In the fight against cancer, at this point in medical science, only one method seems to have any big scale hope for the patient surviving the cancer, chemotherapy. The big downsides to this are the patient isn’t guaranteed to enter remission of the cancer, and worst of all, chemo can be just as lethal as the cancer itself. In the ongoing technological march to a better tomorrow, scientists are searching for ways such that the patient will enter remission, cure the cancer diseases themselves or make the chemotherapy safer for the patient.

Chemotherapy agents are drugs that are administered to a patient in the hopes of helping the chemotherapy as a whole be more effective. The unfortunate downside is these agents can be just as bad as the radiation therapy. For instance, the Alkylating agents directly damage the cell DNA to stop cells, both healthy cells and cancerous cells, from reproducing. The downside to this is it can damage the patient’s bone marrow, and in rare cases, can cause acute leukemia. In other words, the very drug used to treat one cancer can cause another type of cancer to manifest. Future studies hope to find chemo agents that help the chemotherapy be more effective yet don’t harm the patient, if even to a degree. A study published in *Clinical Cancer Research* December 1st, 2015 shows some hope in non-small cell lung cancers. It mentions in both the 80’s and the 90’s, several independent trials of various agents including cisplatin-based chemotherapy agents improved patients’ survival, improved their quality of life and relieved symptoms ([article](http://clincancerres.aacrjournals.org/content/4/5/1087.short)).

Other, alternative therapies are also being tried and some having significant, positive effects on patient survival. *Bloodjournal*, a journal on chronicling hematological research, published an article on March 1st, 2015 on the effects using novel therapies, what they call stem-cell transplantation, versus regular chemotherapies. Among the 387 patients that relapsed after using stem-cell transplantation, they state a clear improvement in the overall survival from time of relapse was seen, 23.9 months versus the 11.8 months of the control group. It also mentions new drugs were used as another experimental agent, and those too showed greater survival times from relapse (30.9 months in the experimental versus 14.8 months for the control) ([article](http://www.bloodjournal.org/content/111/5/2516.full)).

The data used in this study was collected from *The Statistical Analysis of Failure Time Data* by JD Kalbfleisch and R.L. Prentice (1980) with publishing rights going to John Wiley and Sons. The data used comes from the University of Massachusetts’s statistical data website (<http://www.umass.edu/statdata/statdata/>) . The data from the website comes from a larger study carried out by the Radiation Therapy Oncology Group who were looking at patients who specifically had squamous carcinoma in the oropharynx regions, their mouth and throat. The group focused on fifteen target areas across sixteen participating institutions. It’s worth noting the data from the website picked only three of the oropharynx regions studied from the six largest institutions that participated. Patients that entered the study were randomly assigned to one of two groups, those that received radiation therapy alone and those that received radiation therapy along with a chemotherapeutic agent. The purpose of this was to compare the survival time of the two groups and see what, if any, differences existed between them.

The data itself was categorized by 13 variables: the case number, institution, the subjects’ gender, the treatment type (the traditional chemotherapy versus the experimental treatment), the grade, that is how different are the patient’s normal cells in comparison to the cancer cells., one was there was a very notable difference in physical appearance between the two types, all the way down to three where the two were virtually identical, the patient’s age, their current condition based upon a four-point scale ranging from having no disability to restricted work, to requiring assistance with self-care to bed confinement, the site of the cancer, the stage of the primary tumor, the T\_STAGE, if there was any evidence of metastasis, the N\_STAGE, the date the patient entered the study, their status if they were either censored in the study or died and the survival time in days from the day they were diagnosed.

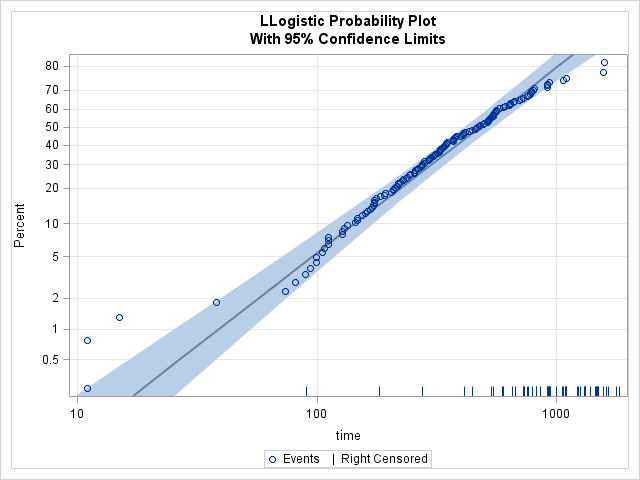
One hundred and ninety-five subjects were used in this experiment. Of them, there were 149 males and 46 females. 100 patients were in treatment group 1, the control, and 95 were in treatment group 2, the group who received the standard treatment plus the experimental agent. At the end of the time of analysis, 142 had died and 53 were censored. Just a note, some of the censoring factors include the patient moving to a different institution not being covered under the study, though those were rare according to the source of the data. Some of course were alive at the end of their study time as well. The average age of the patients was 60 years old with a standard deviation of 11 years. A short summary of these statistics can be found below.

|  |  |
| --- | --- |
| Bar chart of sex | Bar chart of tx |
| Bar chart of status | The SGPlot Procedure |

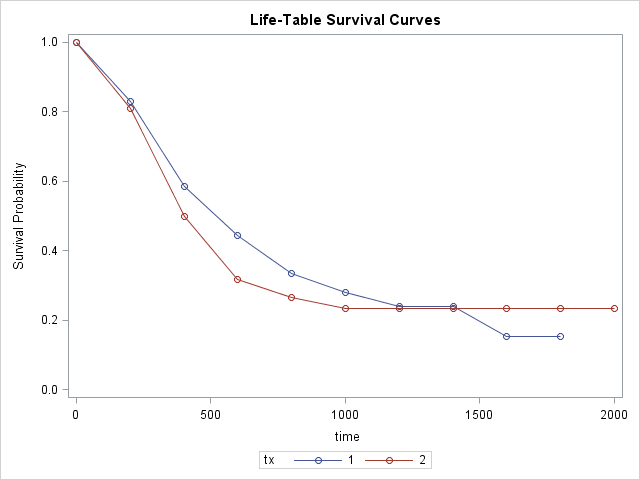
While the big focus on the data was the survival time based upon the treatments (tx), six other variables were tested against the survival time to determine if any of these variables had a significant impact on the survival time. The seven variables are as followed: treatments, grade, age condition (cond), site, t\_stage, and n\_stage.

SAS was used to analyze the data. The data was tested against three distributions, Weibull, log-normal and log-logistic. It was determined that the log-logistic had the best fit for the model. Utilizing the lifereg procedure, the parametric survival analysis, and setting the alpha to 5% the results of our test are as followed.

| **Analysis of Maximum Likelihood Parameter Estimates** | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Estimate** | **Standard Error** | **95% Confidence Limits** | | **Chi-Square** | **Pr > ChiSq** |
| **Intercept** | 1 | 8.1843 | 0.6837 | 6.8443 | 9.5243 | 143.30 | <.0001 |
| **tx** | 1 | -0.1504 | 0.1450 | -0.4345 | 0.1338 | 1.08 | 0.2997 |
| **grade** | 1 | 0.1826 | 0.1165 | -0.0457 | 0.4110 | 2.46 | 0.1170 |
| **age** | 1 | -0.0000 | 0.0067 | -0.0132 | 0.0131 | 0.00 | 0.9941 |
| **cond** | 1 | -0.7599 | 0.1320 | -1.0186 | -0.5011 | 33.13 | <.0001 |
| **site** | 1 | 0.0090 | 0.0608 | -0.1101 | 0.1282 | 0.02 | 0.8821 |
| **t\_stage** | 1 | -0.2773 | 0.0958 | -0.4650 | -0.0896 | 8.38 | 0.0038 |
| **n\_stage** | 1 | -0.1645 | 0.0624 | -0.2868 | -0.0421 | 6.94 | 0.0084 |
| **Scale** | 1 | 0.5442 | 0.0389 | 0.4731 | 0.6260 |  |  |

A plot of the data to determine the goodness-of-fit.

And finally, utilizing the lifetest procedure to obtain the survival curve.



Based upon the results of the lifereg procedure, the two treatments showed no significance between each other, that is, the chemotherapeutic agent had no significant impact on increasing the patients’ survival times. This is further reinforced by the survival curve. If we make the model of our variables we can get a far better, bigger picture of what is going on

The significant variables to the survival time was the patient’s condition, the size of the tumor and if there was any evidence of metastases, and our model reinforces this idea. The biggest value is the patient’s condition at -0.7599 followed by the t\_stage at -0.2773, +.1826 for the grade and n\_stage at -0.1645. With the condition alone, this variable influences the data the data the most.

This leads to our third and final test that we ran. Using the proportional hazard regression procedure (proc phreg) to find the hazard ratio, the results are as follows.

| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **DF** | **Parameter Estimate** | **Standard Error** | **Chi-Square** | **Pr > ChiSq** | **Hazard Ratio** |
| **tx** | 1 | 0.14603 | 0.17798 | 0.6732 | 0.4119 | 1.157 |
| **grade** | 1 | -0.17300 | 0.13769 | 1.5787 | 0.2090 | 0.841 |
| **age** | 1 | -0.0004488 | 0.00874 | 0.0026 | 0.9590 | 1.000 |
| **cond** | 1 | 0.86243 | 0.15534 | 30.8234 | <.0001 | 2.369 |
| **site** | 1 | -0.04460 | 0.07478 | 0.3557 | 0.5509 | 0.956 |
| **t\_stage** | 1 | 0.24402 | 0.13124 | 3.4571 | 0.0630 | 1.276 |
| **n\_stage** | 0 | 0 | . | . | . | . |

Prior to the test, it was found that the proportional hazard of n\_stage was violated, therefore it was stratified in this test and the results show it as zero. That aside, we can utilize the hazard ratios to make quantified statements about the variables. Taking the two remaining significant ones into account, the person’s condition constitutes a 2.3 greater chance they will die per each increase in condition, that is, someone who is at 1 has a 2.3 greater chance of dying versus someone who is at a zero. Similarly, a person has a 1.2 greater chance of dying per increase in the t\_stage.

In summary, while it’s true there do exist chemotherapy agents that do help with the survival time of a patient, which ever one was used in this study appears to be statistically insignificant in regards to patient survivability. Three other variables of interest to us that did appear statistically significant was the patients’ condition, the stage of the cancer and if said cancer had shown any sign of metastases. The American Cancer Society backs our findings up. On an article on non-Hodgkin lymphoma, they mention an international prognostic index. It lists factors that will give a picture of a patient’s prognosis and survival. Of those listed, one of them is the ability to perform daily activities and the other being the stage of the lymphoma ([article](http://www.cancer.org/cancer/non-hodgkinlymphoma/detailedguide/non-hodgkin-lymphoma-factors-prognosis)), stating the more the patient can do on their own and the earlier the cancer is detected, the better the chance of the patient surviving. A June 2011 article of *The King’s Fund*, an English think tank involved in the health system in England, conducted a study on the status quo of England’s efforts to improve its cancer survival rates. While the article focused heavily on comparing England’s cancer survival rate to other countries and how it can improve itself from there, within the article, it too echoes our study’s findings ([*The King’s Fund* article](https://www.kingsfund.org.uk/sites/files/kf/How-to-improve-cancer-survival-Explaining-England-poor-rates-Kings-Fund-June-2011.pdf)) on the patient’s condition and early detection. In conclusion, healthier people and early detection of cancer cells are crucial in increasing ones chances for survival. Further research could be considered in respect to one’s lifestyle, comparing people who scored a zero in the condition category and further dividing those people based upon their daily exercise activities, what they eat, and other things like that. It certainly warrants future investigation.

\*Appendix

/\* Elliott Light

STAT 654 Final Project. Fall 2015 \*/

/\*Variables

1) Case - Case number, i.e. the ID.

2) INST - Participating Institution.

3) SEX - 1=male, 2-female.

4) TX - Treatment: 1=standard, 2=test.

5) GRADE - 1=well differentiated, 2=moderately differentiated, 3=poorly differentiated, 9=missing

6) AGE - Age of subject in years at time of diagnosis.

7) COND - Condition: 1=no disability, 2=restricted work, 3=requires assistance with self care, 4=bed confined, 9=missing

8) SITE - Site of cancer: 1=faucial arch, 2=tonsillar fossa, 3=posterior pillar, 4=pharyngeal tongue, 5=posterior wall

9) T\_STAGE - Stage of the tumor and size: 1=primary tumor measuring 2 cm or less in largest diameter,2=primary tumor measuring 2 cm to 4 cm

in largest diameter with minimal infiltration in depth, 3=primary tumor measuring more than 4 cm, 4=massive invasive tumor.

10) N\_STAGE - 0=no clinical evidence of node metastases, 1=single positive node 3 cm or less in diameter, not fixed,

2=single positive node more than 3 cm in diameter, not fixed, 3=multiple positive nodes or fixed positive nodes.

11) ENTRY\_DT - Date of study entry: Day of year and year, dddyy

12) STATUS - 0=censored, 1=dead

13) TIME - Survival time in days from day of diagnosis. \*/

/\* Will use tx, grade, age, condition, site, t\_stage and n\_stage as independent variables. \*/

/\* Gets the data entered and formatted as needed. \*/

**data** cancer;

infile "C:\Users\Elliott\Documents\STAT654\Final project\pharynx\_copy.csv" missover dlm="," firstobs=**2**;

input case $ inst sex $ tx grade age cond site t\_stage n\_stage entry\_dt status time; /\*status is the censor. \*/

if sex then do; /\* Rewrites the "sex" collum as "male" and "female" as opposed 1 and 2 respectfully. \*/

if sex =**1** then sex='male';

else if sex=**2** then sex='female';

end;

**proc** **print** data=cancer (obs=**5**); /\*Prints the first five observations. This is used mostly to verify the code went through sucessfully. \*/

**run**;

**proc** **freq** data=cancer;/\* Gives us a frequency count of the relavent variables, not necessarily of just the ones we'll use. \*/

tables sex tx grade cond site t\_stage n\_stage status/ nocum; /\*the "nocum" option leaves out the cumulative frequency. \*/

**run**;

/\*Makes individual bar chars of each of the variables.. \*/

goptions device=gif/\*formats the chart outputs \*/

ftext= 'Arial/bold' htext=**16**pt

hpos=**80**

;

**run**;

**proc** **gchart** data=cancer;

title "Gender.";

pattern1 color=lightred;

vbar sex / discrete width=**15** freq;

title "Gender";

**run**;

title;

**quit**;

**proc** **gchart** data=cancer;

title "Treatment Groups.";

pattern1 color=blue;

vbar tx / discrete width=**15** freq;

**run**;

title;

**quit**;

**proc** **gchart** data=cancer;

pattern1 color=darkgreen;

title "Status at the end of the subject's study.";

vbar status/ discrete width=**15** freq;

**run**;

title;

**quit**;

**proc** **means** data=cancer; /\*Gets the average of the ages. \*/

var age;

**run**;

**proc** **template**; /\* Formats the histogram we'll make for "age". \*/

define style style.histo;

parent=styles.listing;

style graphfonts from graphfonts /

'graphdatafont' = ('arial', **16**pt);

'graphlabelfont' = ('arial',**16**pt);

'graphtitlefont'=('arial',**16**pt);

end;

**run**;

ods listing style=histo;

**proc** **sgplot** data=cancer; /\*Makes the histogram. \*/

title "Age Distribution";

histogram age / fillattrs=(color=lightblue); /\*Fills the bars with a lightblue color. \*/

density age / lineattrs=(color=red); /\*Creates a line of the data, age, to determine if it's normally distributed anc colors it red. \*/

**run**;

title;

**quit**;

/\* Begining of the survival analysis. \*/

/\*LIFETEST gets the survival curive and the Kaplan-Meier estimators. \*/

ODS graphics on;

**proc** **lifetest** data=cancer plots=(s,h,ls) method=life; /\*Does the lifetest using the "life" method. \*/

time time\*status(**0**); /\* Our model. The dependent variable, time, with the censor status. \*/

strata tx; /\* Compares the whole data against the two treatment groups, tx. \*/

**run**;

**quit**;

ods graphics off;

/\* Runs the LIFEREG method. Comparing the Weibull distribution. \*/

**proc** **lifereg** data=cancer;

model time\*status(**0**) = tx grade age cond site t\_stage n\_stage / distribution=weibull;

probplot;

**run**;

/\* Runs the LIFEREG method. Comparing the log-normal distribution. \*/

**proc** **lifereg** data=cancer;

model time\*status(**0**) = tx grade age cond site t\_stage n\_stage / distribution=lnormal;

probplot;

**run**;

/\* Runs the LIFEREG method. Comparing the log-logistic distribution. \*/

ods graphics on;

**proc** **lifereg** data=cancer;

model time\*status(**0**) = tx grade age cond site t\_stage n\_stage / distribution=llogistic;

probplot;

**run**;

ods graphics off;

/\* Use the log-logistic distribution. It has the best fit of all the models. \*/

/\* The significant variables are the condition, t\_stage, n\_stage \*/

/\*Utilizing the Cox PH test to obtain th hazard ratio. \*/

ods graphics on;

**proc** **phreg** data=cancer;

model time\*status(**0**)=tx grade age cond site t\_stage n\_stage / ties=efron;

strata n\_stage; /\* We stratify since its prop haz assumption is violated. \*/

assess ph / resample;

**run**;

ods graphics off;