Covariates Implicating Time to Chronic Graft-Versus-Host Disease in Allogeneic Bone Marrow Transplant Recipients with Leukemia in the Clinical Setting

By Azuka Atum

STAT 697 Spring 2023

I. Abstract

Graft-versus-host disease (GvHD) is a multi-systemic disorder that occurs when a transplanted organ's immune cells recognize the host as foreign and attack the recipient's body cells. CGvHD can occur following: 1) allogeneic bone transplantation, 2) following transplantation of solid organs that are lymphoid rich, and 3) after transfusion of un-irradiated blood. Generally, it is diagnosed after symptoms persist or appear 100 days after transplantation. The objective of this analysis is exploration of the factors which impact the time to chronic graft-versus-host disease (CGVHD) in leukemia patients that received an allogeneic bone marrow transplant as it relates to the following clinical characteristics across clinical sites: age, sex, cytomegalovirus immune status (CMV), waiting time to transplant, and disease group. Analysis was conducted using Kaplan-Meier method, generalized gamma Accelerated Time Failure (AFT) model using SAS OnDemand for Academics: SAS Studio. It was found that patients at St. Vincent had a 35% higher time in days to CGVHD compared to the reference level. Conversely, patients at Alfred had a 10.3% decreased time to CGVHD in days compared to the reference level. For patients at OSU, the relative risk of developing CGVHD decreased by 20.4% compared to the reference level. Additionally, leukemia patients who were cytomegalovirus positive had an 11% decrease in time to CGVHD compared to those who were not. Male leukemia patients had a 14% increase in their time to CGVHD compared to female patients across clinical sites. Every one year increase in patient age led to a 0.3% increase in the time to CGVHD. Disease groups also had an impact on the time to CGVHD, with an increase of 15.2% across groups. All covariates were statistically insignificant, suggesting that they may not play a strong role in time to development of the disease.

II. Introduction

II.1 | Chronic Graft-Versus-Host-Disease (CGVHD)

Graft-versus-host disease (CGvHD) is an often fatal, multisystemic disease that occurs when a transplanted organ's immune cells recognize the host as foreign and attack the recipient's body cells. CGvHD can occur following: 1) allogeneic bone transplantation, 2) following transplantation of solid organs that are lymphoid rich, and 3) after transfusion of un-irradiated blood [1]. Generally, it is diagnosed after symptoms persist or appear 100 days after transplantation [2] . In the case of allogeneic bone transplantation, potential adverse reaction is an important marker of survival, mainly because bone marrow transplants are a standard treatment for certain types of blood disorders, such as leukemia [3]. Furthermore, prognosis for recovery can depend on particular risk factors at the time of transplant, such as patient age, patient sex, stage of initial disease, hospital [3]. And, a patient's survival may be impacted post-transplant, as observed when modeling time to development of the CGVHD [4].

II.2 | Clinical Trial

A multicenter trial of patients prepared for transplantation with a radiation-free conditioning regimen. The preparative regimen used in the study of allogeneic marrow transplants for patients with acute myelocytic leukemia (AML) and acute lymphoblastic leukemia (ALL) was a combination of 16 mg/kg of oral BUsulfan (BU) and 120 mg/kg of intravenous cyclophosphamide (Cy). In Copelan et. al (1991) [5], 137 patients (99AML, 38 ALL) were treated at one of four clinical sites: 76 at The Ohio State University Hospitals (OSU); 21 at Hahnemann in Philadelphia (HU); 23 at St. Vincent's Hospital (SVH) in Sydney Australia; and 17 at Alfred Hospital (AH) in Melbourne [6]. Forty-two (42) patients relapsed, forty-one (41) died while in remission, while twenty-six (26) patients had an episode of acute GVHD [7].

II.3 | Data description

According to the original paper by Copelan et. al (1991), the data was collected at four clinical sites from 137 patients [8]. The endpoint of interest for this paper concerned efficacy of busulfan and cyclophosphamide. Data was collected on several factors as observed in figure 7 in the appendix.

II. 4 | Research question

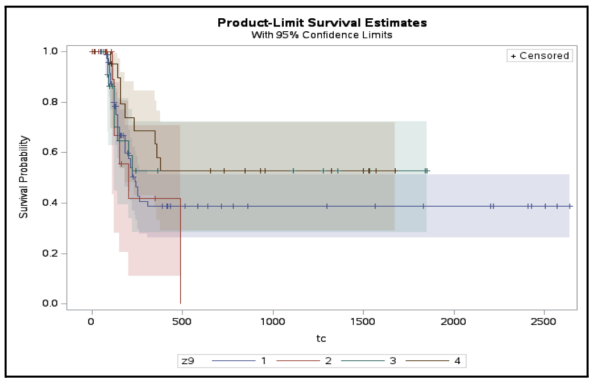
Exploration of the factors which impact the time to chronic graft-versus-host disease (CGVHD) in leukemia patients that received an allogeneic bone marrow transplant as it relates to the following clinical characteristics across clinical sites: age, sex, cytomegalovirus immune status (CMV), waiting time to transplant, and disease group.

III. Statistical Analysis

III.a | Kaplan Meier Estimate

A Kaplan-Meier estimate was performed to assess the survival estimates of each hospital site. Clinical sites are numbered from 1 to 4: 1 = OSU, 2 = AH, 3 = SVH, 4 = HU. The survival curves of each site overlaps considerably, especially in the case of SVH and HU. Additionally, There is crossover observed between OSU and AH sites. From the plot, we can see that the assumption of proportional hazards is violated entirely.

This is further verified by graphically checking the proportional hazards assumption where each of the curves are not in parallel and cross at certain points. Finally, assessing the equality over strata output from the proc lifetest procedure is consistent with present findings, particularly when assessing Log-Rank and Wilcoxson tests at the 5% level. Each had statistically non-significant tests, with p = 0.4397 and p= 0.4011, respectively. A failure of the proportional hazard property suggests that the direction of the effect of clinical site on the time to CGvHD hazard rate changes over time.



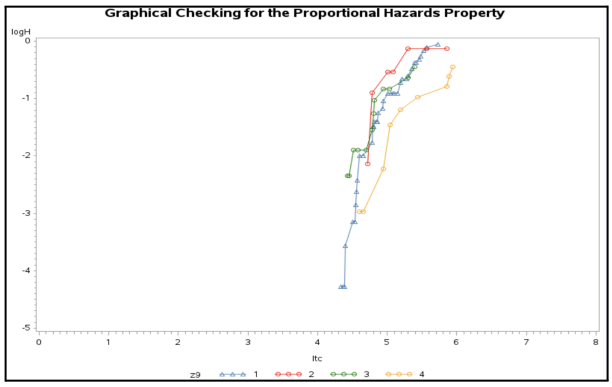


Figure 1 and 2 (left to right): 1) K-M survival curves across four clinical sites, 2) checking PH assumption.

A hypothesis test was performed to assess equality over the strata of different clinical sites.

$$H_0: T_{OSU}(t) - T_{AH}(t) - T_{SVH}(t) - T_{HU}(t) = 0$$

$$H_A: T_{OSU}(t) - T_{AH}(t) - T_{SVH}(t) - T_{HU}(t) \neq 0$$

Where, t= time to CGVHD in days, and

Time to CGVHD at clinical site	**Description**
$T_{OSU}(t)$	is the survival function (time to CGVHD) for The Ohio State University.
$T_{SVH}(t)$	is the survival function (time to CGVHD) for St. Vincent's Hospital (SVH).
$T_{HU}(t)$	is the survival function (time to CGVHD) for Hahnemann University (HU).

$T_{AH}(t)$	is the survival function (time to CGVHD) for Alfred Hospital (AH).

Table 1: List of variables for the hypothesis test.

As you can see, at the 5% level, there is no significant difference in the time to CGVHD between each clinical site. Therefore, we fail to reject the null hypothesis in favor of the alternative.

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	2.7030	3	0.4397
Wilcoxon	2.9391	3	0.4011
Tarone	2.7953	3	0.4243
Peto	2.9125	3	0.4053
Modified Peto	2.9167	3	0.4046
Fleming(1)	2.9329	3	0.4021

Figure 3: More tests for the time to CGVHD across clinical sites. All are statistically insignificant, suggesting there is no difference in time to CGVHD at each clinical site.

III.b | Model Selection

All covariates underwent backwards elimination using Type 3 Analysis of Effects. Insignificant covariates were removed using Wald tests at the 5% level. Since the Kaplan Meier method showed a resolute violation in the proportional hazards assumption, four Accelerated Time Failure (AFT) models were considered: Generalized Gamma, Log-normal, Log-logistic, and Weibull. Assessing Goodness-of-fit across all AFT models using AIC, it was determined that the Gamma model fit the data best.

Figure 4 & Table 2 (left to right): Wald-test screening using Type 3 analysis of effects. AIC of all AFT models is compared.

Type III Analysis of Effects			
Effect	DF	Wald Chi-Square	Pr > ChiSq
g	1	1.1093	0.2922
t1	1	1.5435	0.2141
t2	1	0.8734	0.3500
d1	1	0.2272	0.6336
d2	1	0.0112	0.9157
d3	1	0.2363	0.6269
ta	1	0.0144	0.9043
a	1	0.6774	0.4105
tp	1	6.1798	0.0129
p	1	0.9364	0.3332
z1	1	1.3619	0.2432
z2	1	0.9445	0.3311
z3	1	0.0407	0.8401
z4	1	2.7817	0.0953
z5	1	0.6260	0.4288
z6	1	1.3228	0.2501
z7	1	0.0250	0.8744
z8	1	0.0613	0.8045
z10	1	0.2014	0.6536

Model	AIC
Generalized Gamma	289.398
Lognormal	323.673
Log-logistic	328.875
Weibull	338.022

A likelihood ratio test is performed to compare the reduced gamma model to the original.

	Model Fit Statistics	
Hypothesis	**Criterion**	**Statistic**
H_0: (removed covariates) , df = 11	$-2\,LOG\,L_0$ **(reduced)**	244.148

H_1: (original model), df = 24	$-2\,LOG\,L_1$ **(full)**	241.398
Df1-dfo = 13	$\Lambda = 2.75$	

Table 3: LR test for full gamma model versus reduced model.

The χ^2 test statistic has degrees of freedom df = 13, and at the $\alpha = 0.05$ significance level, a critical value of 22.362. Because the ratio is less than the test statistic, we fail to reject the null hypothesis, which states that the reduced model is adequate enough to fit the data.

III.c | Assumption Check

The assumption of a generalized gamma model for the covariates can be assessed using the probability plot. We can see that the points are more or less following a straight line on the plot, indicating an okay fit for the data.

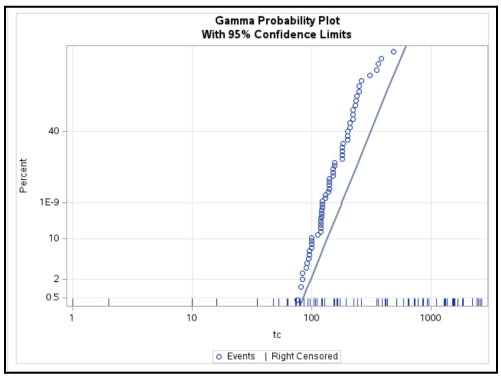


Figure 6: Reduced generalized gamma AFT probability plot model for time to CGVHD.

III.d | Final Model and Interpretation of covariates

The final final fitted model is given by the following from :

$$Log(Y) = 0.1415 * disease\_group + 0.0034 * Age + 0.1295 * patient\_sex - 0.1182 * cmv\_status - 0.2294 * \{clinical\ site = OSU\} - 0.1089 * \{clinical\ site = AH\} - 0.2982 * \{clinical\ site = SVH\} + 0.00 * \{clinical\ site = HA\} + 0.6991 * W + 4.7708$$

Where,

- Disease group is a nominal variable that takes on three values: 1 = ALL, 2 = AML (low risk), 3 = AML (high risk).
- Patient Age is a continuous variable in years.
- Patient sex is a nominal variable that takes on two values: 1=male, 0=female.
- Patient Cytomegalovirus status (CMV) is a nominal variable that takes on two values: 1 = CMV positive, 0 = CMV Negative.
- Clinical site is a nominal variable taking on four values: 1=OSU, 2=AH, 3=SVH, 4=HA.

- W, a scale parameter, is a random variable that follows the log-gamma distribution with W ~ log-gamma(1,1,1) [9]

Based on the final model, after adjusting for disease group, patient age, patient sex, and CMV status, the impact of clinical site on the time to chronic graft-versus-host disease (CGVHD) was examined. In this analysis, Hahnemann was set as the reference level for clinical site. It was found that patients at St. Vincent had a 35% higher time to CGVHD compared to the reference level. Conversely, patients at Alfred had a 10.3% lower time to CGVHD compared to the reference level. For patients at OSU, the relative risk of developing CGVHD decreased by 20.4% compared to the reference level.

Additionally, leukemia patients who were cytomegalovirus positive had an 11% decrease in time to CGVHD compared to those who were not. Male leukemia patients had a 14% increase in their time to CGVHD compared to female patients across clinical sites. Every one year increase in patient age led to a 0.3% increase in the time to CGVHD. Disease groups also had an impact on the time to CGVHD, with an increase of 15.2% across groups. All covariates were statistically insignificant, suggesting that they may not play a strong role in time to development of the disease.

IV. Conclusion and Discussion

From the final model above, each covariate has an effect on the hazard risk for developing CGVHD. After adjusting for disease group, patient age, patient sex, and CMV status, the impact of clinical site on the time to chronic graft-versus-host disease (CGVHD) was examined. In this analysis, Hahnemann was set as the reference level for clinical site. It was found that patients at St. Vincent had a 35% higher time in days to CGVHD compared to the reference level. Conversely, patients at Alfred had a 10.3% lower time to CGVHD compared to the reference level. For patients at OSU, the relative risk of developing CGVHD decreased by 20.4% compared to the reference level.

If you were a patient who was admitted at St Vincent in Australia, who was female, CMV positive, you generally would see an increase of your time in days to CGVHD compared to a male patient from the same clinical site. However, as previously stated, these covariates are not significant at the 5% level (see Figure 8 in appendix). There are a few reasons why many of the covariates were insignificant, most likely owing to the fact that a plurality of covariates were removed via Wald test. Also, it is possible that AFT models are not the best fit for the data. Additionally, over 55% per cent of cases in this study were censored.

Appendix

References:

1. Chronic graft versus host disease. Chronic Graft Versus Host Disease - an overview | ScienceDirect Topics.(n.d.). https://www.sciencedirect.com/topics/medicine-and-dentistry/chronic-graft-versus-host-disease

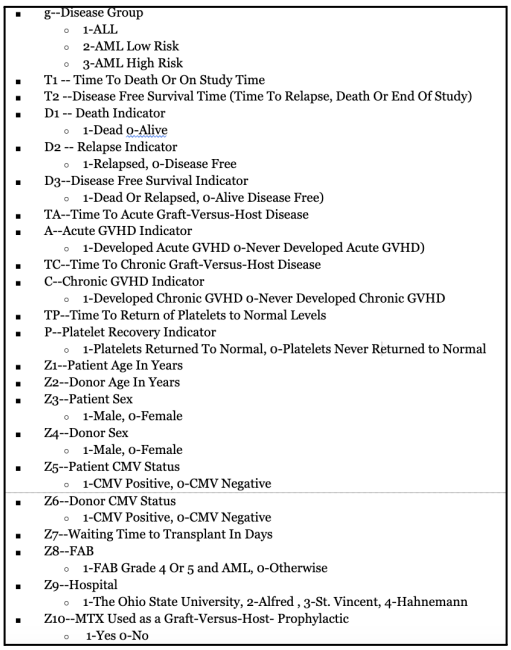
2. Vaillant, A., Modi, P., & Mohammadi, O. (2022, October 10). *Graft Versus Host Disease*. National Library of Medicine. https://ncbi.nlm.nih.gov/books/NBK538235/#:~:text=Graft-versus-host%20disease%20(GvHD)%20is%20a%20systemic,the%20tissues%20of%20the%20recipient.

3-6, 7. Klein, J. P. (2006). Section 1.3 Bone Marrow Transplant in Leukemia. In M. L. Moeschberger (Ed.), *Techniques for Censored and Truncated Data* (2nd ed., pp. 1–538). story, Springer New York, NY.

8. Copelan, E. A., Biggs, J. C., Thompson, J. M., & Crilley, P. (2020, December 21). *Treatment for acute myelocytic leukemia with allogeneic bone marrow transplantation following preparation with Bucy2*. Blood. https://www.sciencedirect.com/science/article/pii/S0006497120852574?via%3Dihub#cesec60

9. Aida, H., Hayashi, K., Takeuchi, A., Sugiyama, D., & Okamura, T. (2022, July 25). *An accelerated failure time cure model with shifted gamma frailty and its application to epidemiological research*. Healthcare (Basel, Switzerland). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9332026/

Figures:



- g--Disease Group
  - 1-ALL
  - 2-AML Low Risk
  - 3-AML High Risk
- T1 -- Time To Death Or On Study Time
- T2 --Disease Free Survival Time (Time To Relapse, Death Or End Of Study)
- D1 -- Death Indicator
  - 1-Dead 0-Alive
- D2 -- Relapse Indicator
  - 1-Relapsed, 0-Disease Free
- D3--Disease Free Survival Indicator
  - 1-Dead Or Relapsed, 0-Alive Disease Free)
- TA--Time To Acute Graft-Versus-Host Disease
- A--Acute GVHD Indicator
  - 1-Developed Acute GVHD 0-Never Developed Acute GVHD)
- TC--Time To Chronic Graft-Versus-Host Disease
- C--Chronic GVHD Indicator
  - 1-Developed Chronic GVHD 0-Never Developed Chronic GVHD
- TP--Time To Return of Platelets to Normal Levels
- P--Platelet Recovery Indicator
  - 1-Platelets Returned To Normal, 0-Platelets Never Returned to Normal
- Z1--Patient Age In Years
- Z2--Donor Age In Years
- Z3--Patient Sex
  - 1-Male, 0-Female
- Z4--Donor Sex
  - 1-Male, 0-Female
- Z5--Patient CMV Status
  - 1-CMV Positive, 0-CMV Negative
- Z6--Donor CMV Status
  - 1-CMV Positive, 0-CMV Negative
- Z7--Waiting Time to Transplant In Days
- Z8--FAB
  - 1-FAB Grade 4 Or 5 and AML, 0-Otherwise
- Z9--Hospital
  - 1-The Ohio State University, 2-Alfred , 3-St. Vincent, 4-Hahnemann
- Z10--MTX Used as a Graft-Versus-Host- Prophylactic
  - 1-Yes 0-No

Figure 7: Description of variables.

Analysis of Maximum Likelihood Parameter Estimates

Parameter		DF	Estimate	Standard Error	95% Confidence Limits		Chi-Square	Pr > ChiSq
Intercept		1	4.7708	0.3779	4.0301	5.5116	159.34	<.0001
g		1	0.1415	0.1051	-0.0645	0.3475	1.81	0.1781
z1		1	0.0034	0.0099	-0.0159	0.0227	0.12	0.7293
z3		1	0.1295	0.1607	-0.1855	0.4445	0.65	0.4203
z5		1	-0.1182	0.1725	-0.4562	0.2198	0.47	0.4930
z7		1	0.0000	0.0002	-0.0004	0.0004	0.01	0.9165
z9	1	1	-0.2294	0.1985	-0.6184	0.1596	1.34	0.2477
z9	2	1	-0.1089	0.3145	-0.7253	0.5075	0.12	0.7291
z9	3	1	-0.2982	0.2639	-0.8153	0.2190	1.28	0.2584
z9	4	0	0.0000	.	.	.	.	.
Scale		1	0.6991	0.1175	0.5029	0.9718		
Shape		1	-2.6704	0.4414	-3.5356	-1.8053		

Figure 8: Maximum Likelihood Parameter Estimates of the final model.

SAS Code

```
/*klein data from section 1.3 for project, stat 697*/
/*load data*/
data bmtp;
      input g t1 t2 d1 d2 d3 ta a tc c tp p z1 z2 z3 z4 z5 z6 z7 z8 z9 z10 @@;
      datalines;
            1 2081 2081 0 0 0 67 1 121 1 13 1 26 33 1 0 1 1 98 0 1 0
                        .
                        .
                        .
                         ;

      proc print data= bmtp; run;

      /*K-M estimation*/
                  proc  lifetest  data=bmtp  method  =  km  plots=(survival(cl),ls,lls)
graphics outsurv=a;
                        time tc*c(0);
                        strata z9 /test=all ;
                        symbol1 v=none color=black line=1;
                        symbol2 v=none color=black line=2;
                        run; *equality over strata;


                    proc  lifetest  data=bmtp2  method  =  km  plots=(survival(cl),ls,lls)
graphics outsurv=a;
                        time tc*c(0);
                        strata z9 ;
                        symbol1 v=none color=black line=1;
                        symbol2 v=none color=black line=2;
                        test g z1 z3 z5 z7 ;
                        run;
```

```
**graphical checking for dist of y;

        data a2;
            set a;
            s=survival;
            logH=log(-log(s));
            lnorm=probit(1-s);
            logit=log(s/(1-s));
            ltc=log(tc);
            run;

    proc print data=a2;
    run;

    title "Graphical Checking for the Proportional Hazards Property";
    proc gplot data=a2;
        plot logit*ltc=z9 logH*ltc=z9 lnorm*ltc=z9;
        symbol1 i=join width=0.5 value=triangle c=steelblue;
        symbol2 i=join width=0.5 value=circle c=red;
        symbol3 i=join width=0.5 value=circle c=green;
        symbol4 i=join width=0.5 value=circle c=orange;
    run;



/*proportional hazards assumption*/

        proc gplot data=a2;
        title "PH Assumption";
        plot logH*tc=z9;
```

```
            run;
                  *fails since they x-over;


      /*aft modeling */

            *gamma;
            proc lifereg data=bmtp;
                  class z9;
                  model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3 z4 z5 z6 z7 z8 z9
z10
                  /dist = gamma;
                  run; *     aic 289.398;

            *log-normal;
            proc lifereg data=bmtp;
                  class z9;
                  model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3 z4 z5 z6 z7 z8 z9
z10
                  /dist = lnormal;
                  run; *aic 323.673;
            /*
            proc lifereg data=bmtp2;
                  class z9;
                  model tc*c(0) = g     z1    z3    z5    z7 / dist = lnormal;
                  run; */

            *log-logistic;
            proc lifereg data=bmtp;
                  class z9;
                  model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3 z4 z5 z6 z7 z8 z9
z10
```

```
      /dist = llogistic;
      run; *aic 326.986;

/*
proc lifereg data=bmtp2;
      class z9;
      model tc*c(0) = g     z1    z3    z5    z7 / dist = llogistic;
      run; */

*weibull;
proc lifereg data=bmtp;
      class z9;
      model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3 z4 z5 z6 z7 z8 z9
z10
      /dist = weibull;
      run; *aic 337.711;

      /*
      proc lifereg data=bmtp2;
      class z9;
      model tc*c(0) =  g z1 z3 z5 z7 / dist = weibull;
      run; */


/* goodness of fit */


            proc lifereg data=bmtp;
                  class z9;
                  model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3 z4 z5
z6 z7 z8 z9 z10
                  / dist = lnormal; probplot;
```

```
run;


/*
proc lifereg data=bmtp2;
class z9;
model tc*c(0) = g z1 z3 z5 z7
/ dist = lnormal; probplot;
run; */


proc lifereg data=bmtp;
      class z9;
      model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3
z4 z5 z6 z7 z8 z9 z10
      / dist = llogistic; probplot;
run;


/*
proc lifereg data=bmtp2;
class z9;
model tc*c(0) = g z1 z3 z5 z7
/ dist = llogistic; probplot;
run;  */


proc lifereg data=bmtp ;
      class z9;
      model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3
z4 z5 z6 z7 z8 z9 z10
      / dist = weibull; probplot;
```

```
                        run;

                        /*
                        proc lifereg data=bmtp2;
                        class z9;
                        model tc*c(0) = g z1 z3 z5 z7
                        / dist = weibull covb ;
                        output out =a2 cdf=f xbeta=xb p=median STD=se; probplot;
                        run;
                        */


                        proc lifereg data=bmtp ;
                              class z9;
                              model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3
z4 z5 z6 z7 z8 z10
                              / dist = gamma  ; probplot;
                        run;

      /*backwards elimination*/
            proc lifereg data=bmtp ;
                              class z9;
                              model tc*c(0) = g t1 t2 d1 d2 d3 ta a tp p z1 z2 z3
z4 z5 z6 z7 z8 z10
                              / dist = gamma  ; probplot;
                        run;

            proc lifereg data=bmtp ;
                        class z9;
                        model tc*c(0) = g t2 d1 d2 d3 ta a tp p z1 z2 z3 z4 z5 z6 z7 z8
z10
                        / dist = gamma  ; probplot;
```

```
run;


*final model;
proc lifereg data=bmtp ;
      class z9 ;
      model tc*c(0) = g z1 z3 z5 z7 z9
      / dist = gamma  ;
run;
```