

The Effect of Acute and Chronic Graft-Versus-Host Disease and Platelet Recovery on Time to Relapse and Time to Death in Leukemia Patients after Bone Marrow Transplantation

BSTA 697 Final Report (Spring 2020)

Ujjwala Myla

**Abstract**

In this study, we model the time to death and the time to relapse for bone marrow transplant patients in the data set from Copelan et al., 1991. We first examine the type-specific hazards model for competing risks and find that covariate coefficients are not equivalent across risk types. We then build separate models for time to death and time to relapse and account for time-dependence of the indicators for acute and chronic graft-versus-host-disease (GVHD) as well as platelet recovery. We find that these indicators are not significant for modeling time to relapse but are significant for time to death. We find that the leukemia disease type (g) and the indicator for French-American-British (FAB) Grade 4 or 5 acute myeloid leukemia (AML) were the most significant covariates for time to relapse. On the other hand, the final model for time to death contains the time-dependent indicator for platelet recovery, donor age (Z2), Z8, and hospital site (Z9).

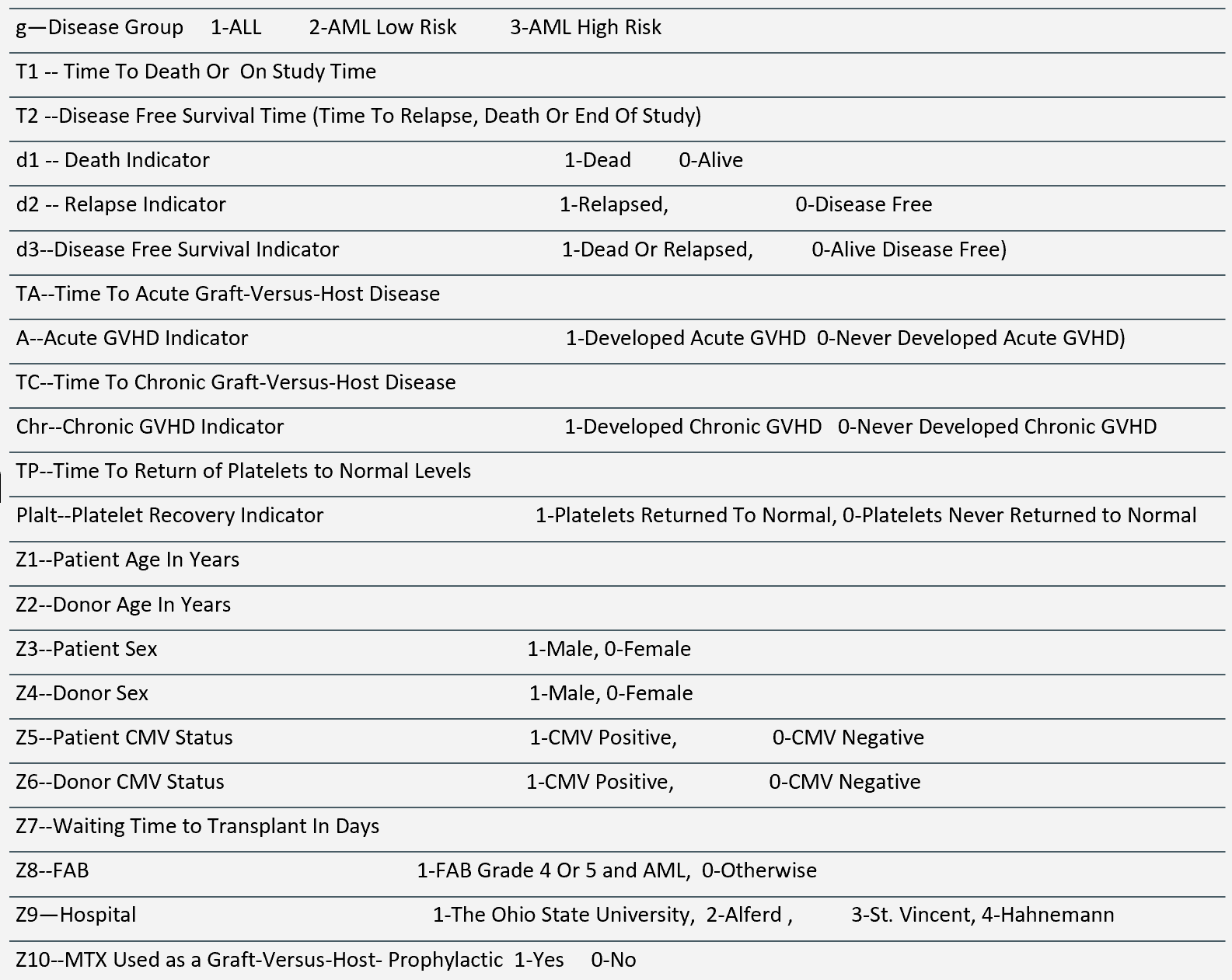
**Introduction**

Bone marrow transplantation is a standard treatment for acute leukemia. The prognosis of the transplant recipient depends on factors known at the time of transplantation and on events that may occur at random times after transplantation.The former are fixed-time and the latter are time-dependent variables. An example of the latter is graft-versus-host-disease (GVHD), a post-transplantation complication which can diminish quality of life and may affect the overall survival of the transplant patient. GVHD can be acute or chronic, and a patient may develop both. Furthermore, GVHD may be antileukemic and reduce the risk of relapse. Another post-transplantation event is the return of platelet count to normal levels. Platelet recovery is associated with a favorable prognosis.

The data consists of 137 leukemia patients who received bone marrow transplants between Mar 1, 1984 and June 30, 1989. The study was conducted at four different hospitals, two in the U.S. and two in Australia. Here, relapse of leukemia or death were defined as failures, and the survival times of patients for these competing risks were coded as time to death (T1) and disease free survival time (T2; time to the earlier of either relapse or death). The maximum follow up time was 7 years. Including T1 and T2, there were 22 variables (Table 1).

Several covariates warrant elaboration. The variable for type of leukemia, g, contains three levels: acute lymphoid leukemia (ALL), acute myeloid leukemia (AML) low risk first remission, and AML high risk second remission or untreated first relapse. Of the three types of leukemia, ALL has the worst prognosis and AML low risk has the best. Secondly, Z8 is an indicator of the French-American-British (FAB) classification of AML tumors as Grade 4 or 5. The presence of FAB Grade 4 or 5 AML may increase risk for death or relapse (Klein and Moeschberger 3-6). Finally, CMV status variables (Z5 and Z6) refer to patient and donor cytomegalovirus immune status.

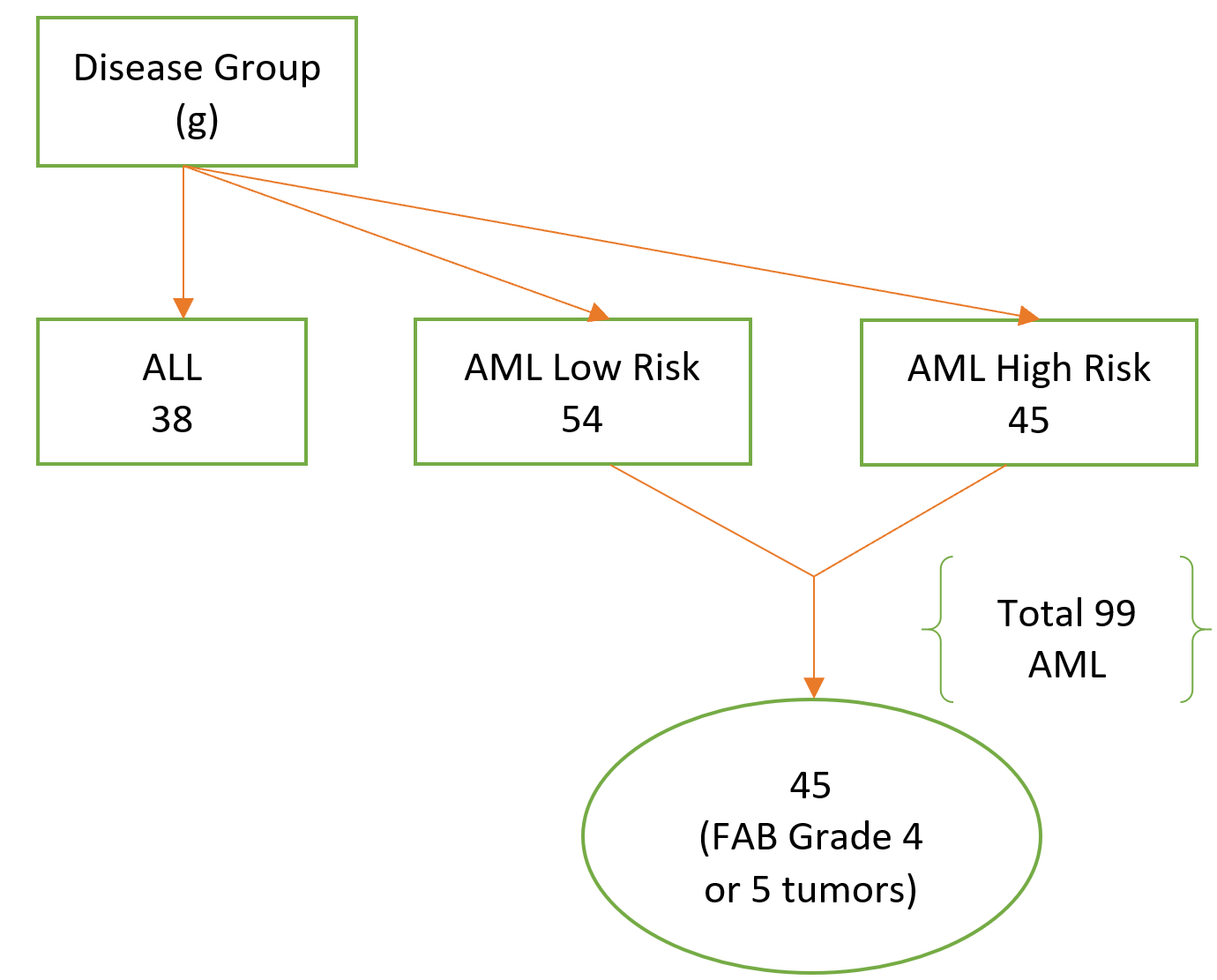
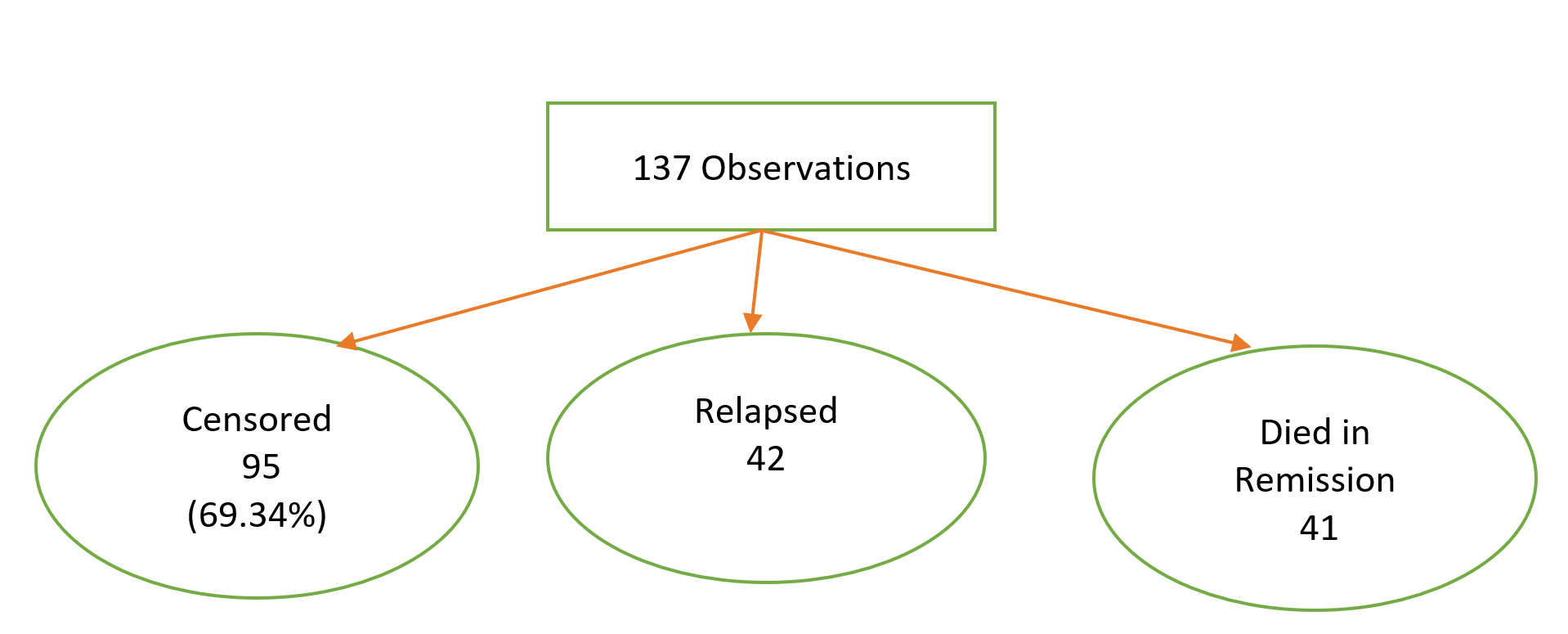
In this study, we examine the competing risk events (time to relapse, time to death, and disease free survival time) and find that while the proportional hazard property holds when we compare time to relapse and time to death to the time to failure due to either event, the coefficient of at least one covariate differs between the event types. Hence, we model each event separately. We focus on time to death and time to relapse and examine the effect of the time-dependent indicators for acute GVHD (A), chronic GVHD (Chr), and platelet recovery (Plalt) on these competing risks. We find that acute and chronic GVHD are not significant covariates for either time to death or time to relapse but that platelet recovery is significant for modeling risk of death.

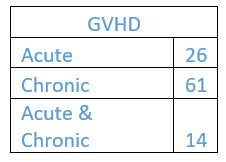


**Table 1:** Definition of variables

**Data Exploration and Analysis**

Overall, of the 137 observations, 95 (69.34%) were censored; 42 patients relapsed, and 41 died in remission. There were 38 ALL, 54 AML low-risk, and 45 AML high risk patients. Of the 99 AML patients, 45 had FAB Grade 4 or 5 tumors. Twenty-six patients had acute, 61 had chronic, and 14 had both acute and chronic GVHD. (**Fig 1**). Seventeen patients either relapsed or died in remission without platelets returning to normal levels.

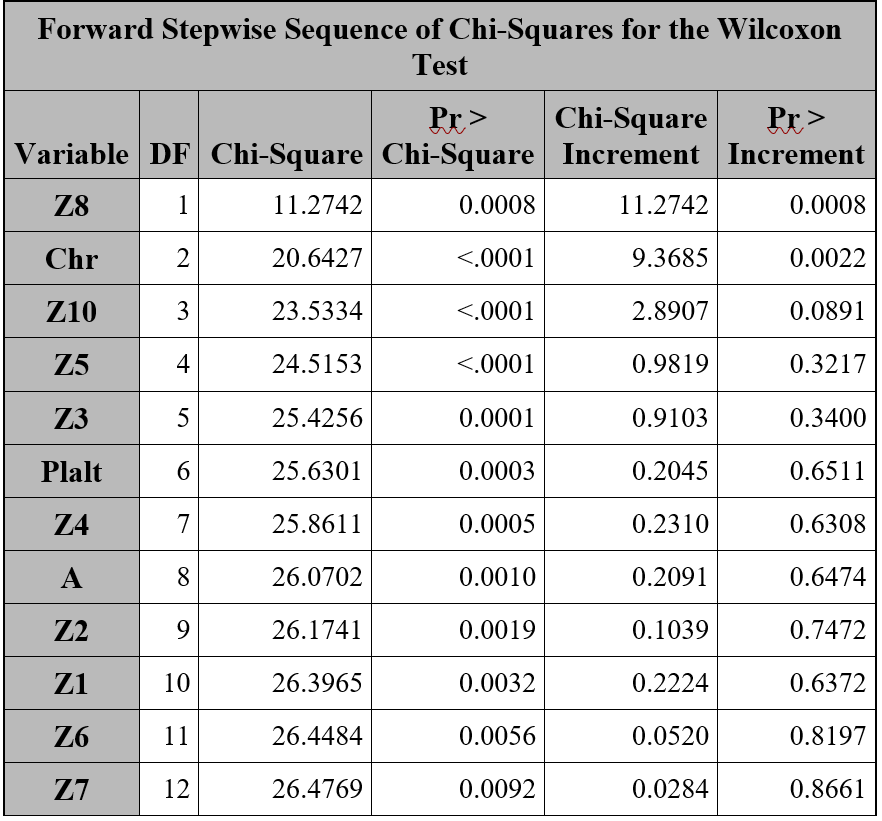
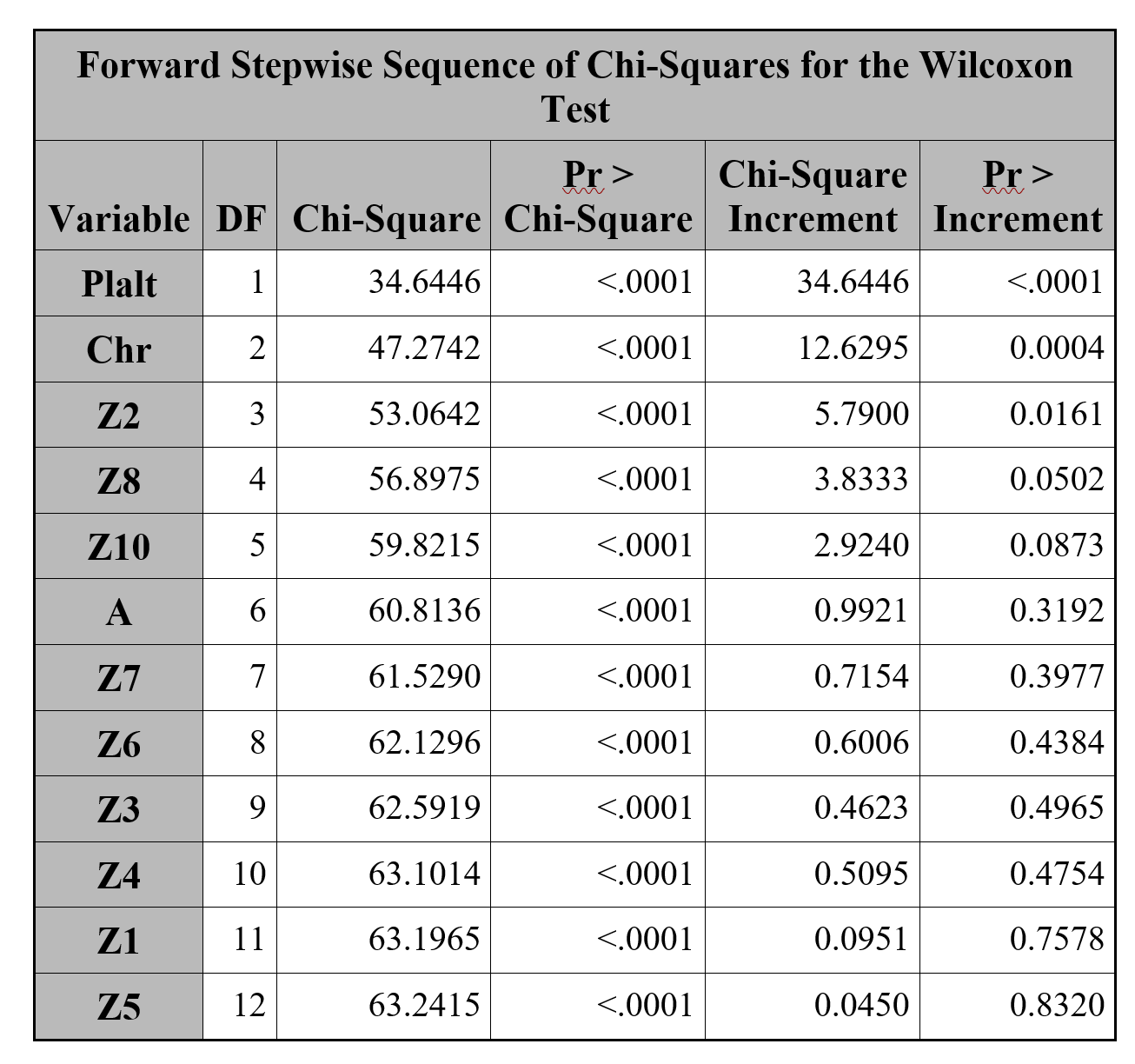




**Figure 1:** Overall Observation(top left),Disease group statistics(top right),# of patients with GVHD types (bottom)

Nonparametric Tests

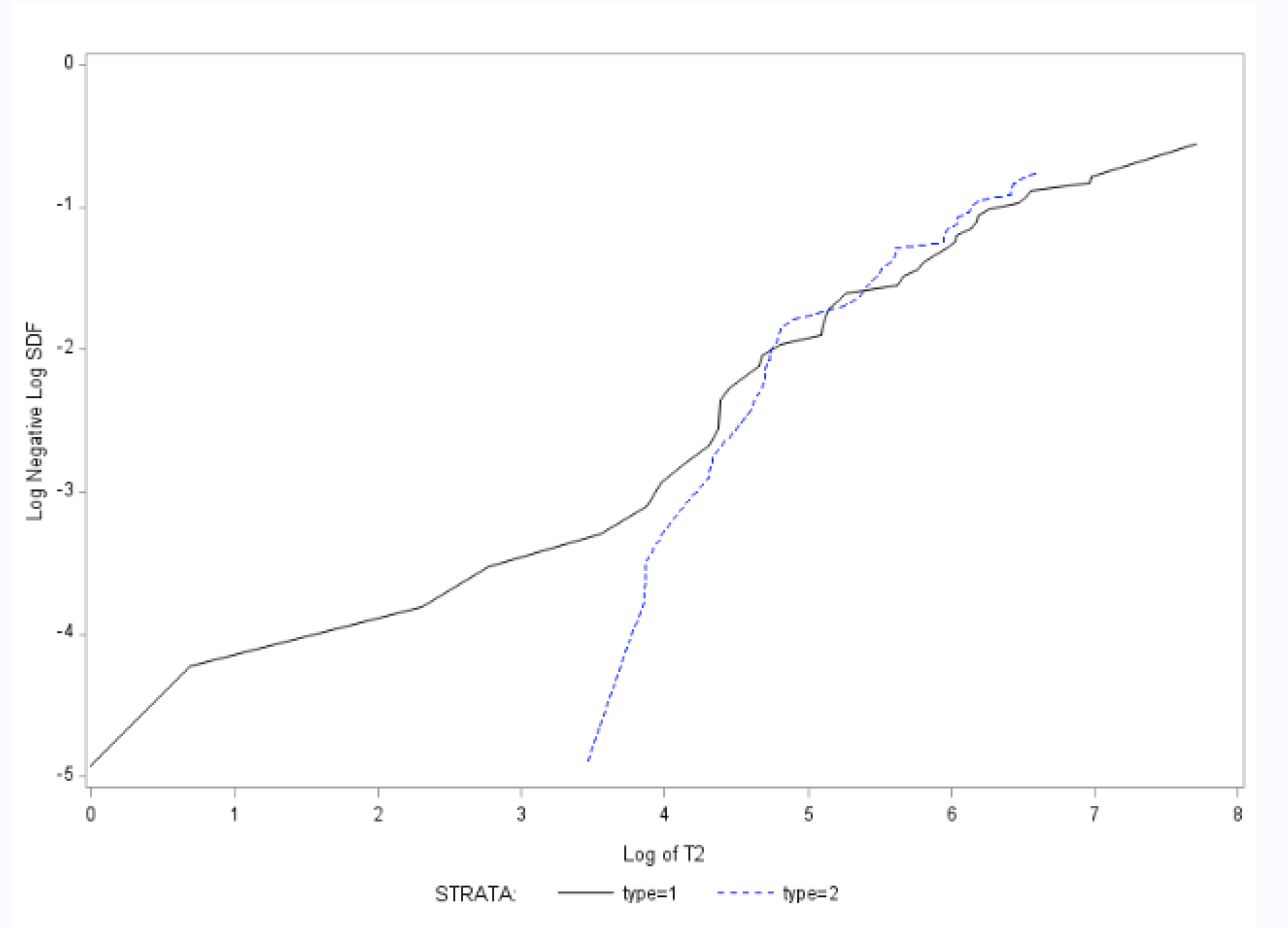
We first explored the data using non-parametric methods. Using the Wilcoxon test, we assessed the impact of covariates on Kaplan-Meier survival functions. For time to death, there were significant differences among levels when stratifying on platelet recovery (Plalt), chronic GVHD (Chr), disease group (g), and hospital site (Z9). Furthermore, survival times varied significantly with donor age (Z2). For time to relapse, stratifying on chronic GVHD (Chr), FAB Grade 4 or 5 AML (Z8), and disease group (g)(**Fig2**).

****

**Figure 2:** Wilcoxon test results using Forward Stepwise for time to death (T1) (left), and time to relapse (T2\*D2) (Right)

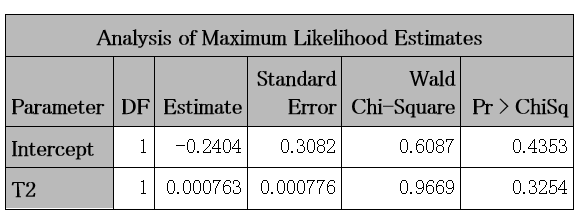
Competing Risks:

In order to study competing risk events, we modeled the cause-specific hazard functions. We first examine whether the hazard functions without covariates for the two events are proportional using the log T2 vs. log H plot, where T2 is the time until the earlier of relapse or death and H is the cumulative hazard. In the plot, Type 1 is the time to death and Type 2 is the time to relapse (**Fig 3**). The curves are similar, suggesting the proportional hazard property may be satisfied.

****

**Figure 3:** Log cumulative Hazard plot of T2 with strata on type variable

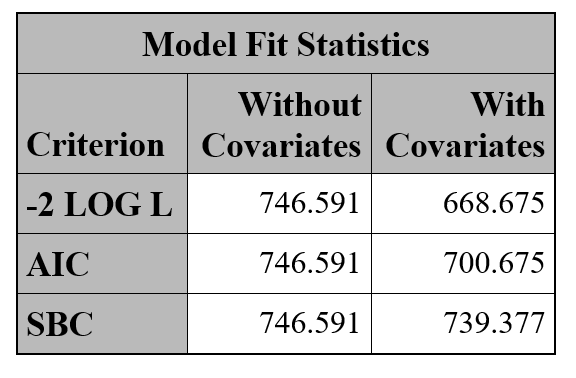
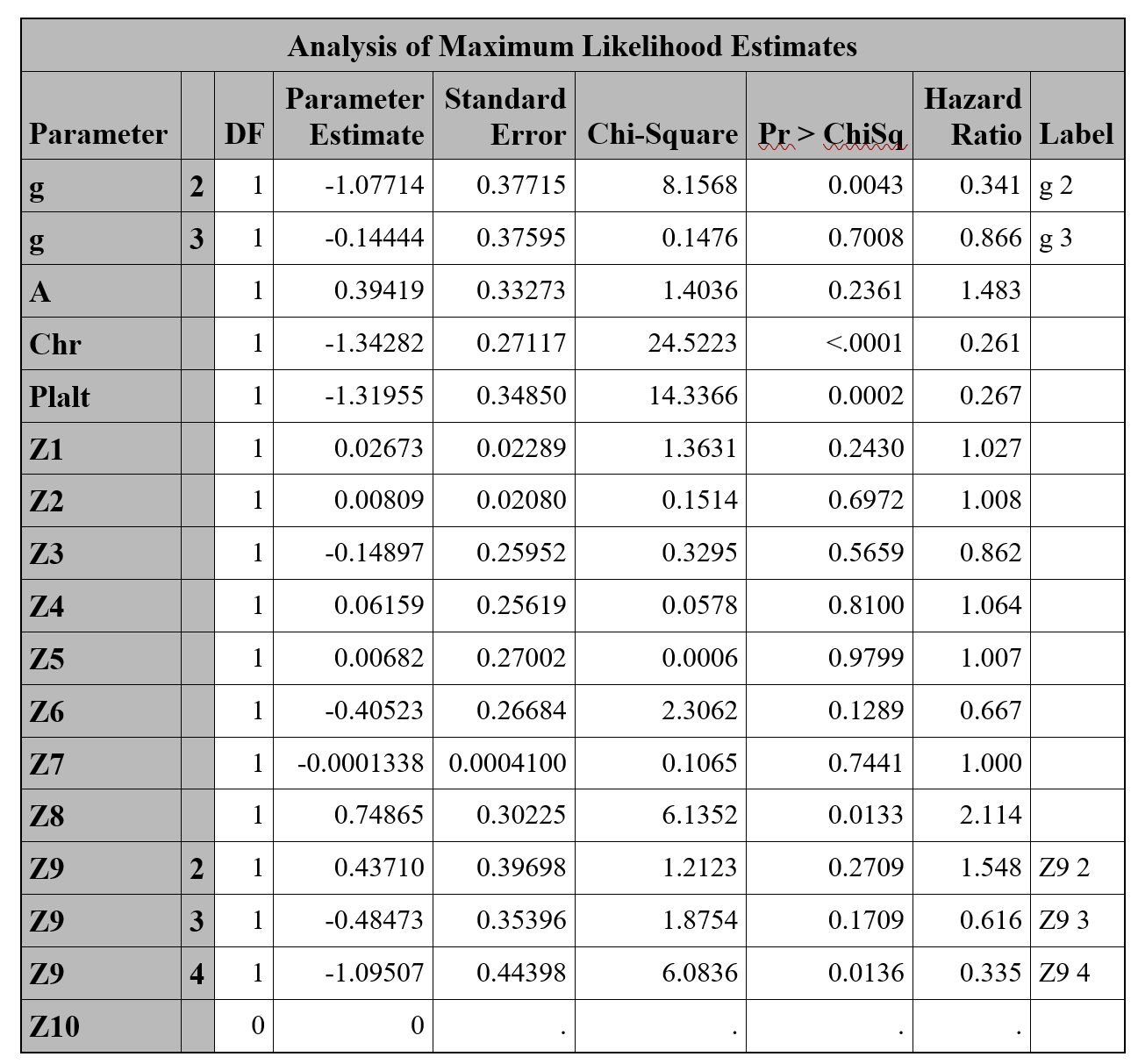
To test the proportional hazard property for the baseline hazard functions formally, we created the combined event indicator “d\_all,” which takes the value 0 for censored observations, 1 for deaths, and 2 for relapses to capture all possible event types. The logistic regression of d\_all on T2 yields a non-significant p-value (0.9669) for T2 **(Fig 4)**. Thus, the proportional hazard property is satisfied.



**Figure 4:** Maximum Likelihood Estimates of Competing risk model

We next examined the effect of covariates on the survival of the two types. For the particular cause of interest cause-specific models were fit as the regular cox models. For each risk event of interest, the competing risk was treated as censored observations. For example, when modeling time to death, relapses were treated as censored observations and vice versa. When modeling disease free survival, both death and relapses were treated as uncensored observations.

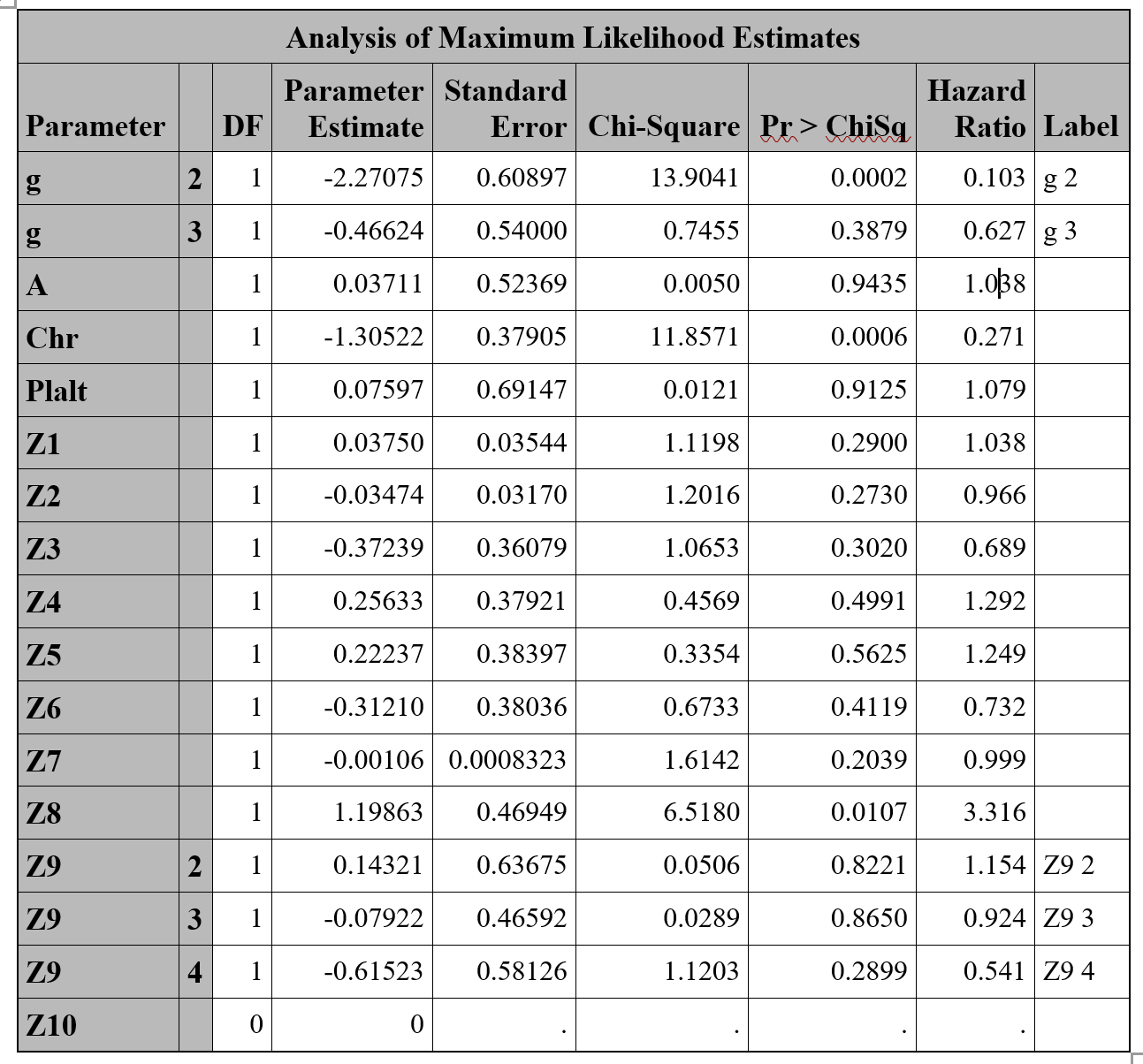
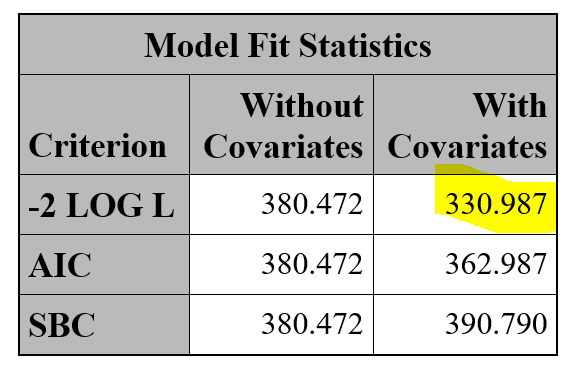
**Survival analysis for disease free survival**

Both death and relapses are considered uncensored. The time variable used for modeling is T2.(**Fig 5)**

**Figure 5:** Fit statistics for disease free survival model (right) and estimates for ‘disease free model’ (left)

**Survival analysis for deaths**

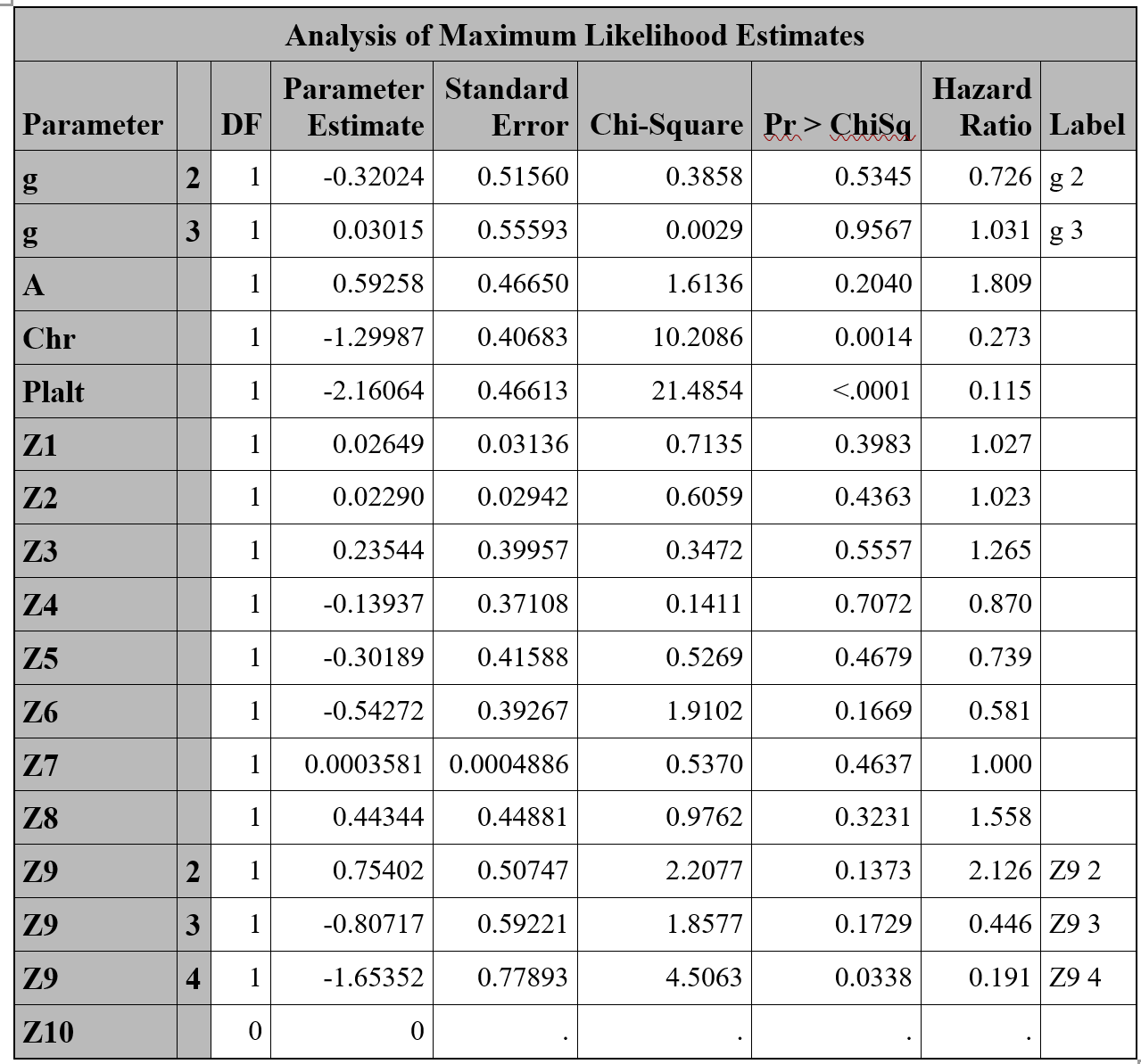
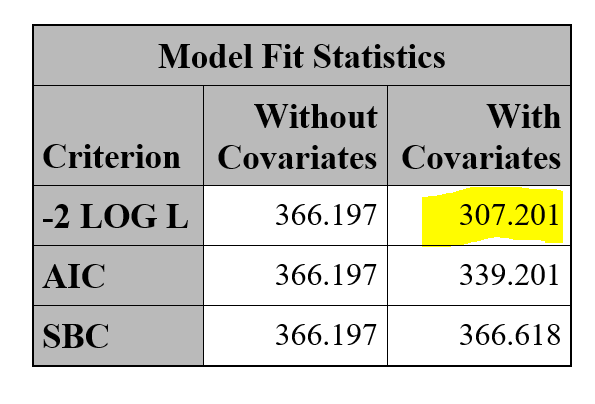
Relapses are considered censored data. The time variable used for modeling is T1.(**Fig 6)**

****

**Figure 6:**Fit statistics of survival model for deaths (right) and estimates for ‘time to deaths’ (left)

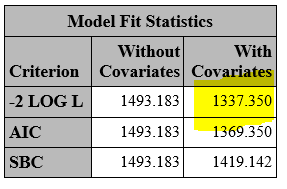
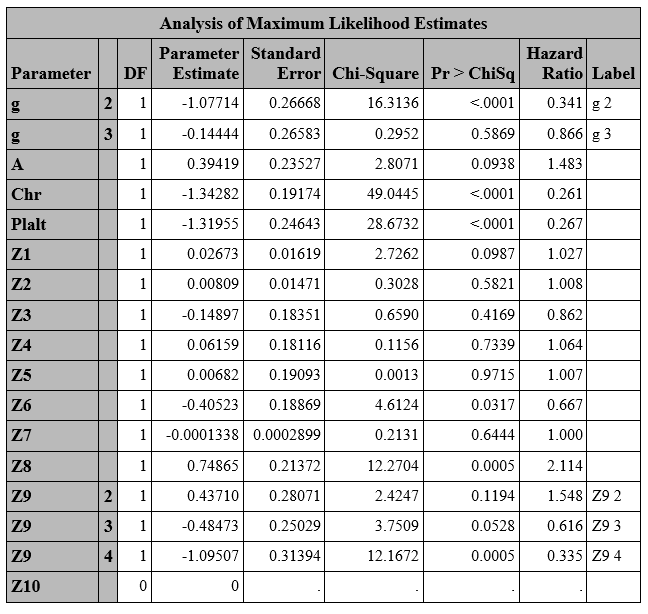
**Survival analysis for relapse**

Deaths are considered censored observations. The time used for modeling is T2.(**Fig 7)**

******Figure 7:** Fit statistics of survival model for relapse (right) and estimates for ‘time to relapse’ (left)

**Likelihood Ratio Test**

As the parameter estimates of covariates are clearly different for the three competing risks. In order to see if the differences are due to random sampling, we perform a likelihood ratio test. First, this required creating the dataset “combine,” which contains the indicator “type” to track the type of event. This allows us to compare the cox hazard models for death alone and relapse alone and still use the Efron method for resolving ties for all models. The following table shows the model fit for death and relapse simultaneously with the different baseline hazard functions but having the same covariate coefficients for the two event types.(**Fig 8**)



**Figure 8:**Fit statistics of survival model for the combined types (right) and estimates for ‘combined model’ (left)

The likelihood ratio test tests the following hypotheses:

**Ho**: all covariate coefficients of the two types are the same

**Ha**: some covariate coefficient of one type is different to that of the other

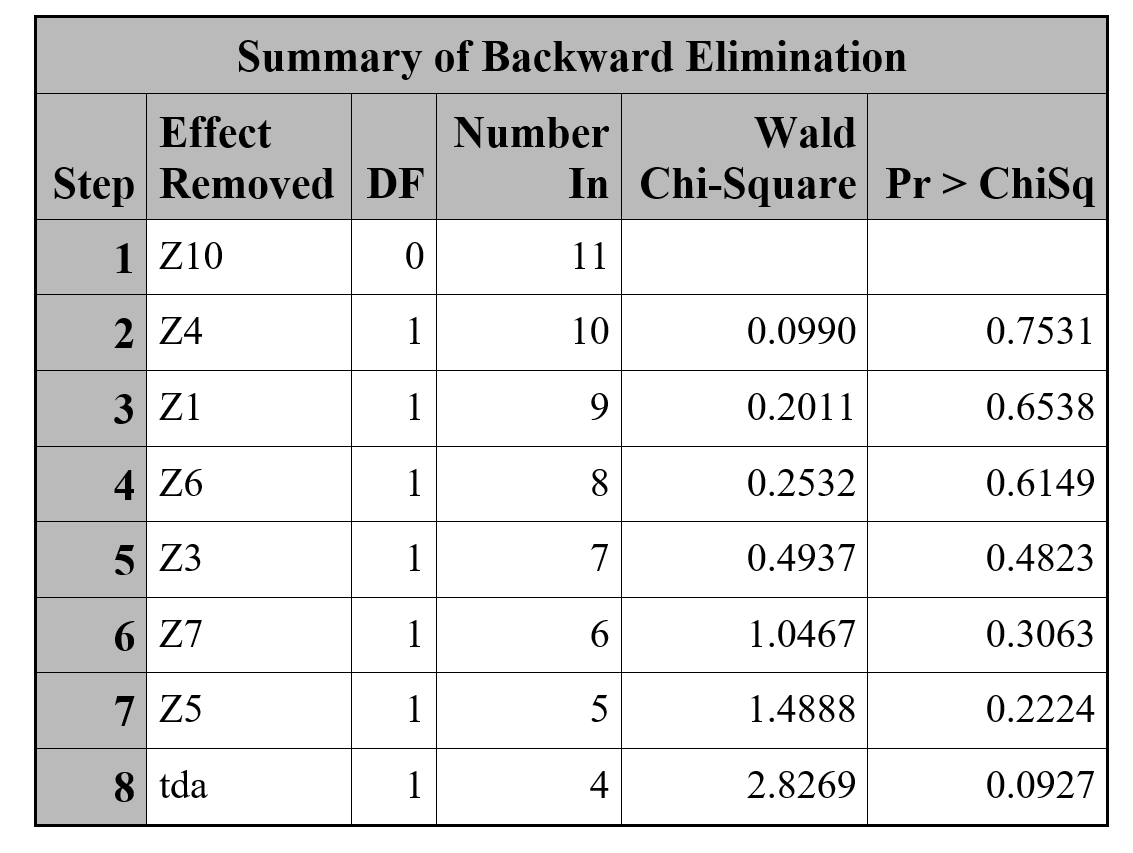
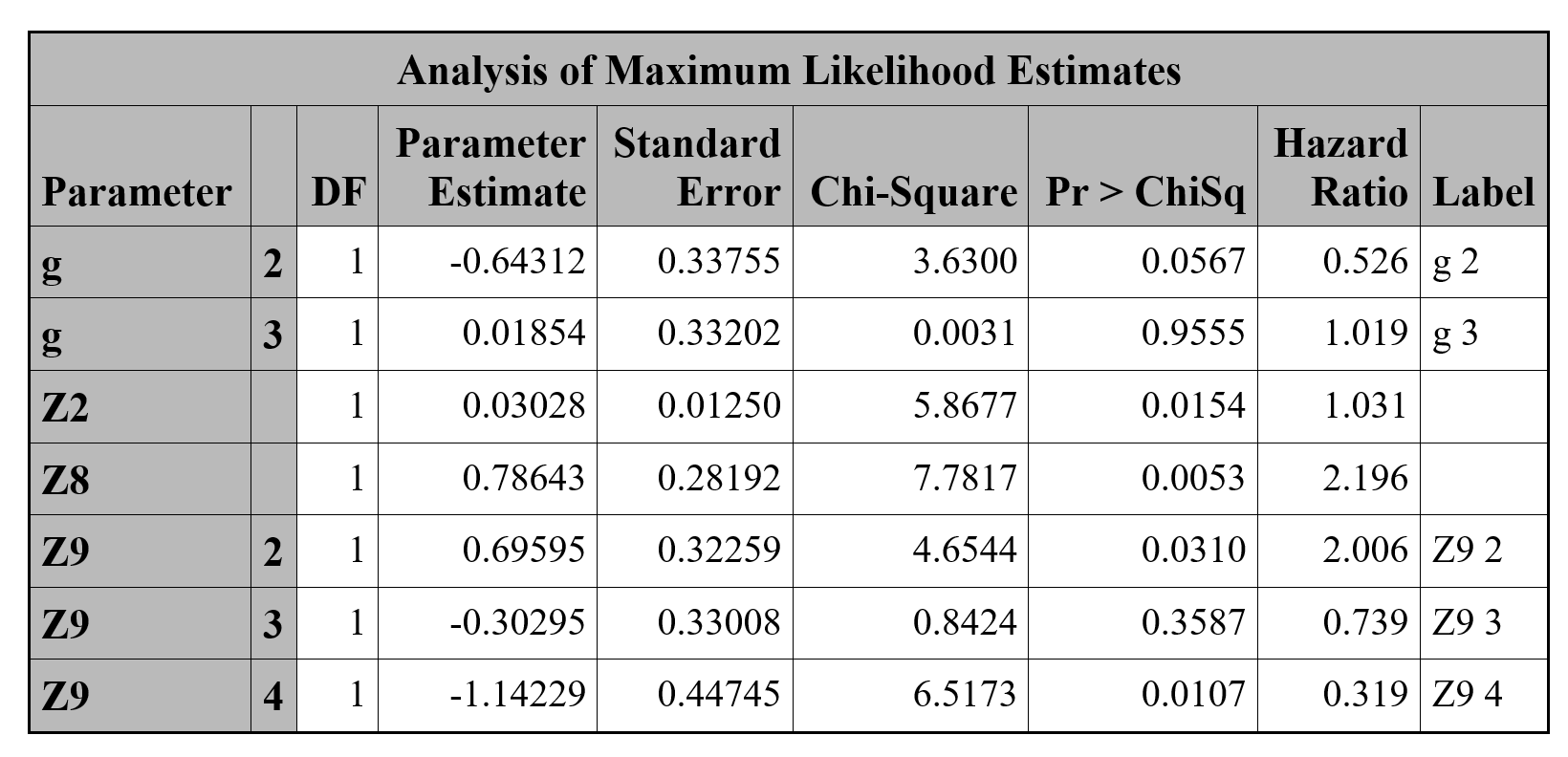
The likelihood ratio statistic is 1337.350-(330.987+307.201) = 699.162 with distribution with 17 degrees of freedom (Z10 has 0 df because of redundancy). The Chi-square probability of this statistic (p-value) is 0, suggesting strong evidence against the null hypothesis. Therefore, the covariate coefficients differ across the two event types.

Model with time-dependent covariates

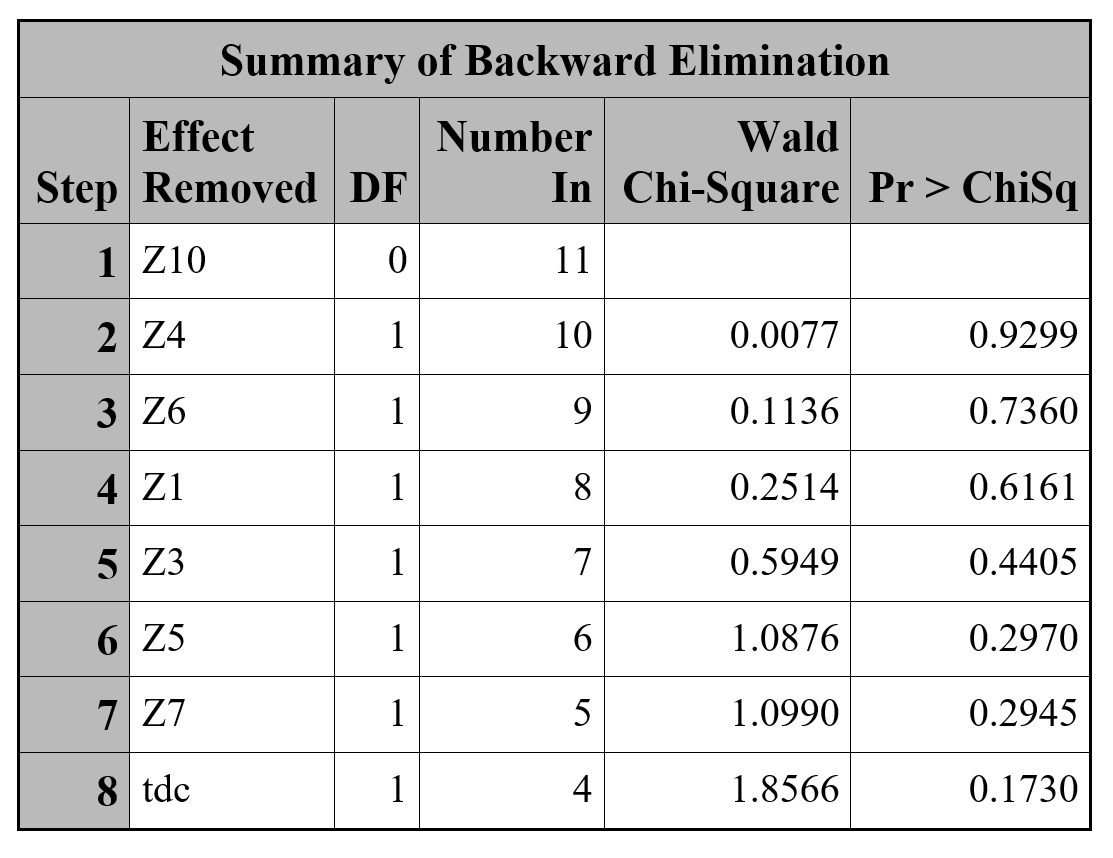
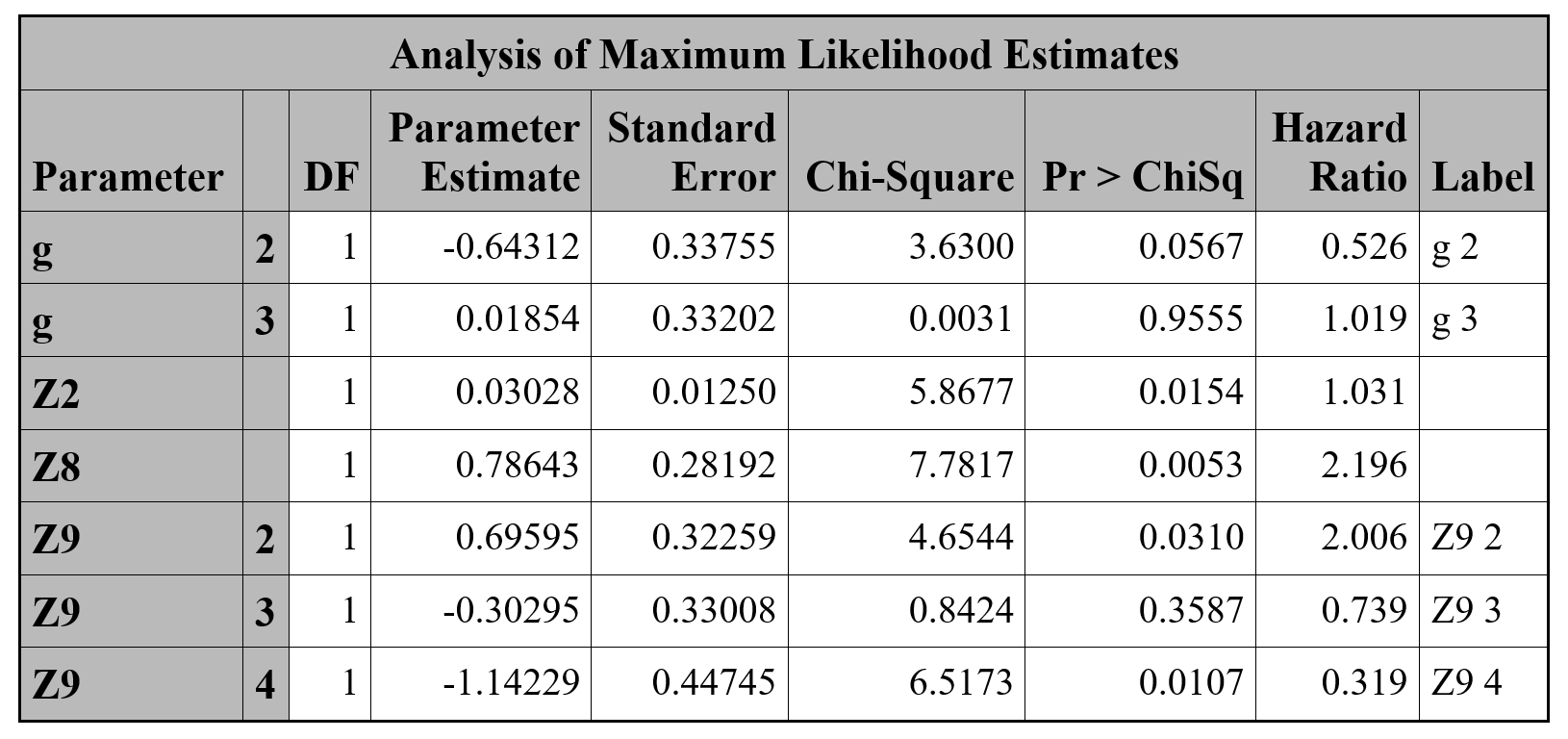
Given that model for each risk event is distinct, we construct a separate model for each event, focusing on time to death alone and time to relapse alone. For model selection, we tested the effects of each time-dependent covariate (acute GVHD, chronic GVHD, platelet recovery) individually. When there are only fixed-time variables, we can use the counting process with resampling to assess PH assumption on the remaining covariates. Since censoring is >40% (69% censored), we used Cox-Snell residual to assess goodness of model fit instead of deviance residuals. The counting process PH assessment and Cox-Snell residuals method are not available for models with time-dependent covariates.

Time to death using time-dependent covariates

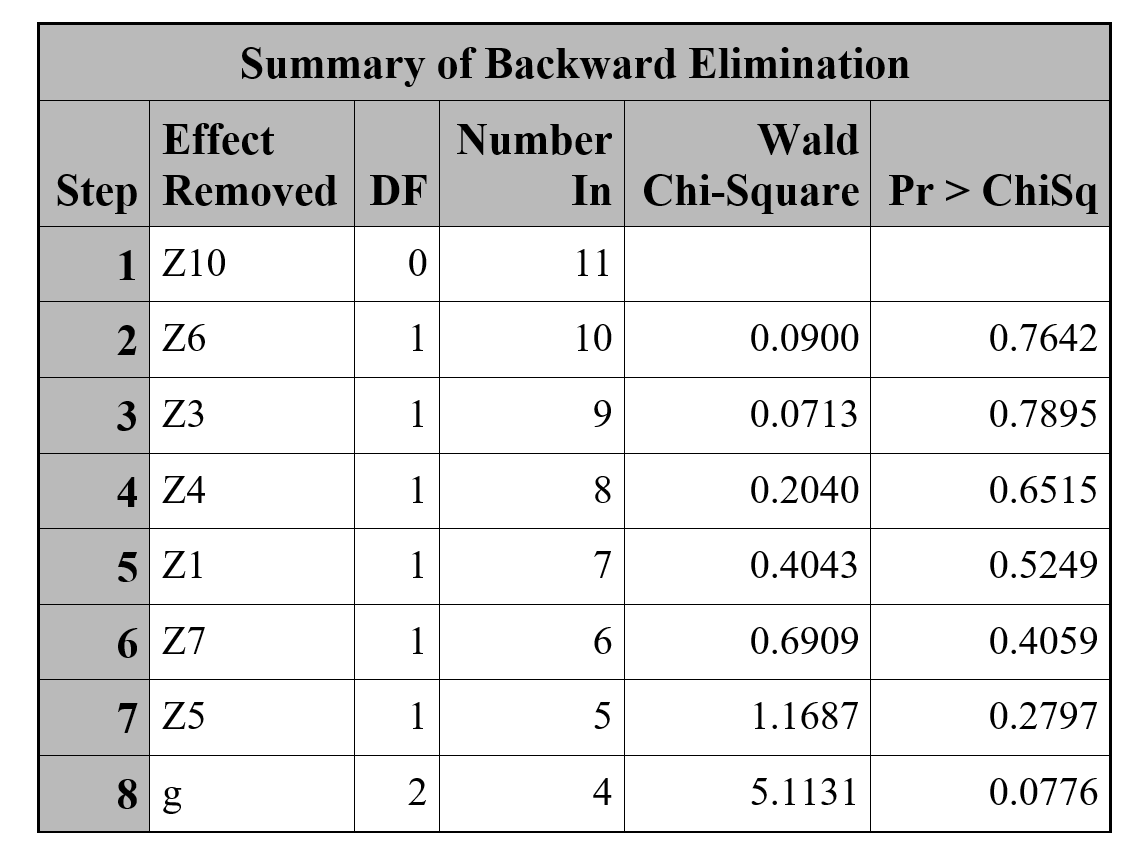
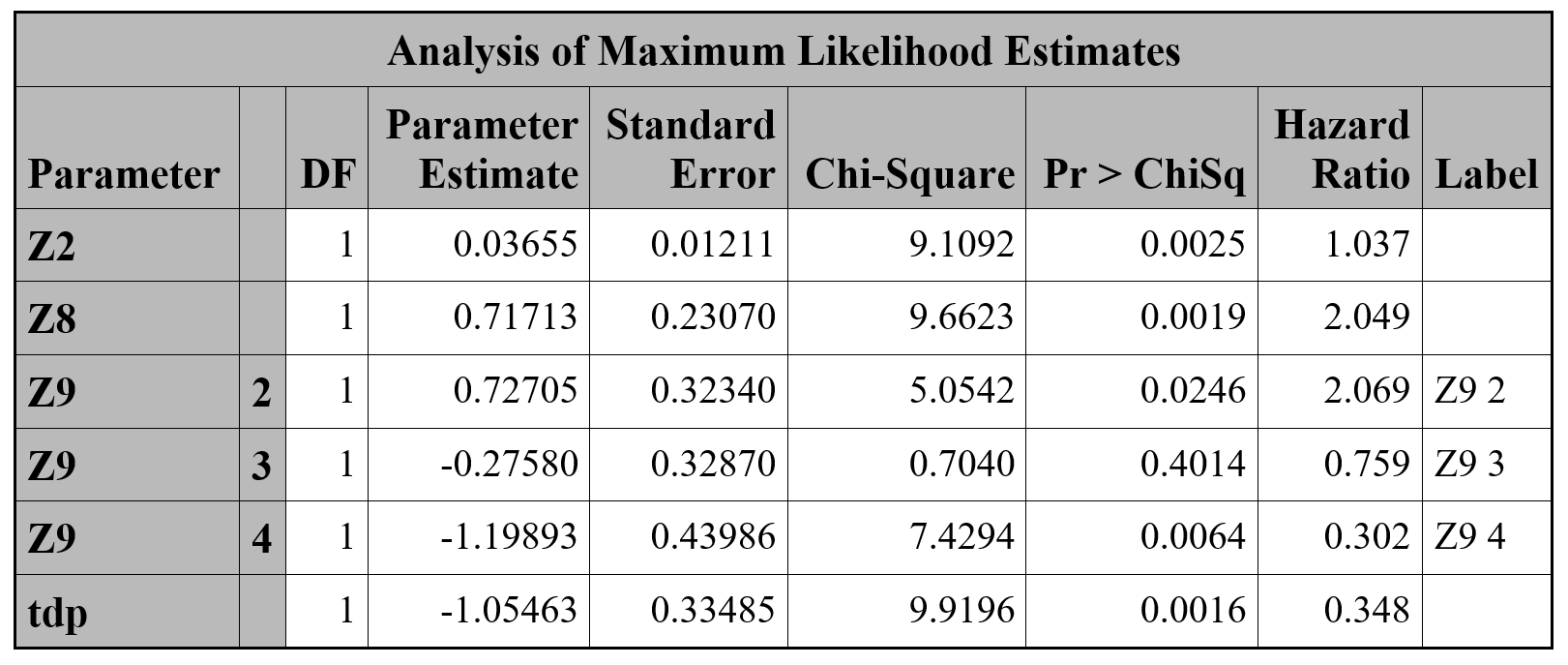
In time to death analysis, neither acute nor chronic GVHD (tda and tdc, respectively) are significant covariates; they are removed by backward elimination, where Wald Chi-square p-value for tda is 2.8269 (**Fig 9**)and for tdc is 1.8566 (**Fig 10**). However, platelet recovery (tdp) is significant (p-value=0.0016). (**Fig 11**) The covariates Z2, Z8, and Z9 are significant in all three time-dependent models except the covariate g in time dependent model (tdp).



**Figure 9:**  Results of backward selection process (for tda) with Efron approximation (right),and resulting model estimates (left).



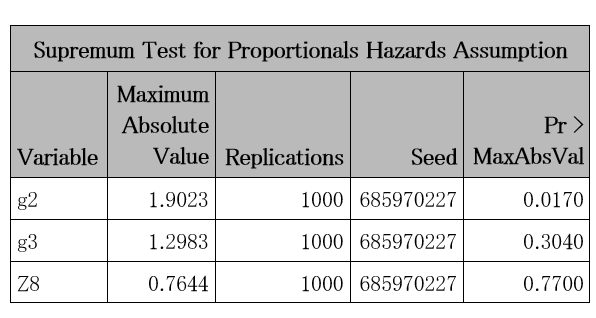
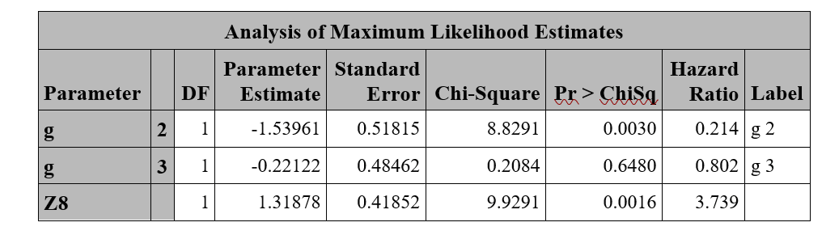
**Figure 10:** Results of backward selection process (for tdc) with Efron approximation (right),and resulting model estimates (left).

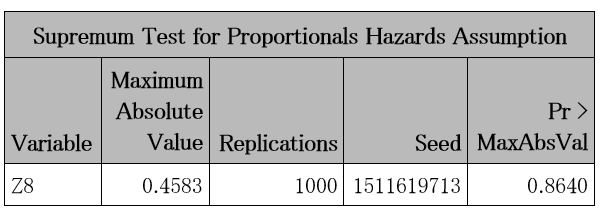
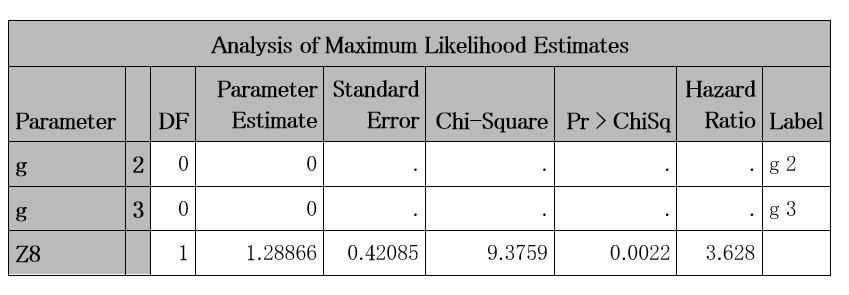


**Figure 11:** Results of backward selection process (for tdp) with Efron approximation (right),and resulting model estimates (left).

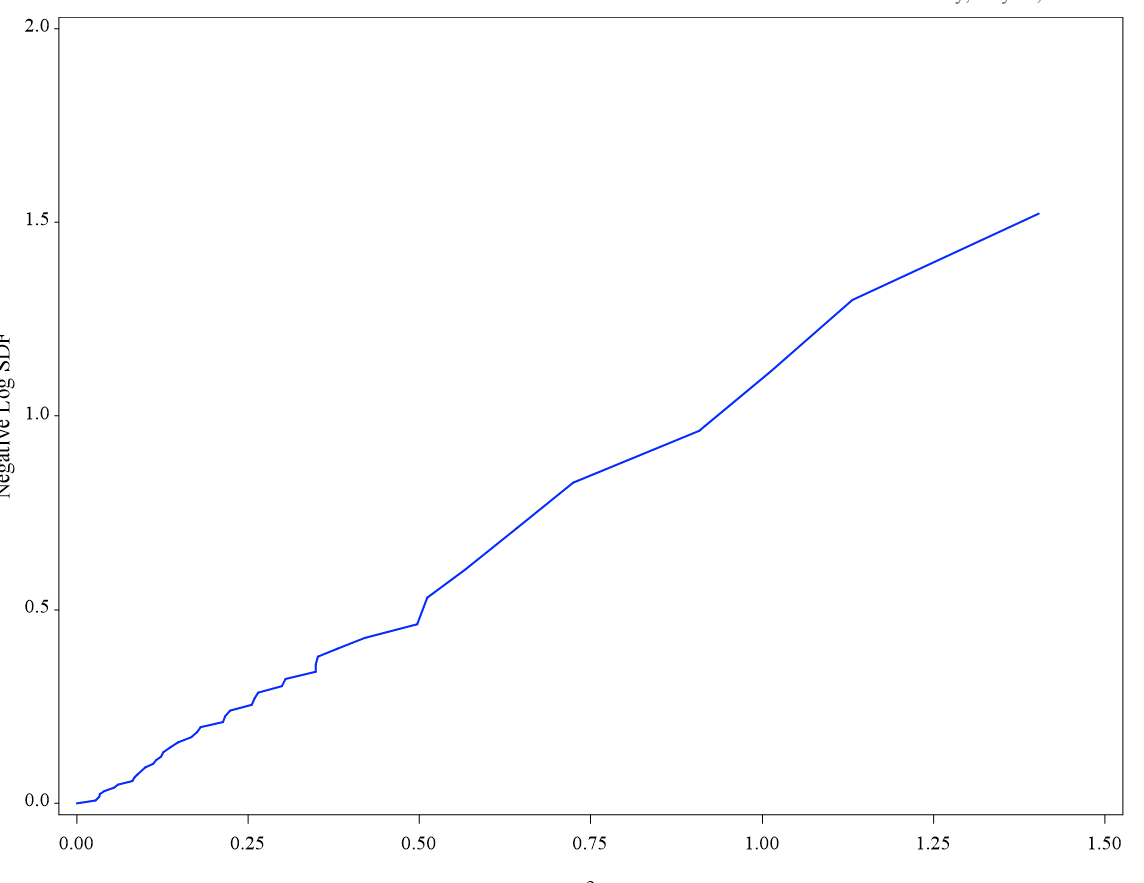
Time to Relapse

A similar model selection was performed on time to relapse. All the three models eliminated the time-dependent covariates along with the other insignificant variables except the covariates g and Z8 (**Fig 12**).

**Figure 12:**  Parameter estimates for time to relapse without strata on disease group (g) and supremum test for PH violation.



**Figure 13:**  Parameter estimates for time to relapse with strata on disease group (g) and supremum test for PH violation.



**Figure 14:** Cox-snell residual plot for the time to relapse

After assessing PH with the counting process, g has a level (g=2; AML low risk) that significantly violates the PH assumption (p-value = 0.014; Fig #). Hence, we stratified by g to derive the final model (Fig #). PH assessment for this model shows that Z8 satisfies the PH assumption (p-value=0 .86; Fig #). The Cox-Snell residual plot shows essentially a straight line through the original at a 45 degree angle with the x-axis (Fig #). Thus, the final model is a reasonable fit.

Interpretation of Covariates:

For time to death analysis, we have the following hazard model:

hx(y) = h0(y)\*exp{βTX}

log hx(y) = log h0(y) - 1.055\*tdp + 0.037\*Z2 + 0.717\*Z8 + 0.727\*(Z9=2) - 0.276\*(Z9=3) - 1.199\*(Z9=4)

When all other covariates are held constant, each covariate in the model can be interpreted in the following way. Firstly, a patient whose platelet has recovered has about 0.35 times the risk of death as that before recovery. Secondly, for each additional year in donor age, the risk of death increases by a factor of 1.037. Thirdly, the risk of death of a patient with FAB Grade 4 or 5 AML is 2.05 times that of one without it. Finally, compared to a patient at the Ohio State University Hospital (Z9=1), a patient at Alfred Hospital (Z9=2; Australia.) has a 2.07 times higher risk of death, a patient at St. Vincent Hospital (Z9=3; Australia) has a 0.759 times lower risk of death, and a patient at Hahnemann Hospital (Z9=4; U.S.) has a 0.302 times lower risk of death. The use of methotrexate (MTX) as Graft-Versus-Host-prophylaxis was dropped because of redundancy with the indicator for hospital site equal to the two Australian hospitals, which were the only two to useMTX.

For time to relapse analysis, we have the following hazard model:

log hx(y) = log h0(y) + 1.289 \*Z8

For a given strata of AML disease group (g=2 or 3), the risk of relapse for the patient with Z8 (FAB Grade 4 or 5 AML) is estimated to be 3.63 times that of a patient with Z8 at a significance level of 0.0022.

**Discussion**

In this study, we found that while the hazard functions for time to death and time to relapse were not significantly different when we do not consider covariates, the functions were markedly different (p-value = 0) when we include covariates. For time to death, model selection yielded donor age (Z2), indicator for FAB Grade 4 or 5 AML (Z8), and hospital site (Z9). Model selection for time to relapse yielded a model that only contained the disease group (g) and FAB indicator (Z8) covariates. The final model stratified the disease group to alleviate PH violation.

For both time to death and time to relapse, the fixed-time indicator Chr for chronic GVHD was significant in the Wilcoxon test but not in the time-dependent analysis. This apparent contradiction is probably due to the time-dependence of Chr. Using the fixed-time version of Chr skews the results because death precludes chronic GVHD and likely precludes relapse. Here, Chr is a consequence of rather than a contributor to death or relapse. Although Klein and Moeschberger suggest that acute and chronic GVHD may be anti-leukemic, the analysis in this study finds no evidence of this when accounting for time-dependence of the GVHD indicators.

In addition to evaluating PH assumption and goodness of fit for the time to death hazard model, we also need to consider whether our assumption of noninformative censoring is reasonable. In fact, censoring may be informative; patients who relapse may be more likely to die, and patients who die could have been at a higher risk for relapse had they been alive. Unfortunately, there is currently no definitive method to test this assumption.

In conclusion, this study suggests that bone marrow transplantation data can provide a rich opportunity for developing and applying survival analysis.