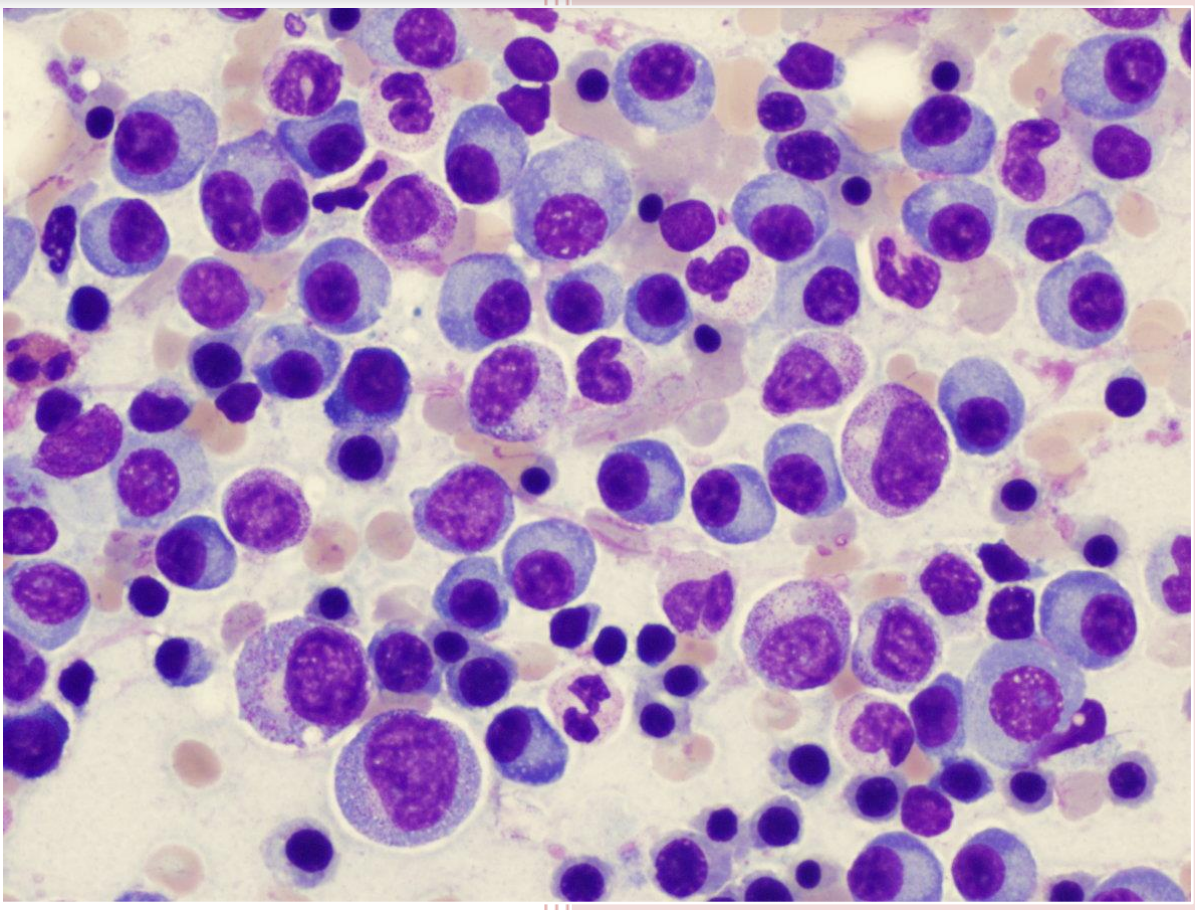
A Survival Study on Multiple Myeloma



Samuel Young Annan

The University of Iowa

12/1/2010

Photo credit: www.uthoustonim.blogspot.com

Introduction

Although a relatively rare form of cancer, Myeloma is the most common type of plasma cell and bone cancer in adults. Myeloma begins when a plasma cell undergoes malignant transformation and forms clumps in the bone marrow through self-division. They may damage certain organs and the solid part of the bones at different sites causing "Multiple Myeloma" (MM). The cells make excessive deficient monoclonal antibodies called 'M' reducing the patient's ability to fight infections. While bones may not be involved at onset, osteolysis is likely to set in later making bones susceptible to fracture. As a result of osteolytic lesion, blood calcium increases by great proportions. Tissue breakdown is an MM trademark and accelerated levels of calcium and lactate dehydrogenate are both indicators of tissue breakdown. Due to high rate of turnover of cancer cells, the destroyed cells elevate LDH activity making LDH levels an interesting variable in the study of disease progression. In other findings, high LDH levels were present in cases where disease progress rate was very fast with mean survival rate of about three months.

[1]Research has also shown that multiple and complex chromosomal abnormalities (both structural and numeric) are present in MM cases. Reports describe disparate clinical courses for patients with specific chromosomal/ cytogenic abnormalities. What is probably the most important 'prognostic' parameter is the beta$_2$-microglobulin level (β2M). [2] β2M is a light chain protein that is elevated in all lymphoproliferative diseases such as myeloma. Other [3]researchers have concluded that β2M is a valuable marker for assessing initial tumor mass and response to chemotherapy for patients with MM.

When bone marrows become crowded due to overproduction of myeloma cells, lesser quantities of normal blood cells are produced resulting in lower levels of hemoglobin. The Durie and Salmon stage III MM classification suggests the presence of Hemoglobin value less than 8.5 g/dl in the final stage of the disease. Serum albumin is the most abundant protein in blood plasma and low levels of the albumin are correlated with severity of MM. The reason for this is not clear although it is to be expected considering that there are huge protein losses in MM patients. Another possible explanation is the fact that calcium bonds with albumin. It is also known that the level of albumin is inversely related to β2M levels. The levels of albumin, hemoglobin, LDH and β2M will be studied for the ability to act as markers of disease progression for patients in the data presented. While the causes of the disease remains unknown, research has shown that certain risk factors increase the chance of the disease developing in people

although they do not guarantee that a person will develop myeloma. They include aging (age over 65), race (highest myeloma occurrence found in African Americans and lowest rates in Asian Americans), gender (the ratio of MM in men to women is about 3:2), family history of MM and long exposure to certain chemicals and materials. Many other suspected risk factors are under study such as the effect of genetic alterations, diet, or obesity on MM risk. In this study, risk factors will be used to model survival of patients to determine which factors are significant. Once a person is diagnosed with the disease, treatment begins usually with steroids like Dexamethasone (Dex) which is described by the International Myeloma Foundation as the most effective single agent for treatment of multiple myeloma when given in large doses. Dex can be used alone to prevent white blood cells from infected areas and actually kill myeloma cells before stem cell transplant or after the return of the disease. It may also be used at different doses in combination with other drugs (e.g. Immunomodulators such as Thalidomide) to increase the effectiveness of those drugs. It is believed that the combinations can provide an improved outlook for patients. Since 1996 Thalidomide has been used at an increased rate with Dex due to marked improvement in the survival rate of treated patients. This study will compare the survival rates of two groups; one treated with Dex alone and the other treated with Dex and Thalidomide to ascertain that observation.

Methods

Data was collected on 318 individuals at their enrollment into the study which began in 2000. At the time of last contact, 128 (about 40% of the initial risk set) of these individuals had died with a mean survival of 3.2 years. The mean time spent in the study for censored individuals was about 7.3 years. Demographic variables, including race, gender and age, are summarized in Table 1 below. Age at last contact ranged from 31 to 82 with mean 61 years.

	Race				**Age**			**Gender**		
	White	Black	Hispanic	Unknown	≤ 50	$50-65$	≥ 65	Male	Female	**total**
Frequency	281	30	4	3	93	173	52	178	140	318
Percentage	88.36	9.43	1.26	0.94	29.25	54.40	16.35	55.97	44.03	100

Table 1 Demography Descriptive Statistics

A preliminary analysis of the survival data is shown in Table 2 in an attempt to get a glimpse of what the data presents. The third and fourth rows represent the mean time to death for those who died and the mean time to last contact. From the results, the control group seems to have a higher failure rate that the treatment group although

acceleration to death seem faster for the treatment group. The amount of time spent under observation on average was comparable for both groups and so the relatively low failure percentage may not be attributable to a short exposure. Considering the varied proportions of racial groups in the study, the effect of race will be left for further analysis. It is however quite clear that failure rate increases monotonously with increasing age with the greatest acceleration to death in the 50 to 65 year bracket. Also, men show a higher failure rate and lower acceleration to death than women. Possible correlation between variables may present alternative explanations for these trends.

	Treatments		**Race**				**Age**			**Gender**		**ALL**
	Ctrl	Treat	White	Black	Hispanic	Unknown	≤ 50	$50-65$	≥ 65	Male	Female	
N	164	154.00	281	30	4	3	93	173	52	178	140	318
% Failure	45.12	35.06	40.93	40.00	0.00	33.33	31.52	42.20	50.00	43.82	35.71	40.25
$\mathbf{E(T^{\circ} \mid T^{\circ} < c)}$	3.48	2.88	3.22	3.10	0.00	5.36	4.12	2.50	3.17	3.47	2.85	3.23
$\mathbf{E(T^{\circ} \wedge c)}$	5.57	5.79	5.63	5.70	7.18	7.86	6.29	5.51	5.13	5.71	5.63	5.68

Table 2 Comparison of Survival Data among Various Categorical Variables

In-depth analysis will be made with Time to last contact as the outcome variable will all variables included in the initial model. Bivariate analysis on predictor variables will be used to test equality across strata using the Log-Rank test in conjunction with Kaplan-Meier survival curves. An accelerated failure time parametric model will be fit to the data and compared with the Kaplan Meier Survival curve to check for fitness. Maximum likelihood estimates of parameters for the various coefficients of the predictors will be tested for significance to help decide which variables to retain in the final model. Also, the semi-parametric Cox proportional hazard model will be fitted to the data as a possible alternative model. The Cox PH model will be used to compare the treatment group to the control group after checks for proportionality have been performed to be sure that the underlying assumptions of the model have not been violated. The Cox PH model will be tested for goodness-of-fit. The primary statistical software used was R.

Results

The rate of disease progression is of interest in this analysis and the International Staging System (ISS) will be used. Ideally, the ISS should be used in addition to the Durie-Salmon staging criterion however the present data does not have information about calcium levels, bone lesions, or urinary light chain excretion. The ISS combines albumin levels with β2M and classifies disease progression in three stages. The staging system is presented in Appendix A1. The variable ISS has three levels (1-3) for the three stages (I-III). The control group (coded 1) represents those

treated at inception with only Dex while the Treatment group (coded 0) represents those treated also with thalidomide. The failure indicator is defined as 1= death and 0=censored. A covariate for cytogenic abnormalities gives the code 1 for patients with abnormalities present and 0 otherwise. A summary of the level of genetic risk (1 or 0 for high or low respectively) based on the patient's history will be added to the model. Ages are categorized into three groups; 1 – at most 50 years, 2 – between 50 and 65 years, and 3 – at least 65 years. Males will be coded 1 and females with a 0 for the gender variable. Due to very low numbers of Hispanics and unknowns (total of 7 out of 318) in the data, the analysis will attempt to compare whites and blacks mainly.

Figure 3 on the next page is a non-exhaustive pictorial display of what the information collected in the data. Looking at the K-M survival curves, we find that a lot of the trends meet expectations. The ISS curves show that multiple myeloma patients in stage I of the disease have a better survival trend than those in stages II and likewise, the stage II curve is better than the stage III survival curve. The curves also suggest that, all things equal we should expect female patients with the disease to survive longer than male patients and survival rate to decrease with higher age categories. Patients whose genetic profile suggest a high risk of contracting the disease have a distinctively lower survival trend; similarly, patients with chromosomal irregularities have a much lower survival curve as compared with those without those abnormalities. Race is however much more difficult to decipher, considering firstly that there are only a few observations from the Hispanic and unidentified racial categories. Among blacks and whites though, the curves seem to confirm numbers from the demographic summaries in Table 2 on the previous page. The race curves seem to move in concert; however the relatively small sample size for blacks (30) could skew analysis from reality.

The analysis on the next few pages will delve into these observations to see how significantly they affect survival of patients as a whole and attempt to arrive at a model (either parametric or semi-parametric) that best summarizes the data. Other variables that are of interest are LDH levels, time to progression of the disease, duration of induction prior to bone marrow transplantation, length of time of drug use for both treatment groups, time to the first significant response to treatment, and the dosage of Dexamethasone used throughout the induction period.

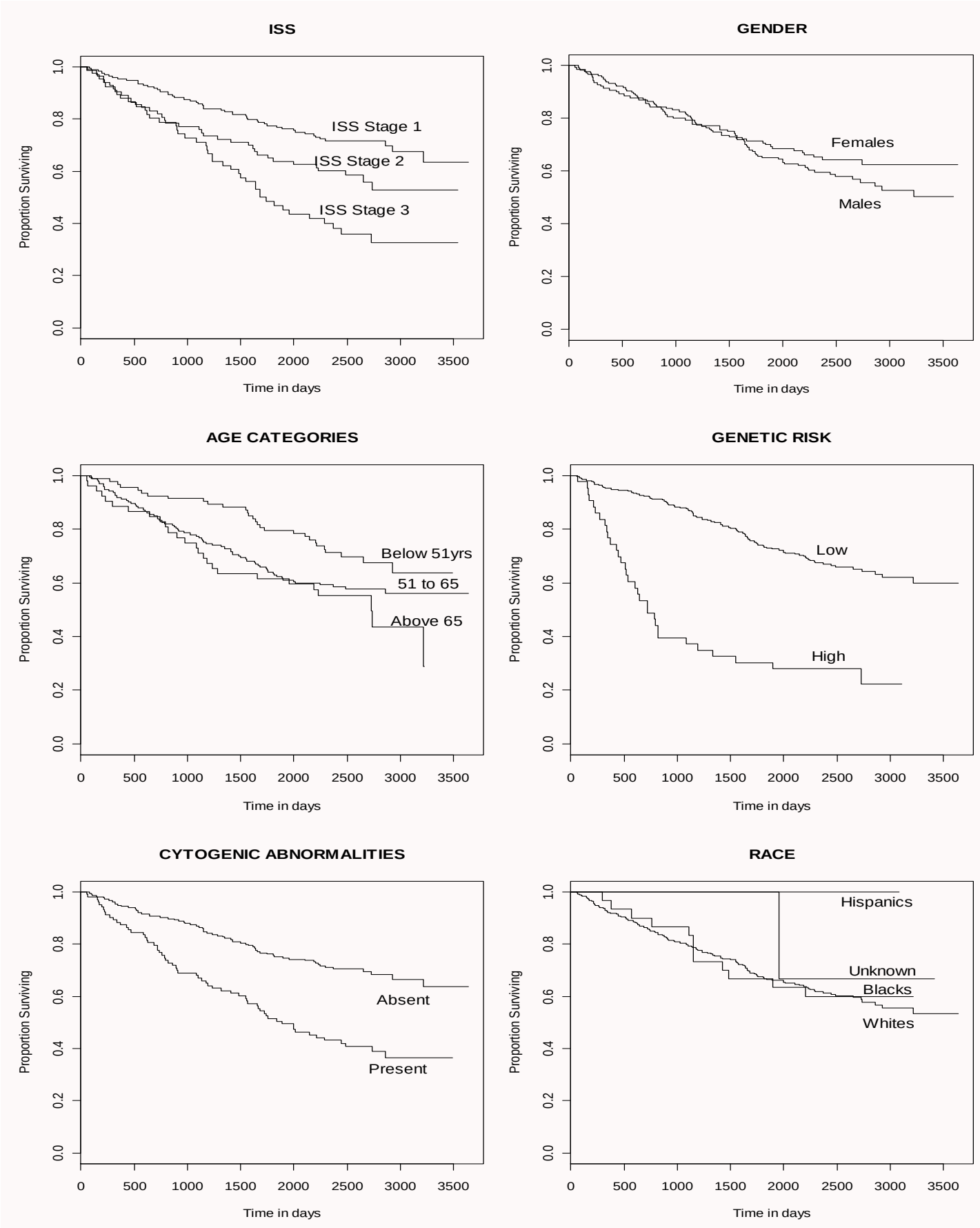


Figure 3 Kaplan-Meier Survival Curves for Various Categories

Accelerated Failure Time models have general form of $lnT = \alpha + \beta' X + \sigma W$, where W is a distribution that has a parameterization that includes a scale parameter. A visual comparison of the KM curve and the Exponential and Weibull curves in Figure 1(a, b) shows a reasonable fit for both models and the two distributions are very similar with lambda values approximately equal for both distributions. The choice of the exponential and Weibull distributions is based on the fact that they are the only two parametric distributions that exhibit the property to be parameterized both as AFT models or proportional hazards models. The assumption that the hazard function is always monotonic (based on the Nelson-Aalen curve in appendix A2) is being made to allow for the use of Weibull.

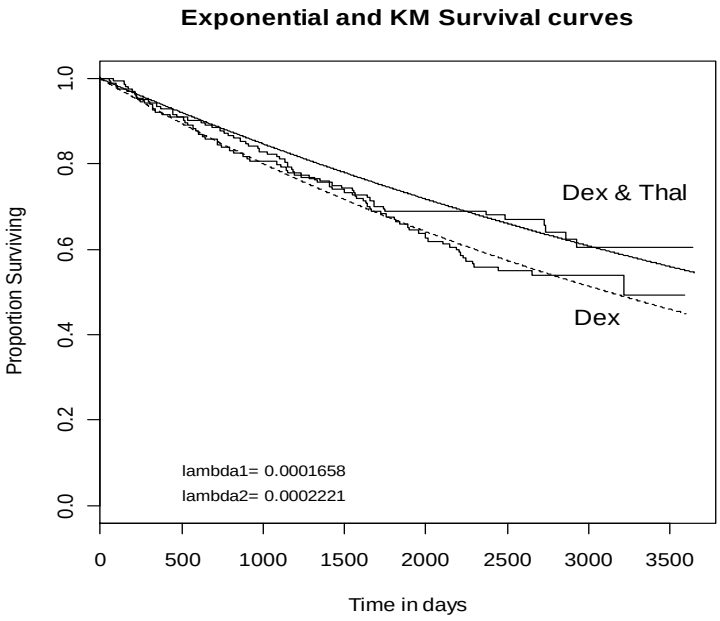


Figure 2a KM and Exponential Survival Curves by Treatment

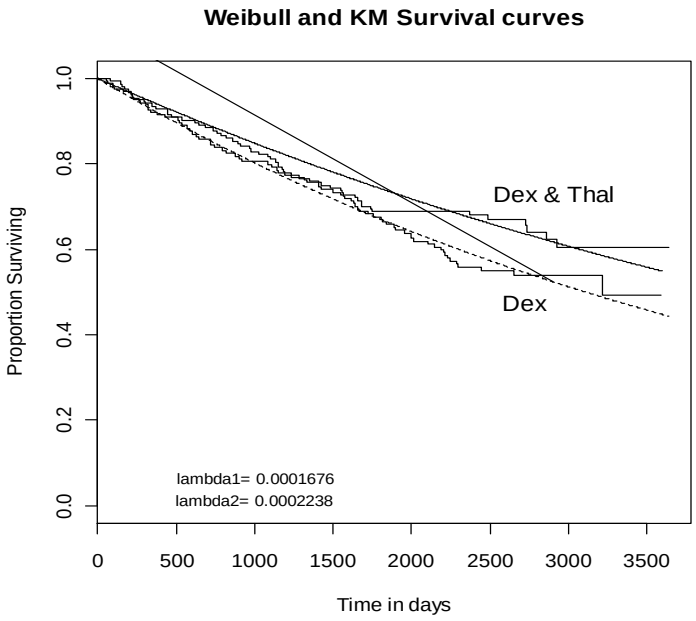


Figure2b KM and Weibull Survival Curves by Treatment

A Weibull regression analysis on the data for various models with and without the ISS variable yielded the following results shown in table 4(a & b). The shaded cells represent statistically significant variables in the model.

Model	-2*Loglik	$\hat{\sigma}$	Treat	Cytogenic	LDH	ISS 3	Gen Risk	Age Cat 2
1	2441.00	0.99	-0.2891					
2	2416.60	0.96	0.3632			-1.0538		
3	2387.20	0.93	-0.4086			-0.6705	-1.0859	-0.3927
4	2440.00	0.99	-0.2101					
5	2388.20	0.92	-0.2464	-0.6666	-0.0029			
6	2384.60	0.92	-0.274	-0.6557	-0.0027	-0.5111		
7	2385.00	0.92	-0.3764	-0.721	-0.003	-0.7216		-0.3737*

Table 4a Weibull Regression Analysis of Multiple Myeloma Data with ISS

The table summarizes models due to the sheer size of some of them and shows the most significant variables along with their log-likelihood values. When treatment, ISS, age categories, genetic risk and cytogenic abnormalities were added to the model, σ was close to $1(\gamma = 1)$ suggesting an Exponential or Weibull fit. However, when time to progression, age categories and duration of Dex usage were included in the model, treatment and ISS were not significant and σ was about 0.5 $(\gamma = 2)$ making both the Exponential and Weibull distributions unlikely candidates for the model. Based on the log-likelihood values, the best models are 7 (in the presence of ISS and with *trt* significant) and 8(in the absence of ISS). A model with all variables accepts the same variables as model 14 with a marginal improvement in the log-likelihood (1300.8). In model 7 *agecat1* is marginally significant (sig at 7% level).

Model	-2*Loglik	$\widehat{\sigma}$	Treat	Progress	Agecat1	Agecat2	DurDex	DurThal
8	486.00	0.40	0	7.42E-04	-0.808	-0.862	4.26E-04	4.34E-06
9	520.80	0.45	0	8.64E-04	-0.788	-0.822	3.92e-04	2.66E-05
10	1308.00	0.46	0.1284	7.4E-04	-0.3406	-0.5974	3.44E-04	
11	1309.60	0.46	0.1286	7.2E-04	-0.3268	-0.5843	3.63e-04	
12	534.00	0.55	0	0.0015	-0.7006	-0.7805		0.0002
13	534.40	0.55	0	0.015	-0.6964	-0.8397		0.0002
14	1307.00	0.46	0.085	0.0007	-0.366	-0.6108	3.52E-04	

Table 4b Weibull Regression Analysis of Multiple Myeloma Data without ISS

There seem to be little reason to choose either model over the other and so both equations will be stated. The final equation for model 7 is

$$ln(time) = 9.99 - 0.376 * trt - 0.721 * cyto - 0.003 * ldh - 0.722 * iss3 - 0.374 * agecat2 - 0.4 * agecat3 + 0.9W$$

and for model 8,

$$ln(time) = 6.17 + 0.0007 * progt - 0.0006 * respt + 0.135 * iss2 - 0.381 * iss3 - 0.001 * ldh - 0.615 * iss3 - 0.808 * agecat2 - 0.862 * agecat3 + 0.4W.$$

Like all AFT models, these time models are very sensitive to the choice of the distribution of the hazard function. Another look at the Kaplan Meier Survival curves and Nelson-Aalen Cumulative Hazard functions for the two groups (figure 1 of Appendix A2), shows that the group treated with Dex only starts off with higher survival rates prior to time$(\hat{S}_{PL})$ 1237(0.773) after which it moves closely with the Thal group until the touch again at 1423(0.74) and finally separate after 1657(0.714) with Dex having a consistently lower survival rate. The cumulative hazards model

tells the same story with the Thal group exhibiting a faster acceleration to death in the first 1237 days and then maintaining a relatively lower hazard rate consistently after time 1657 days. The one-year survival rate is 0.92 for the treatment group and 0.93 for the control group however the three-year survival rate for the treatment group is higher at 0.68 against 0.6 for the control group. The curves move together after $t = 1657$, and it would be interesting to investigate the difference between the two groups further. To do that a Cox Proportional Hazards model will be fit to the data.

An analysis of the maximum likelihood estimates of the regression coefficients is given below;

Model	LRT(p-val)	Treat	Cytogenic	LDH	ISS2	ISS 3	Gen Risk	Age Cat 2
1	1.1E-01	0.282						
2	1.46E-03							
3	1.2E-09	0.730			0.368	0.703	1.170	0.431
4	4.1E-09	0.256	0.716	0.003				
5	8.3E-09	0.288	0.709	0.003	0.274	0.545		
6	1.2E-09	0.402	0.780	0.003	0.320	0.545		0.411*
Model	**LRT(p-val)**	**Treat**	**Progress**	**ISS3**	**Agecat1**	**Agecat2**	**DurInd**	**DurDex**
7	0.000	-0.160	-0.002	-0.094	0.686	0.946	0.005	-0.001
8	0.000	-0.372	-0.002	0.195	0.629	1.087		-0.001
9	0.000	-0.340	-0.002		0.687	1.191	-0.001	-0.001
10	0.000	-0.342	-0.002		0.658	1.158		-0.001
11	0.000	-0.523	-0.002		0.664	1.083	-0.001	
12	0.000	-0.575	-0.002		0.554	0.962		

Table 5 Cox Regression Outputs

The outcomes of the Cox regression tested at the 5% significance level are summarized as follows;

- In the absence of the variable '*agecat*', *iss3* was significant. The interaction between *trt* and *iss* (not shown in table) was significant only if no other variable was present.
- Time to response ('*respt'*), Hemoglobin ('*Hgb*'), gender ('*sex'*), race ('*race'*) and Duration of thalidomide ('*Dur0'*) were not significant in any model.
- Albumin and beta$_2$-microglobulin were not significant when ISS was included in the model. This is suspected to be due to the fact that ISS contains information about the two variables leading to interactions.

- Genetic risk was not significant except in the absence of other strong variables like Duration of Dex use and Treatment.
- Treatment and Duration of Dex were not both significant in any model. A cox regression of 'trt', 'dur1' and their interactions had only 'dur1' being significant favoring model 10 over 12.
- Using the LRT criterion, model 6 is the best model in the presence of '*iss*'; forcing treatment into the model, model 12 is the best in the absence of '*iss*'. Adding duration of induction '*durind*' to the model does not improve model 12.
- The best fitting models are therefore; (in presence of ISS)[*agecat coefficients were marginally significant]

  Model 6: $\lambda(t|x) = \lambda_o(t) * \exp\{0.402 * trt + 0.780 * cyto + 0.003 * ldh + 0.320 * iss2 + 0.545 * iss3 + 0.411 * agecat2 + 0.504 * agecat3\}$

  and (in the absence of ISS)

  Model 10: $\lambda(t|x) = \lambda_o(t) * \exp\{-0.342 * trt - 0.002 * progt + 0.658 * agecat2 + 1.158 * agecat3 - 0.001 dur1\}$

  Model 12: $\lambda(t|x) = \lambda_o(t) * \exp\{-0.575 * trt - 0.002 * progt + 0.544 * agecat2 + 0.962 * agecat3\}$

Covariates		Treat	Cyto	LDH	ISS2	ISS3	Agecat1	Agecat2	ProgT	Dur1	Global
Model 6	Rho	0.1718	0.0171	-0.1137	-0.0337	0.1223	-0.1448	-0.1018			**NA**
	Chi-Sq	4.0198	0.0384	1.9072	0.1532	2.0393	2.7641	1.4235			**13.5589**
	p-value	0.0450	0.8447	0.1673	0.6955	0.1533	0.0964	0.2328			**0.0596**
Model 6**	Rho		0.0198	-0.1084	-0.0316	0.1239	-0.1359	-0.0951			**NA**
	Chi-Sq		0.0517	1.7354	0.1326	2.0645	2.4247	1.2531			**7.5893**
	p-value		0.820	0.188	0.716	0.151	0.119	0.263			**0.270**
Model 10	Rho	0.262					-0.112	-0.036	0.343	0.375	**NA**
	Chi-Sq	5.931					1.096	0.115	4.818	4.802	**19.212**
	p-value	0.0149					0.2951	0.7344	0.0282	0.0284	**0.0018**
Model 12	Rho	0.337					-0.169	-0.119	0.522		**NA**
	Chi-Sq	9.32					2.20	1.18	14.96		**19.73**
	p-value	0.00227					0.13765	0.27796	0.00011		**0.00057**

Table 6 Test of Hazard Proportionality for selected models

To help decide which of the model best fits the data, the Cox-Snell residuals were generated and plotted for the three models in Figure A3 in the appendix. From the residual plots, model 10 seems to be the best fit of the three models; followed by model 6. The tails deviate strongly from the 45^{o} line, however considering that they represent only a few observations no suggestive interpretations will be made to explain the deviations. Using the Cox PH model implicitly assumes that the hazards for various covariates are proportional. The selected models cannot be used without checking for these assumptions. Table 6 on the previous page represents results from a hazard proportionality test. The R function cox.zph takes the coxph object as its argument and provides test statistics for the proportionality test along with the p-values. The components of the function make a matrix based on Schoenfeld residuals. Of the three models, model 6 was the only one which satisfied the test for most covariates including the global test. The proportionality of the treatment covariate is however in only marginally significant at 5% significance level. Model 6** is based on model 6 after stratifying on treatment groups. The plots in Appendix A4 are graphical checks for the proportionality assumptions in model 6**. The graphs show that for all covariates proportionality can be assumed since the plots are all roughly horizontal. The influence function delta-beta plots did not show any individual to be particularly influential enough to skew results of the analysis and so no further case-deletion analysis were made. A deviance plot on model 6 stratified on treatment shows two groups of points without any clear outliers (see Appendix A5).

Stratification based on treatment was done because of the belief that treatment groups differ significantly and that modeling the effect of the other coefficients, with treatment silenced, can help check the veracity of the proportionality claim. Model 6 does not contain time dependent variables and so further proportionality test will not be done for the hypothesis that $\zeta = 0$ in $\lambda(t|x) = \lambda_o(t)e^{\beta X + \zeta X \psi(t)}$ where $\psi(t)$ is a function of time. Model 6 is therefore the favorite model and is given by;

$$\lambda(t|x) = \lambda_o(t) * \exp\{\beta_1 * trt + \beta_2 * cyto + \beta_3 * ldh + \beta_4 * iss2 + \beta_5 * iss3 + \beta_6 * agecat2 + \beta_7 * agecat3\}$$

Discussion

In an attempt to understand the relationship between various risk factors and survival time, an AFT model was fit to the data to explain time to death for multiple myeloma patient by assuming a distribution for the hazard function (and survival function for that matter). The Cox Proportional Hazards model however does not make any

assumptions about the shape of the hazard function and concerns itself with the effects of the covariates on time till death. It assumes a parametric form of the effects and that the effects will be the same at all values of time. For both models, an ISS variable was created for every patient in the study. Also, ages of the participants at the time of entry were categorized into three groups; less than 50, between 50 and 65, and more than 65. All analysis was done in R and MS Excel®. Based on past research[2] the following was expected;

- Hazard rate will increase (survival will decrease) with higher covariate values of cytogenic abnormalities, lactase dehydrogenase, ISS, Age (Age Categories), genetic risks and will reduce with treatment of Thalidomide.
- Hazard rate will increase with lower values of dosage of the drug dexamethasone, duration of induction, duration of dexamethasone use and duration of thalidomide use.
- Hazard rates will be higher for men than women of the same race and higher for blacks than whites.

The analysis done in this study goes to confirm each of these expectations. It is worth mentioning that treatment seemed to be strongly associated with duration of treatments; whenever they both appeared in a model, '*trt*' was found to be insignificant. In a way, the presence of the duration of the individual treatments masked the effect of the treatment variable whenever both were present. That is to be expected since they are all telling similar stories. Also, race and gender did not play a very significant role in determining acceleration to death. For race, it may be worth noting that the relatively low representation of blacks, Hispanics and others in the data makes the comparison weak if even notable. Time to disease progression was directly and strongly correlated with time to death. Fitting the parametric AFT model, the data suggests that the exponential or Weibull could be used as approximated distributions for the underlying hazard. The favorite model (7) can be interpreted as having a maximum log(time) to failure of approximately 10 and being negative correlated with *'trt', 'cyto', 'ldh', 'iss3', 'agecat2', 'agecat3'*. Model 7 was based on the likelihood values of the fitted models. For the Cox models, selection was based on the likelihood ratio test. Since most tests had an approximate value of zero for the likelihood ratio test, other factors were taken into consideration during selection including size of model, number of significant covariates, residual diagnostics using Cox-Snell. Finally a proportionality test was performed for the three favorites from the Cox regression. The selection of model six as the final model was because it produced the best p-values for the proportionality test after the other two failed. Model 6 shows that

the hazard rate can be described as increasing for higher values of *'trt', 'cyto', 'ldh', 'agecat', 'iss'*. These were the same covariates in the final AFT model and they tell the same stories; the combination of Thalidomide and Dexamethasone treatment (0) significantly improves survival time (reduces hazard relative) relative to the Dexamethasone treatment (1), presence of cytogenic abnormalities and lactase dehydrogenase increases the hazard rate, while patients in the last two age categories (age more than 50 yrs.) and those in the second and final stages of the disease were seen to accelerate faster towards death. [4]In the study by the Arkansas Cancer Research Center, Little Rock, researchers tracked the action of Thalidomide and found that nearly all patients showed some reduction in paraprotein, a result which the lead researcher described as "stunning". That research was done with only Thalidomide as treatment, however in this study we find that the combination therapy still works favorably compared to Dexamethasone only. The hazard ratios for various groups reveal that a patient of the Dex only treatment has a hazard rate about 1.5 times that of a patient being treated under the combination therapy. Other interesting ratios were the following;

- All things equal, a patient with cytogenic abnormalities has a hazard rate about twice of a patient without abnormalities.
- Although LDH was a significant variable in the model, the hazard rate when present is only about 0.3% higher than when absent.
- A patient classified as ISS3 has a hazard rate more than twice that of a patient under ISS1 and about 1.6 times one under ISS2. A ISS2 patient has a ratio about 1.4 times a patient under ISS1.
- Though the hazard rate doesn't increase by a great deal between agecat 3 and 2, they are both significantly higher (1.7 and 1.5 times bigger) when compared with agecat1 .

Other similar results can be found in table 6 in the appendix.

In all, the findings in this study, though not exhaustive, agree in large part with other studies conducted in the past. Information about blood calcium and level of osteolytic lesions could help to get a stronger staging criterion and an even wider array of models can be tested with various distributions (AFT) and PH models.

References

1. CYTOGENETIC ABNORMALITIES IN MULTIPLE MYELOMA

   Author: Rafael Fonseca, Lionel J.A. Coignet, Gordon W. Dewald

   Publication: Hematology/Oncology Clinics of North America

2. MULTIPLE MYELOMA: RECOGNITION AND MANAGEMENT

   Author: Enitza D. George, M.D., and Richard Sadovsky, M.D., M.S.

   Publication: State University of New York Health Science Center

3. SERUM BETA-2-MICROGLOBULIN IN MULTIPLE MYELOMA: RELATION TO PRESENTING FEATURES AND CLINICAL STATUS.

   Author: Bataille R, Magub M, Grenier J, Donnadio D, Sany J.

4. EXTENDED SURVIVAL IN ADVANCED AND REFRACTORY MULTIPLE MYELOMA AFTER SINGLE-AGENT THALIDOMIDE: IDENTIFICATION OF PROGNOSTIC FACTORS IN A PHASE 2 STUDY OF 169 PATIENTS.

   Author: Barlogie B, Desikan R, Eddlemon P, Spencer T, Zeldis J, Munshi N, Badros A, Zangari M, Anaissie E, Epstein J, Shaughnessy J, Ayers D, Spoon D, Tricot G.

   Publication: Myeloma and Transplantation Research Center, University of Arkansas School for Medical Sciences

5. COMPARISON OF CASE-DELETION DIAGNOSTIC METHODS FOR COX REGRESSION

   Author: Hsiao-Mei Wang, Michael P. Jones, Barry E. Storer

   Publication: Statistics in Medicine Volume 25, Issue 4, pages 669–683, 28 February 2006

Appendix

A1 Staging Systems

- Stage I: β2-microglobulin (β2M) < 3.5 mg/L, albumin >= 3.5 g/dL
- Stage II: β2M < 3.5 and albumin < 3.5; or β2M >= 3.5 and < 5.5
- Stage III: β2M >= 5.5

A2 Kaplan Meier Curves for the Two Treatment Groups

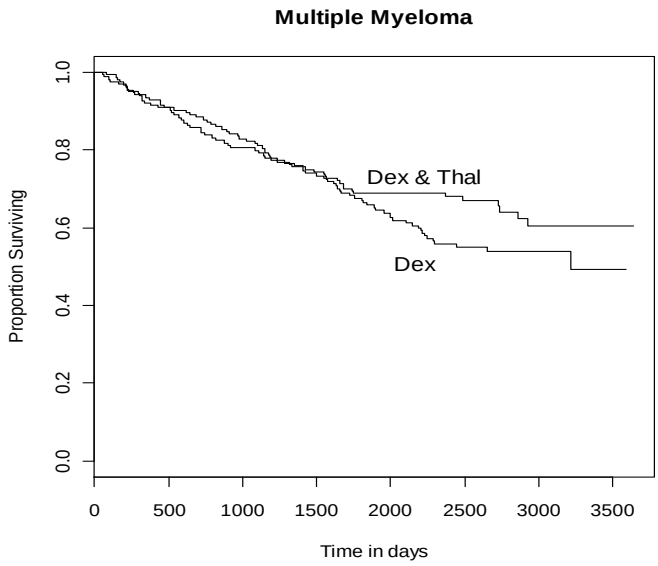


Figure A2a Kaplan Meier Curves for Treatment Groups

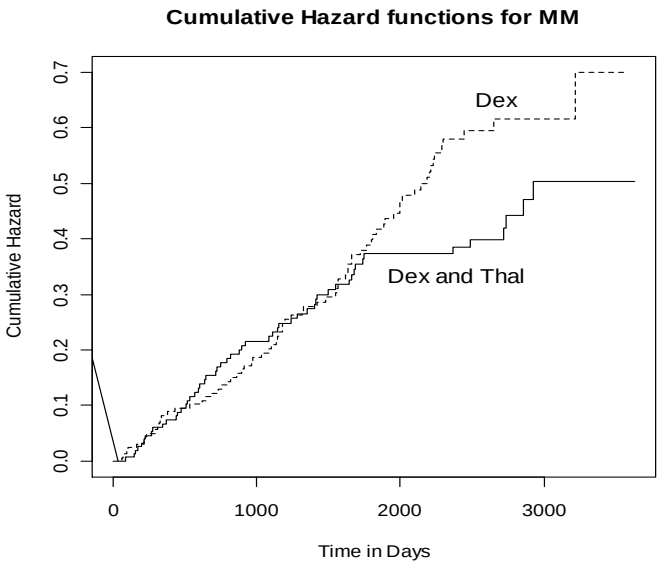


Figure A2b Nelson-Aalen Cumulative Hazard functions for Treatment Groups

A3 Cox-Snell Residual Plots for Selected Models

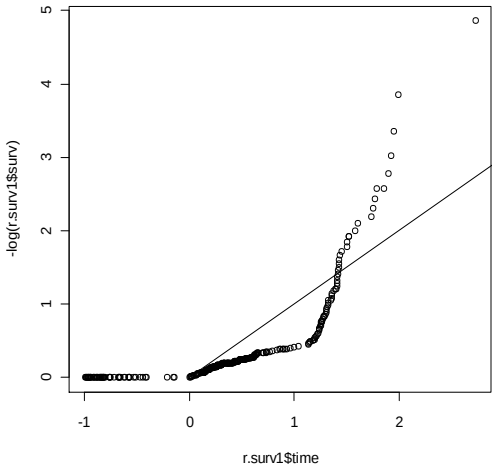


Table A3a Cox-Snell residuals for model 6

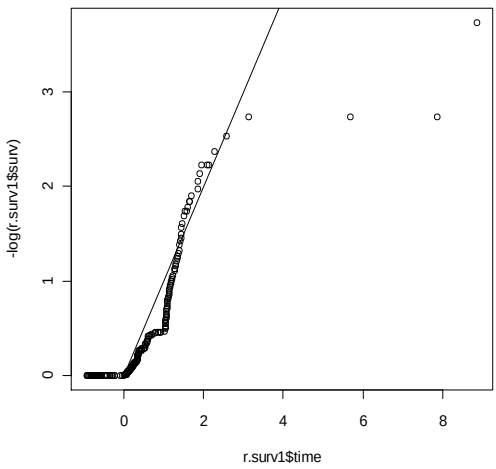


Table A3b Cox-Snell residuals for model 10

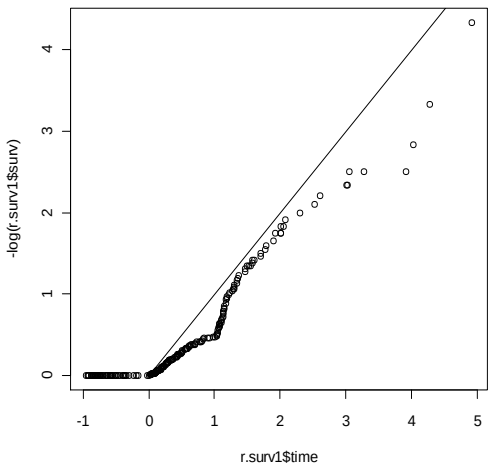


Table A3c Cox-Snell residuals for model 12

A4 Graphical Checks for Proportional Hazards Assumptions

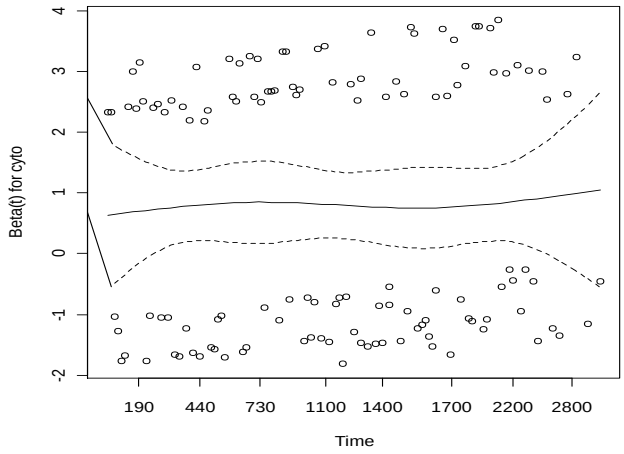


Figure A4a PH Assumption test for 'cyto'

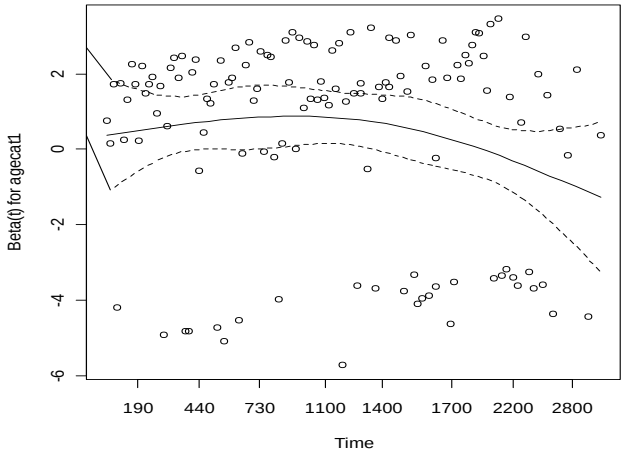


Figure A4b PH Assumption test for agecat1'

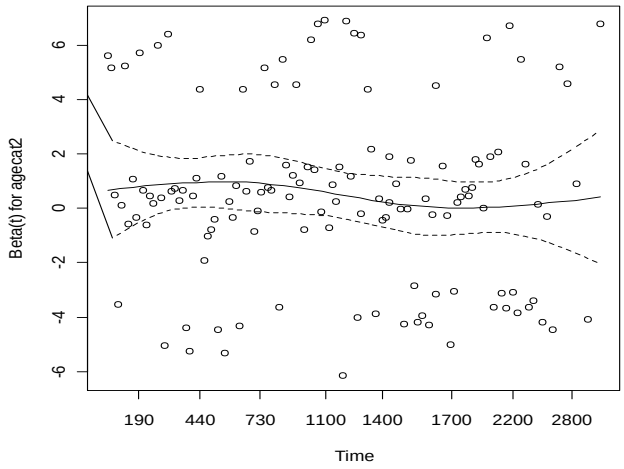


Figure A4c PH Assumption test for 'agecat2'

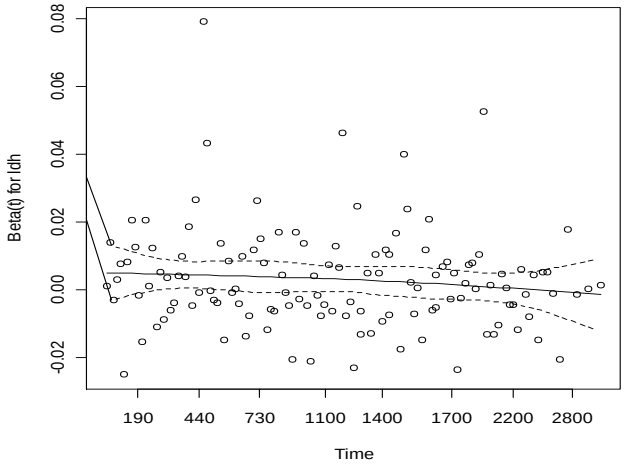


Figure A4d PH Assumption test for 'ldh'

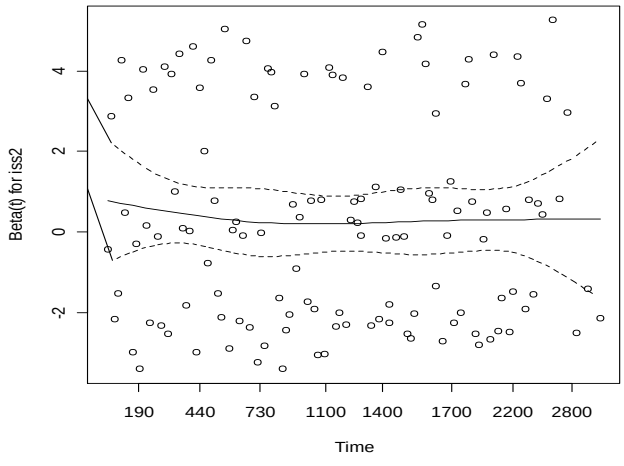


Figure A4e PH Assumption test for 'iss2'

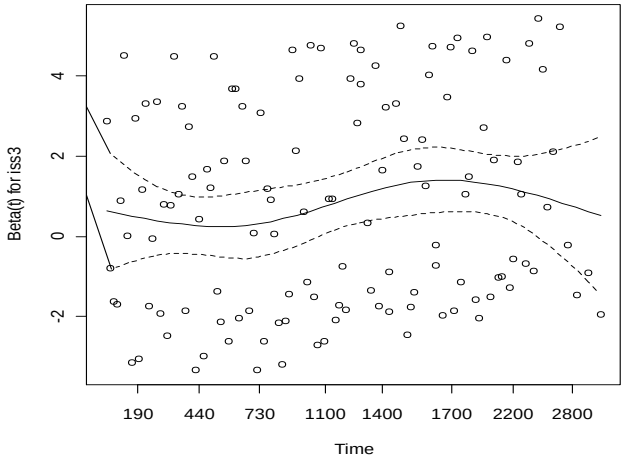


Figure A4f PH Assumption test for 'iss3'

A5 Deviance Plot for $Surv(time, status) \sim strata(trt) + cyto + agecat + ldh + iss$ (Model 6**)

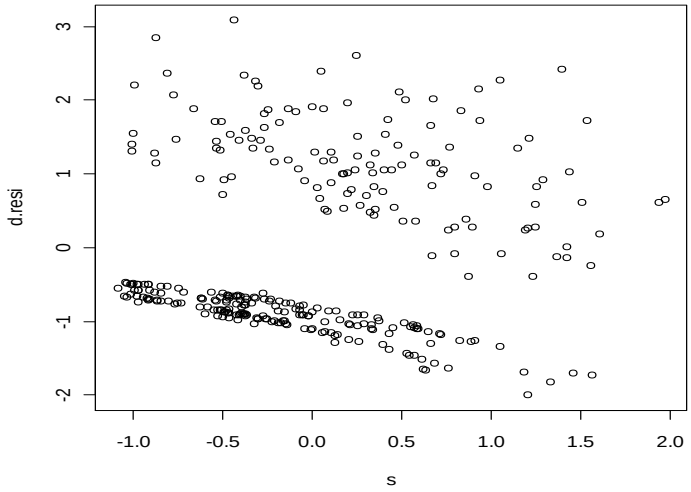


Figure A5 Deviance plot for Model 6**

A6 Hazard Ratios

Comparisons	Ratios
Thal vs. Dex	1.494811333
Cyto: Presence vs. Absence	2.181472265
LDH: Presence vs. Absence	1.002904209
ISS2 vs. ISS1	1.377127764
ISS3 vs. ISS1	2.16842262
ISS3 vs. ISS2	1.574597997
Agecat2 vs. Agecat1	1.508325357
Agecat3 vs. Agecat1	1.655329363
Agecat3 vs. Agecat2	1.097461735
Unhealthy vs. Healthy[6]	7.853034362
With LDH & Cyto vs. Without	2.217321642

Figure A6 Hazard Ratios for various comparisons

[6] Healthy is only relative, and refers to a patient without LDH presence or Cytogenic abnormalities and who is categorized in Agecat 1 and ISS1. Comparison is done holding treatment constant.