c). Meta-analysis of risk factors for cutaneous melanoma: III. Family history, actinic damage and phenotypic factors, *European Journal of Cancer, 41*(14): 2040-59.

MacKie, R. M., Aitchison, T., Sirel, J. M., McLaren, K., & Watt, D. C. (1995). Prognostic models for subgroups of melanoma patients from the Scottish Melanoma Group database 1979-86, and their subsequent validation. *Br J Cancer 71*(1): 173-176.

Mackle, R. M., Hauschild A., & Eggermont, A. M. M. (2009). Epidemiology of Invasive Cutaneous melanoma, *Annals of Oncology 20*(Suppl 6): vi1-vi7.

Markovic, S. N., et al., (2007). Malignant melanoma in the 21st century, Part 1: epidemiology, risk factors, screening, prevention, and diagnosis. *Mayo Clinic Proceedings, 82*(3): 364-80.

Melanoma Research Foundation. (2011). *Melanoma Statistics & Facts, 2011*, Retrieved October 23, 2012 from http://www.melanoma.org/learn-more/melanoma-101/melanoma-statistics-facts

Naldi, L., Lorenzo, I. G., Parazzini, F., Gallus, S., & La Vecchia, C. (2000). Pigmentary traits, modalities of sun reaction, history of sunburns, and melanocytic nevi as risk factors for cutaneous malignant melanoma in the Italian population, *Cancer, 88*(12): 2703-2710

National Cancer Institute. *Staging*, (2011a, January 11). Retrieved October 23, 2012 from http://www.cancer.gov/cancertopics/wyntk/skin/page9

National Cancer Institute. *What do you need to know About Melanoma and other Skin Cancers*, (2011b, January 11). Retrieved October 23, 2012 from http://www.cancer.gov/cancertopics/wyntk/skin/page4

National Cancer Institute. (2012). *Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Research Data (1973-2009)*,

National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2012, based on the November 2011 submission.

Sahin, S., Rao, B., Kopf, A. W., Lee, E., Rigel, D. S., Nossa, R., Rahman, I. J., Wortzel, H., Marghoob, A. A., & Bart, R.S. (1997). Predicting ten-year survival of patients with primary cutaneous melanoma: a corroboration of a prognostic model. *Cancer 80*(8): 1426-1431.

Schuchter, L., Schultz, D. J., Synnestvedt, M., Trock, B. J., Guerry, D., Elder, D. E., Elenitsas, R., Clark, W. H., & Halpern, A. C. (1996). A prognostic model for predicting 10-years survival in patients with primary melanoma: The Pigmented Lesion Group. *Ann Intern Med*. *125*(5): 369-375.

Singer, J. D. & Willett, J. B. (1993). It's about Time: Using Discrete- Time Survival Analysis to Study Duration and the Timing of Events. *Journal of Educational Statistics 18*(2): 155-195.

Soong, S.-J., Zhang, Y., & Desmond, R. (2003). Models for predicting outcome. In: Balch, C. M., Houghton, A. N., Sober, A., & Soong, S.-J., Eds. *Cutaneous melanoma* (4th ed., pp. 77-90). St. Louis: Quality Medical Publishing.

Soong, S.-J., Ding, S., Coit, D., Balch, C. M., Gershenwald, J. E., Thompson, J. F., Gimotty, P., & the AJCC Melanoma Task Force. (2010). Predicting the Survival Outcome of Localized Melanoma: An Electronic Prediction Tool Based on the AJCC Melanoma Database, *Annals of Surgical Oncology 17*(8): 2006-2014.

Svetomir, N. M., Erickson, L. A., Rao, R. D., Weenig, R. H., Pockaj, B. A., Bardia, A., Vachon, C. M., Schild, S. E., McWilliams, R. R., Hand, J. L., Laman, S. D., Kottschade, L. A., Maples, W. J., Pittelkow, M. R., Pulido, J. S., Cameron, J. D., Creagan, E. T., & the Melanoma Study Group of the Mayo Clinic Cancer Center. (2007). Malignant Melanoma in the 21st Century, Part 1: Epidemiology, Risk Factors, Screening, Prevention, and Diagnosis. *Mayo Clinic Proceedings*, *82*(3): 364-380.

Xie, H., McHugo, G., Drake, R., & Sengupta, A. (2003). Using Discrete-Time Survival Analysis to Examine Patterns of Remission From Substance Use Disorder Among Persons With Severe Mental Illness. *Mental Health Services Research 5*(1): 55-64.