Acute Leukemia Report

Nada Ali, Machi Kaneko, Dante Ramirez, Jiayu Sun

2023-03-15

Introduction

Acute leukemias are blood and bone marrow cancers that include eight subtypes of acute myeloid leukemia (AML) and three subtypes of acute lymphocytic leukemia (ALL). Treatment for acute leukemia may include bone marrow transplant. For some people, bone marrow transplantation may cure their disease, but recovery after transplantation is a complex process. Prognosis after bone marrow transplantation likely depends on both patient factors and donor factors and is dynamic given the occurrence of events such as acute graft-versus-host disease (aGVHD) and recovery of normal platelet counts. For the purposes of our study, we will consider relapse of leukemia or patient death during remission as failure of transplantation events. We aim to conduct an exploratory analysis of patient factors, donor factors, and clinical events that may be predictive of survival after allogeneic transplant. Specifically, we aim to:

- Provide an estimate of disease-free survival in our study population
- Describe measured baseline variables of patient age, patient sex, donor age, donor sex, patient cytomegalovirus (CMV) immune status, donor CMV immune status, wait time from diagnosis to transplant, and use of methotrexate prophylaxis among disease groups and among French-American-British (FAB) subtypes of leukemia in our study population
- Explore whether any of the measured baseline variables are associated with differences in disease-free survival
- Explore whether occurrence of aGVHD is associated with improved disease-free survival or decreased risk of relapse to determine whether it is an important prognostic event
- Explore whether any of the measured baseline variables are associated with differences in disease-free survival among patients who develop aGVHD
- Explore whether use of prophylactic methotrexate is associated with the risk of developing aGVHD, adjusting for potential confounding from other baseline variables
- Explore whether recovery of normal platelet levels is associated with improved disease-free survival or decreased risk of relapse
- A total of 137 patients were enrolled after allogeneic bone marrow transplantation at four different hospitals in the United States and Australia between March 1, 1984 and June 30 1989. All patients were followed until death or end of the study.

Methods

Terminating Event and Time

Our team was tasked with identifying associations and trends in survival times with two separate terminating events. In directives 1-5, where our objectives revolved around disease free survival time, our team identified *deltadfs* as the terminating event of interest. This terminating event is a binary indicator of death or relapse of disease among our patient sample. In directives 6-7, the research question shifted to the developmental risk of acute graft-versus-host disease (aGVHD) in our patient sample, and thus, the binary indicator of aGVHD onset *deltaa* was chosen as our terminating event.

Approaching the data with the research question of disease free survival time in mind, our team decided to classify *tdfs*, the time in days until death, relapse, or censoring, as the time argument when analyzing directives 1-5. The research question of directives 6 and 7 handles the time until onset of acute graft-versus-host disease, and thus *ta*, the time in days until onset of aGVHD was used as the time argument in the respective survival objects.

Significance and Family-wise Error

Given our fairly low clinical sample size of $n = 137$ patients, our tests of significance will have lower power of detecting true differences in groups if they exist. Based on this and observations of significance from Thiese, Ronna, and Ott (Thiese et al. 2016), we deliberately set our significance level to an unconservative $\alpha = .1$. This large significance level allows us to account for our small sample size, in addition, it will allowing us to correct for the family-wise error we will encounter through multiple testing without over-correcting with a practically unrejectable p-value. We chose to correct for family-wise error through a Bonferroni correction. In directives 2 and 3, we run 8 log-rank hypothesis tests of significance. Thus, our family-wise error adjusted significance level of rejection will be $\frac{.1}{8} = .0125$.

Test Statistics

To test the survival time and onset time differences between groups, we will be using the log-rank test as an *a priori* choice of statistic based on what we believe are appropriate scientific considerations. We don't believe there to be any basis on which to weigh early terminations more heavily, thus we do not see the Gehan-Breslow generalized Wilcoxon statistic as appropriate. Additionally, we expect to see proportional hazards among our disease and FAB groups. Thus, the log-rank test statistic will be used as our method of nonparametric testing of equal survivorship between groups.

Results and Discussion

Descriptive Statistics

This data set includes 137 patients, of which 39% are censored for our chosen terminating event of death or relapse of disease. The mean age is approximately 28 years old, and there are 58.4% males in the data set. 49.6% of our patients have positive cytomegalovirus (CMV) immune status,

while 29.2% report prophylactic use of methotrexate (MTX). We also collected similar biomarker predictors on our donor sample, and for the same covariates of mean donor age, proportion male, and proportion CMV status, we measured ~ 28 years old, 64.2%, and 42.3% respectively.

Directive 1

When attempting to *estimate the disease-free survival time for patients enrolled in this study*, we first plotted *(figure 1)* the non-parametric Kaplan-Meier estimator in an attempt to gain knowledge of the underlying shape of our data. The estimated median survival time was **481 [363, 748] days**. The curve made us hopeful to use a parametric estimator. We began with a Weibull and generalized gamma estimate, which we plotted *(figure 2)* in conjunction with the Kaplan-Meier to eye-test it's effectiveness at modeling the data. Using a likelihood ratio test, we can empirically answer if the Weibull is an appropriate simplification of the generalized gamma, in which case we will use the Weibull's parameter estimates. The LRT test statistic can be computed as

$2 \times (657.77 - 650.19) \approx 15.16$

which signifies with a p-value 9.88×10^{-5} that the Weibull distribution is not an appropriate simplification of the generalized gamma with our data. We can see that with the scale $p = 0.59[0.48, .70]$ and shape $\lambda = .0007[0004, 0009]$ parameters of the Weibull distribution resulting in a large maximized loglikelihood value of -657.77, we do not have a good fit. However, the parameter estimations of the gamma don't fair much better given the maximized loglikelihood value of -650.19. The parameters for the generalized gamma are as follows, $\mu = 6.23[5.52, 6.94]$ $\sigma = 2.31[1.89, 2.74]$, and $Q = -0.4[-1.06, .27]$. Given the poor likelihood values for even the most robust of parametric estimators, we elect a non-parametric estimate in the form of the Kaplan-Meier estimator. This allows us to provide estimates like the median survival time as described prior.

Directive 2

Disease

Subgrouping by disease groups, we note some interesting features that indicate ease of analysis by means of proportional hazards visible in the survival curves in *figure 3*. According to the table in *table 1*, we notice approximately equal counts of each disease group in our patient sample, with an average of 46 patients per group. The occurrence of censoring is also retentively proportional between disease groups, with an average of 28 censored observations per group. Summary statistics such as mean age per group were calculated in the same table. Proportions of groups with biomarkers such as aGVHD, CMV (both in patient and donors), and MTX, were calculated to identify potential associations between these baseline measurements and the disease grouping classification. A chi-square test of independence was preformed between selected variables which we determined to differ more so than just random noise we would expect from a low sample data set. These variables we tested against disease group were *mean age, count males, count CMV, count MTX, count Hospital, count donor males, and count donor CMV*. Of these covariates, only *count MTX* and *count Hospital* had significant p-values of .045 and .0099 respectively, while only *count Hospital* had a p-value below our family-wise error rate corrected significance level $\alpha = .0125$. Thus, we determine this variable is associated with disease grouping.

Looking at our plot *(figure 3)* we see Disease Group 2 has the highest median survival time by a wide margin at **2204 [641, NA] days**. Disease Groups 1 and 3 have much lower median survival times at **418 [192, NA] days** and **183 [390] days** respectively. Using the logrank test of differences in subgroups, we observed a **p-value of .001**, which is lower than our family-wise error rate corrected significance level, and thus we reject the null hypothesis and say there is evidence of association between disease grouping and survival time.

FAB

Subgrouping by FAB leukemia classification, we note even better indications of proportional hazards between groups as seen in *figure 4*. According to the table in *table 2*, the dichotomy of distribution between these two classifications is a bit more apparent. We observe 92 patients NOT in the FAB classification of leukemia, and 45 patients with this classification. Censoring is approximately equal in counts, with 48 observed deaths or relapses in our FAB 0 classification, and 35 in our FAB 1 group. Again, summary statistics such as mean age per group were calculated in a *table 2*. Proportions of groups with biomarkers such as aGVHD, CMV (both in patient and donors), and MTX, were calculated to identify potential associations between these baseline measurements and FAB leukemia grouping classification. A chi-square test of independence was preformed between selected variables which we determined to differ more so than just random noise we would expect from a low sample data set. These variables we tested against disease group were *mean age, count males, count CMV, count MTX, count Hospital, count donor males, and count donor CMV*. Of these covariates, only *count MTX* and *count donor CMV* had significant p-values of .06 and .09 respectively, but neither of these had a p-value below our family-wise error rate corrected significance level $\alpha = .0125$. Thus, we determine no variables are associated with FAB leukemia classification.

Looking at our plot *(figure 4)* we see FAB Classification 0 has the highest median survival time at **704 [390, NA] days**. FAB Classification 1 has the lower median survival time at **272 [164, 456] days**. Using the logrank test of differences in subgroups, we observed a **p-value of .0037**, which is lower than our family-wise error rate corrected significance level, and thus we reject the null hypothesis and say there is evidence of association between FAB classification and survival time.

Directive 3

In *figures 5-10*, our team plotted every meaningful factor grouping to identify potential associated differences in disease-free survival. We observed two different trends. In the all but one of the figures *(figures 6, 7, 7, 8, and 10)*, using the logrank test of differences in disease free survival, we measured p-values between the range of **.092 to .97**. Many of these p-values are much larger than our family-wise error rate corrected significance level $\alpha = .0125$. So, in an empirical and purely exploratory sense we won't assume any association between disease free survival time and *patient sex, patient CMV status, patient MTX status, donor sex, and donor CMV status*.

The one grouping which did result in significant p-values based on the logrank test of differences in disease free survival was *figure 9*, where we grouped by the four different hospitals. This grouping resulted in a **p-value of .0034**, well below our our family-wise error rate corrected significance level. There is a notable contrast in sample sizes between these four groups, with Hospital 1, Hospital 2, Hospital 3, and Hospital 4 having sample sizes of $n = 76, 17, 23,$ and 21 respectively. Additionally, we have major inequities in the amount of censored observations, with $50, 13, 13,$ and 7 for the four

respective hospitals. This explains why there is no median survival time observed for Hospital 4, as not enough event have occured in our data set for that particular subgroup. For the remaining hospitals, we have very similar median survival times at **467 [381, 667] days** and **466 [129, NA] days** for Hospitals 1 and 3 respectively. Hospital 2 by far had the lowest median survival time at **107 [53, 332] days**. Perhaps the variable *hospital* is a proxy for location, seeing as there four hospitals were based in different locations, and perhaps this explain these sharp contrasts in sample size, events, and survival time. Additionally, we see that Hospitals 1 and 4 did not have any patients adminstered with MTX, which could also be telling of interaction between these two variables on the association with disease free survival time.

hospital	mtx
1	0
2	17
3	23
4	0

Directive 4

In order to characterize the relationship between occurrence of aGVHD after transplantation and disease-free survival, we fit the following cox-proportional hazards model adjusting for age, patient cmv and donor cmv status stratifying by hospital.

$$h(t|\text{gvhd.tv(t)}, \text{age}, \text{cmv}, \text{donorcmv}, \text{hospital}) = h_0(t|\text{hospital})e^{\beta_1 gvhd.tv(t) + \beta_2 age + \beta_3 cmv + \beta_4 donorcmv}$$

where gvhd.tv(t) is the time-varying covariate for aGVHD status at time t and $h0(t|\text{hospital})$ denotes the distinct baseline hazard function at time t corresponding to each hospital. Because in this model we estimate the hazard of the composite outcome of relapse and death, we did not perform competing-risk analysis.

When estimating the association of aGVHD and risk of relapse, we employed the Fine-Gray model to account for death as a competing risk to risk of relapse. We used the crr function in R.

We estimate that individuals who developed acute GVHD by a given time had a 59% higher hazard of death or relapse than individuals who did not develop acute GVHD by that time, with fixed age, patient and donor CMV immune status and hospital.. The estimated hazard of relapse or death comparing individuals who developed acute GVHD by a given time to individuals who did not develop acute GVHD by that time is **1.59 (95% CI: 0.89, 2.83, p=0.116)**.

The hazard of relapse is estimated to be approximately **35.8% lower** in patients who developed acute GVHD by a given time compared to those who did not develop acute GVHD by that time, accounting for death as a competing risk, with fixed age, patient and donor CMV immune status and hospital. The hazard ratio is estimated to be **0.642** with an associated **95% CI of (0.242, 1.70, p=0.37)**.

Given these findings, we can conclude that there is no evidence to suggest that the development of acute GVHD is an important prognostic event.

Directive 5

We analyzed the baseline characteristics and disease-free survival in the subset of the study population who developed acute GVHD in the observation period. We fit a survival object to this subpopulation and used log-rank tests to determine if baseline characteristics, including patient sex, donor sex, patient CMV immune status, donor CMV immune status, disease group, FAB classification, methotrexate prophylaxis or hospital, were associated with differences in disease-free survival.

None of the following baseline characteristics were found to have a statistically significant association with disease-free survival among patients who developed aGVHD following transplant at a family-wise adjusted significance level of $=.0125$ using log-rank test: patient sex **(p=0.6)**, donor sex **(p=0.7)**, patient CMV status **(p=0.6)**, donor CMV status **(p=0.2)**, disease group **(p=0.08)**, FAB classification **(p=0.4)**, methotrexate prophylaxis **(p=0.3)** and hospital **(p=0.1)**.

We noted that among the subgroup of patients who developed acute GVHD, death and relapse events occurred early in the time course. We also noted that many of the Kaplan-Meier curves cross early in the time course. Given these observations, we recognized that the proportional hazards assumption was not met, and that use of the log-rank test may not appropriately test the difference in disease-free survival based on the baseline characteristics. Use of the Gehan-Breslow-Wilcoxon test was considered. While this test does not require consistent hazard ratios, the assumption that one group has a consistently higher hazard must hold. This was not deemed appropriate as most survival curves cross. In future studies the Fleming Harrington test could be applied.

Directive 7

Directive 6

Limitations and Conclusion

Our exploratory analysis suggests that survival among patients who undergo allogeneic transplantation for acute leukemia is at least 1 year. When evaluating differences in measured baseline variables between disease groups, we found that there was an association between the hospital at which the patient was recruited and their disease group. We accounted for this by stratifying our cox proportional hazards models of survival by hospital. Additionally, we accounted for multiple comparisons by utilizing a Bonferroni correction and found no association between disease free survival and any of the measured baseline variables. We also found that aGVHD and recovery of normal platelet levels are not associated with improved disease free survival or decreased risk of relapse, suggesting that they are not important prognostic events. Additionally, our data suggest that use of prophylactic methotrexate is not associated with the risk of developing aGVHD. Two limitations of this analysis are the small sample size of our study population and the large proportion of censored survival times. These two limitations may make it challenging to identify confounding variables which would alter our unadjusted hazard estimates. To address this, future work could include recruitment of a larger sample size and longer follow up time.

References

- Thiese MS, Ronna B, Ott U. P value interpretations and considerations. J Thorac Dis. 2016;8(9):E928-E931. doi:10.21037/jtd.2016.08.16

- Marubini E, Valsecchi MG. Analysing Survival Data from Clinical Trials and Observational Studies. John Wiley & Sons; 2004.

- Moore DF. Applied Survival Analysis Using R. Springer International Publishing; 2016. doi:10.1007/978-3-319-31245-3

- Collett D. Modelling Survival Data in Medical Research. 3rd ed. Chapman and Hall/CRC; 2015. doi:10.1201/b18041

- Breslow NE, Day NE. Statistical Methods in Cancer Research: The Design and Analysis of Cohort Studies. IARC; 1987.

- Kleinbaum DG, Klein M. Survival Analysis: A Self-Learning Text. Springer; 2012. doi:10.1007/978-1-4419-6646-9

- Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data. Springer; 2003. doi:10.1007/b97377
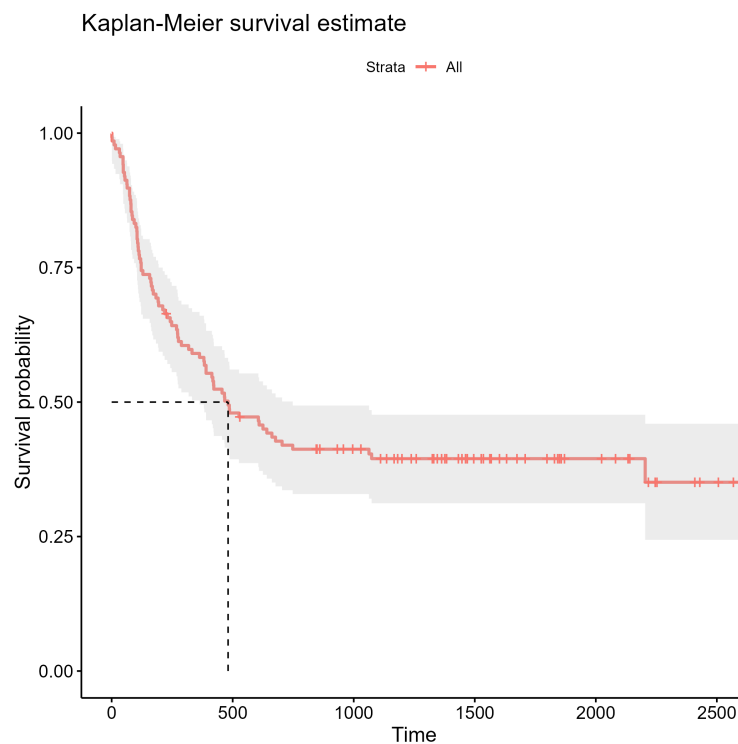
Tables and Figures

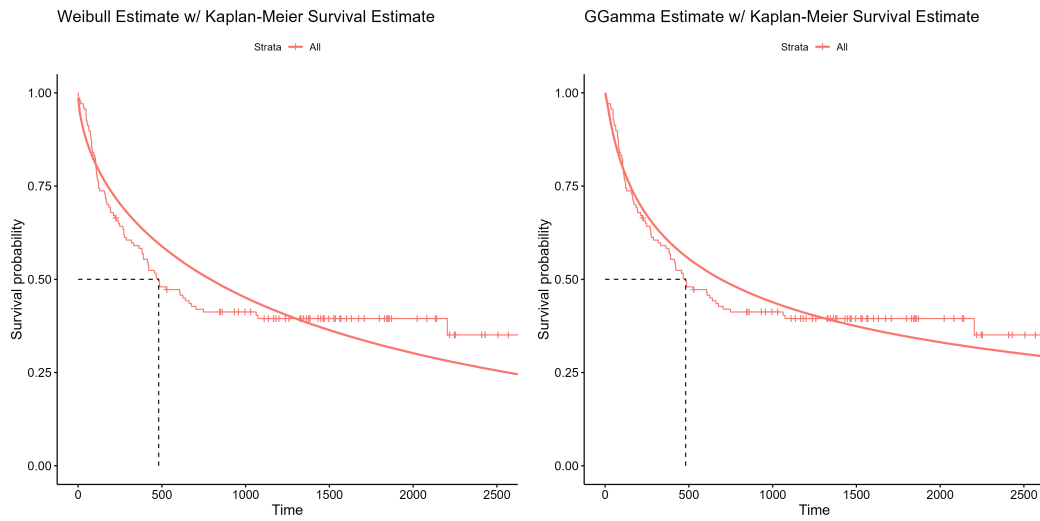Figure 1: Kaplan-Meier Estimate, Disease Free Survival Time

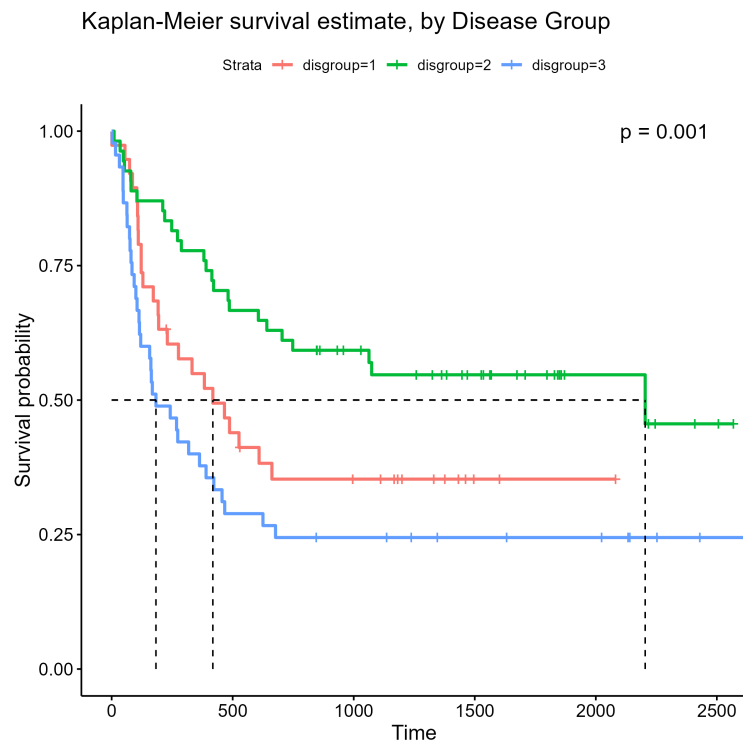Figure 2: Weibull and GGamma Estimates, Disease Free Survival Time

Figure 3: Kaplan-Meier Estimate, Disease Free Survival Time

Table 1: Call: s_bmt ~ disgroup Chisq = 13.803722 on 2 degrees of freedom, p = 0.001006

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
disgroup=1	38	24	21.85	0.2112	0.2893
disgroup=2	54	25	39.97	5.604	11.01
disgroup=3	45	34	21.18	7.756	10.53

|	Disease Group 1	Disease Group 2	Disease Group 3
mean age	24.421	29.407	30.444
sd age	7.295	8.764	11.220
count males	26.000	30.000	24.000
prop males	0.684	0.556	0.533
count females	12.000	24.000	21.000
prop females	0.316	0.444	0.467
count cmv	15.000	26.000	27.000
prop cmv	0.395	0.481	0.600
count mtx	17.000	12.000	11.000
prop mtx	0.447	0.222	0.244
count hospital	64.000	118.000	81.000
mean donorage	26.789	28.074	29.933
sd donorage	8.933	9.245	12.057
count donormales	26.000	34.000	28.000
prop donormales	0.684	0.630	0.622
count donorcmv	17.000	22.000	19.000
prop donorcmv	0.447	0.407	0.422
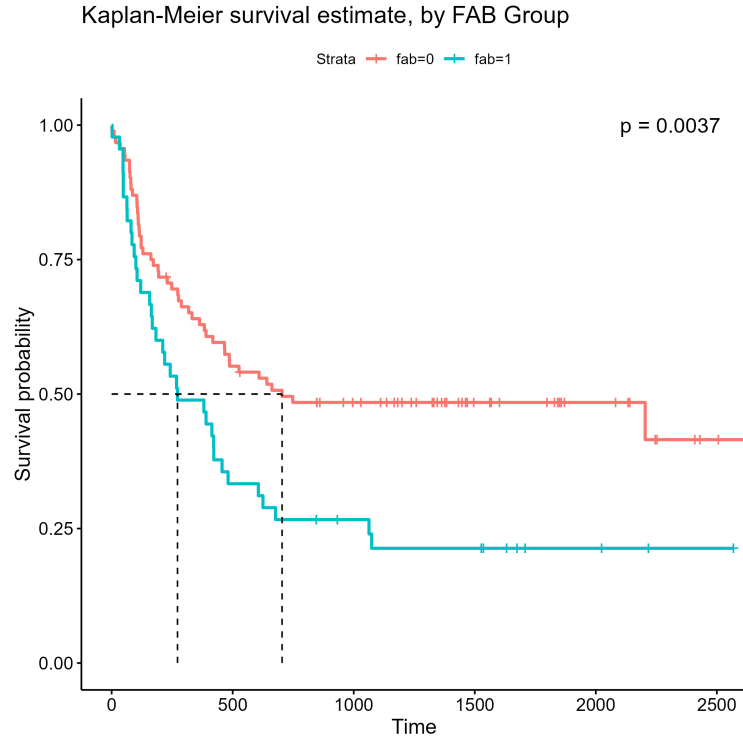
Figure 4: Kaplan-Meier Estimate, Disease Free Survival Time

Table 2: Call: s_bmt ~ fab Chisq = 8.435337 on 1 degrees of freedom, p = 0.003680

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
fab=0	92	48	59.83	2.337	8.435
fab=1	45	35	23.17	6.034	8.435

|	FAB Classification 0	FAB Classification 1
mean age	28.598	27.889
sd age	9.478	9.810
count males	56.000	24.000
prop males	0.609	0.533
count females	36.000	21.000
prop females	0.391	0.467
count cmv	44.000	24.000
prop cmv	0.478	0.533
count mtx	32.000	8.000
prop mtx	0.348	0.178
count hospital	178.000	85.000
mean donorage	29.000	26.956
sd donorage	9.669	11.133
count donormales	58.000	30.000
prop donormales	0.630	0.667
count donorcmv	44.000	14.000
prop donorcmv	0.478	0.311
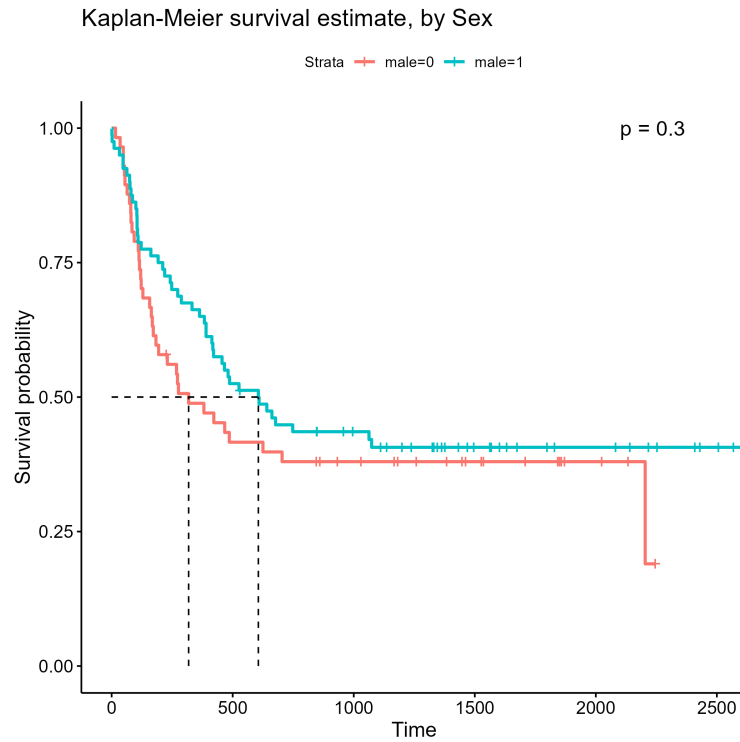
Figure 5: Kaplan-Meier Estimate, Disease Free Survival Time

Table 3: Call: s_bmt ~ male Chisq = 1.078766 on 1 degrees of freedom, p = 0.298974

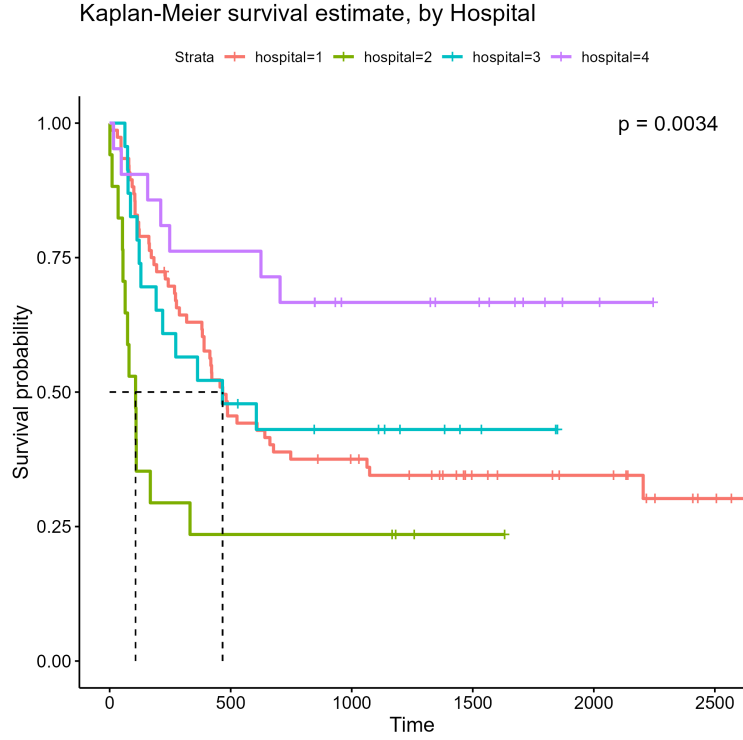|	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
male=0	57	36	31.42	0.6662	1.079
male=1	80	47	51.58	0.4059	1.079
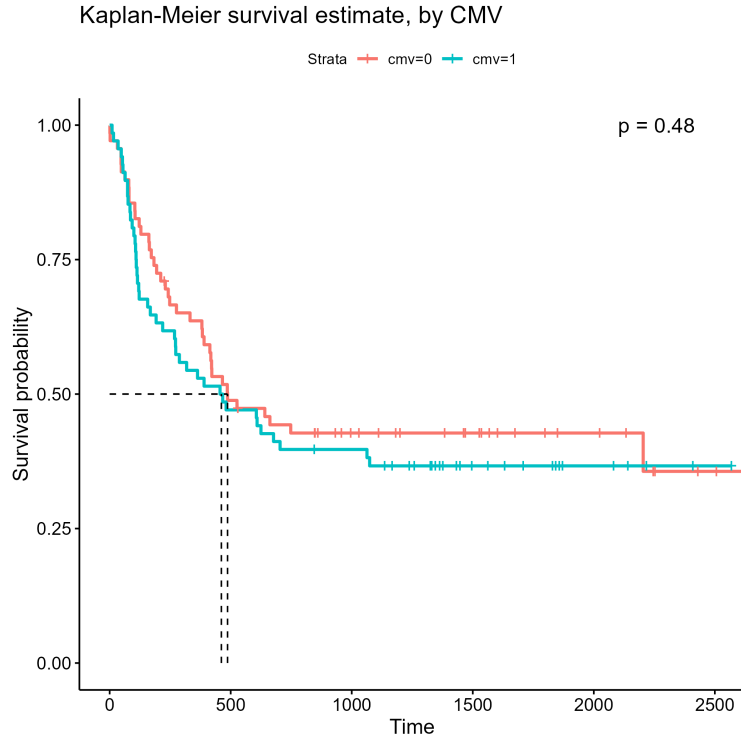
Figure 6: Kaplan-Meier Estimate, Disease Free Survival Time

Table 4: Call: s_bmt ~ cmv Chisq = 0.497423 on 1 degrees of freedom, p = 0.480635

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
cmv=0	69	40	43.2	0.2375	0.4974
cmv=1	68	43	39.8	0.2579	0.4974
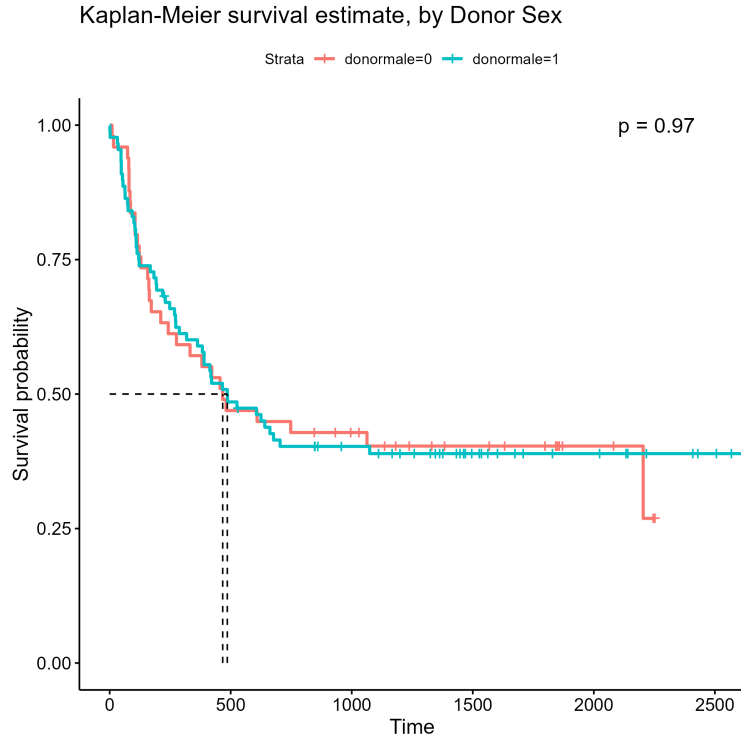
Figure 7: Kaplan-Meier Estimate, Disease Free Survival Time

Table 5: Call: s_bmt ~ donormale Chisq = 0.001359 on 1 degrees of freedom, p = 0.970591

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
donormale=0	49	30	29.84	0.0008686	0.001359
donormale=1	88	53	53.16	0.0004875	0.001359

Kaplan-Meier survival estimate, by Donor CMV

Strata — donorcmv=0 — donorcmv=1

p = 0.84
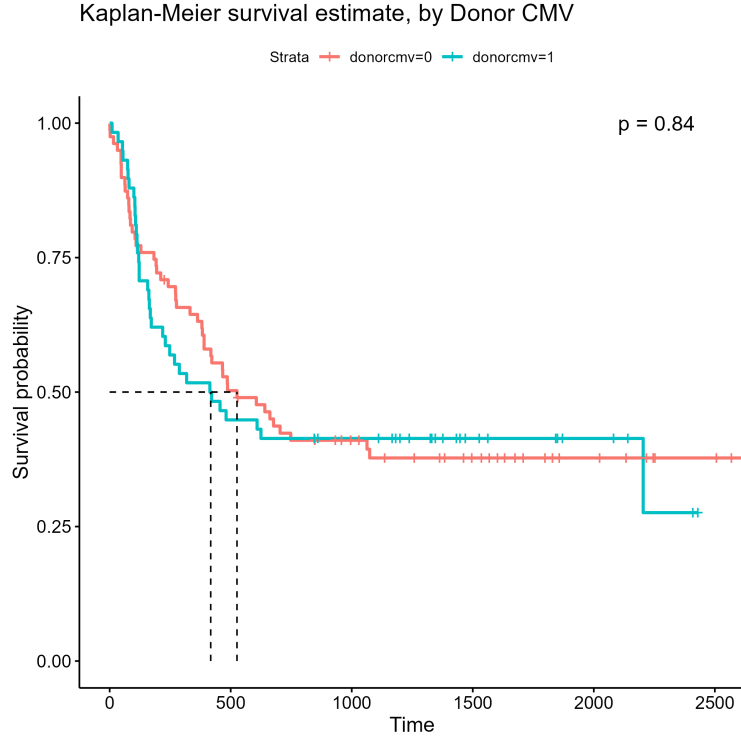
Figure 8: Kaplan-Meier Estimate, Disease Free Survival Time

Table 6: Call: s_bmt ~ donorcmv Chisq = 0.043347 on 1 degrees of freedom, p = 0.835073

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
donorcmv=0	79	48	48.93	0.01772	0.04335
donorcmv=1	58	35	34.07	0.02544	0.04335

Figure 9: Kaplan-Meier Estimate, Disease Free Survival Time

Table 7: Call: s_bmt ~ hospital Chisq = 13.680494 on 3 degrees of freedom, p = 0.003374

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
hospital=1	76	50	47.71	0.1101	0.2613
hospital=2	17	13	5.905	8.524	9.258
hospital=3	23	13	13.62	0.02779	0.03339
hospital=4	21	7	15.77	4.879	6.076
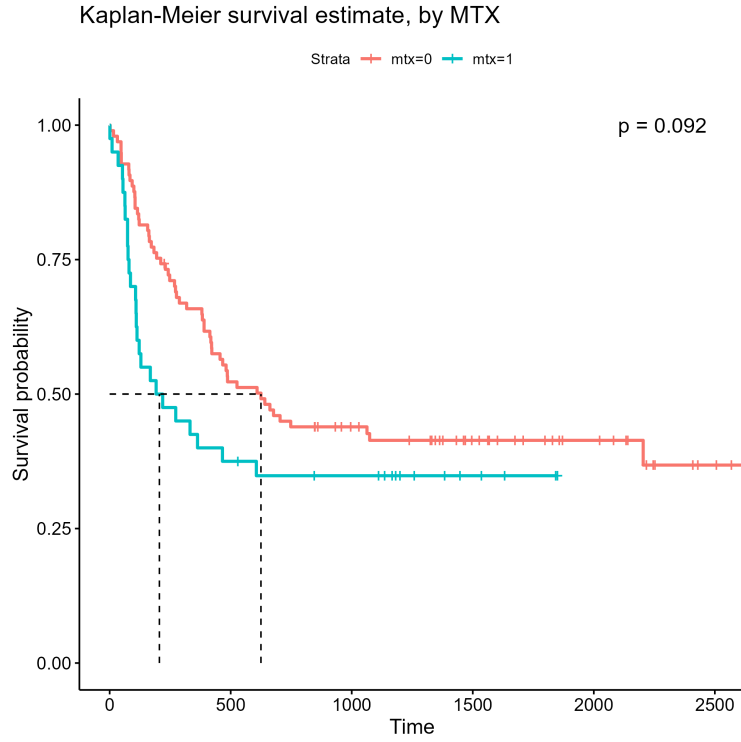
Figure 10: Kaplan-Meier Estimate, Disease Free Survival Time

Table 8: Call: s_bmt ~ mtx Chisq = 2.838053 on 1 degrees of freedom, p = 0.092056

	N	Observed	Expected	(O-E)^2/E	(O-E)^2/V
mtx=0	97	57	63.48	0.6614	2.838
mtx=1	40	26	19.52	2.151	2.838
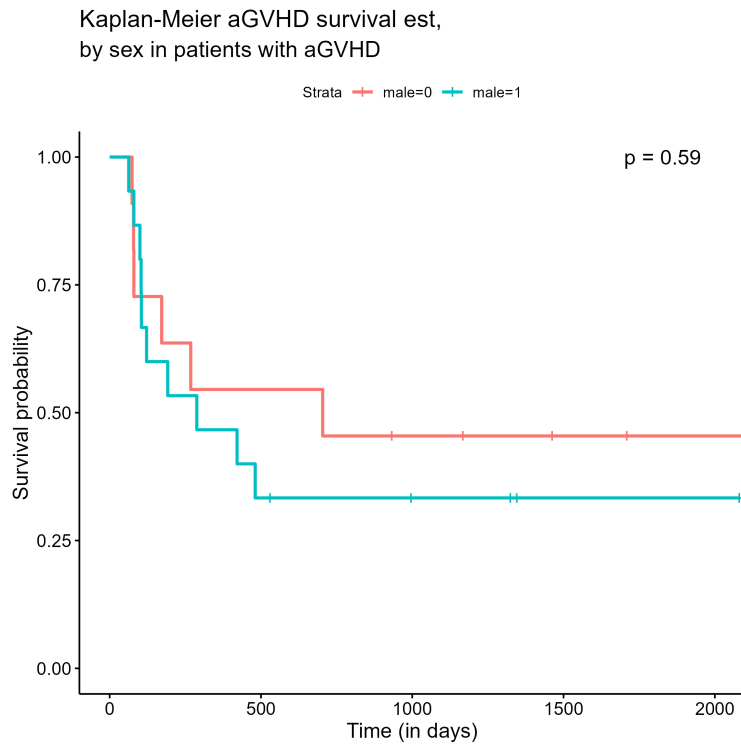
Figure 11: Kaplan-Meier GVHD Survival Time
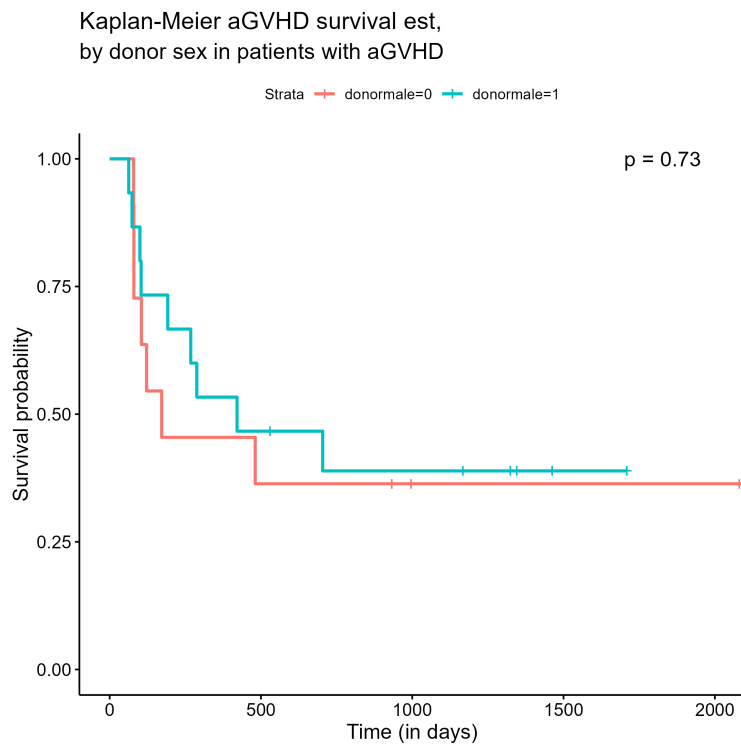
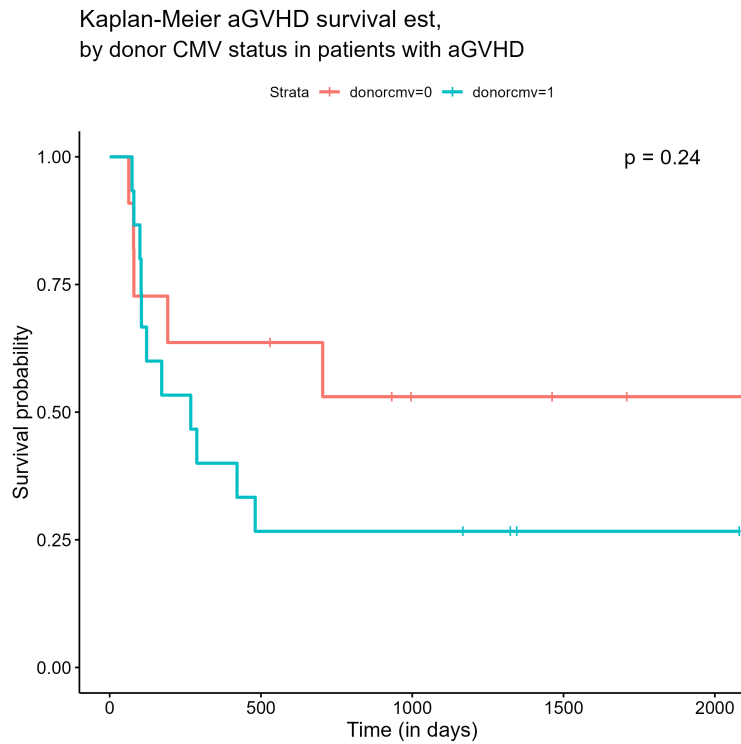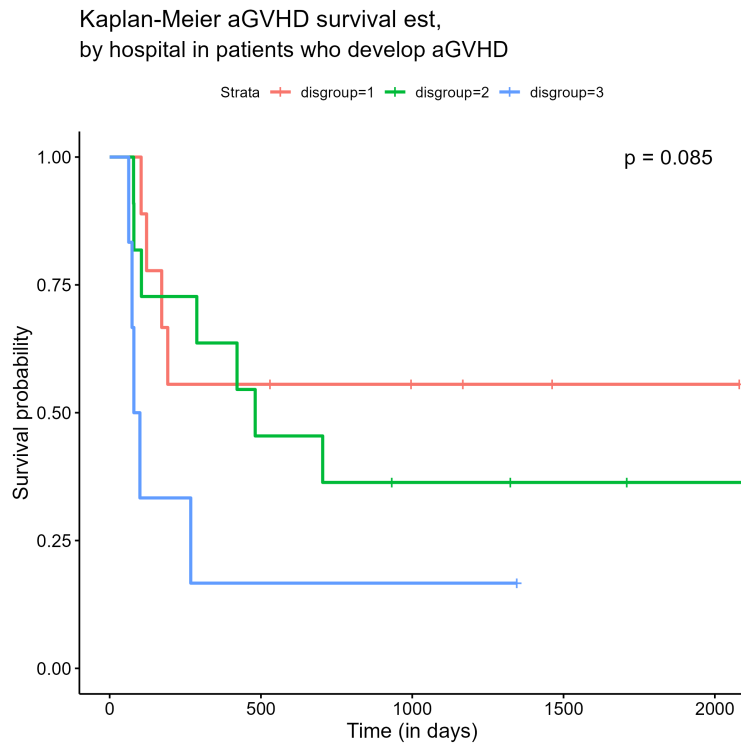Figure 12: Kaplan-Meier GVHD Survival Time

Kaplan-Meier disease-free survival est,
by patient CMV status in patients who develop aGVHD

Strata — cmv=0 — cmv=1

p = 0.64

Survival probability

Time (in days)
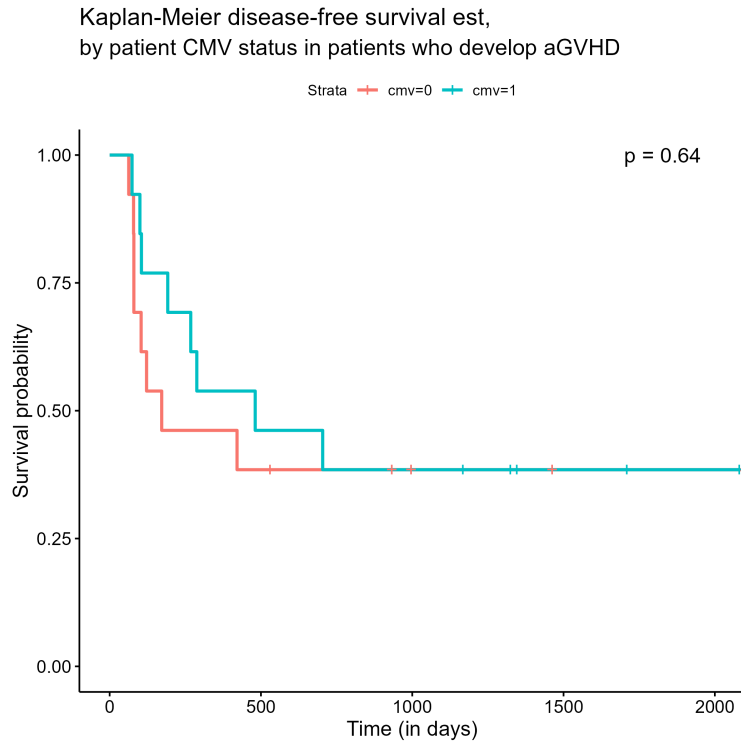
Figure 13: Kaplan-Meier Estimate, Disease Free Survival Time

Kaplan-Meier aGVHD survival est,
by donor CMV status in patients with aGVHD

Strata — donorcmv=0 — donorcmv=1

p = 0.24

Survival probability

Time (in days)

Figure 14: Kaplan-Meier GVHD Survival Time

18

Figure 15: Kaplan-Meier GVHD Survival Time
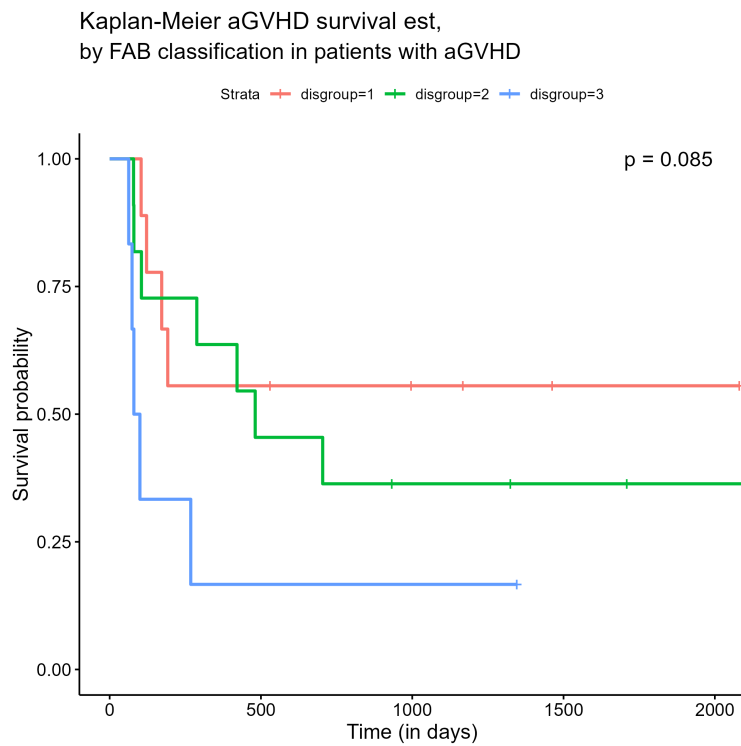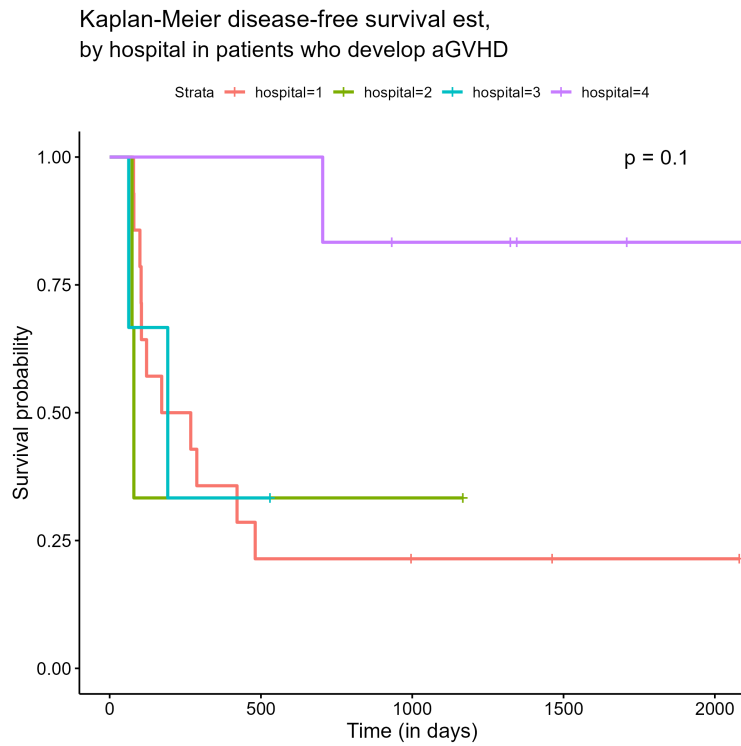

Figure 16: Kaplan-Meier GVHD Survival Time

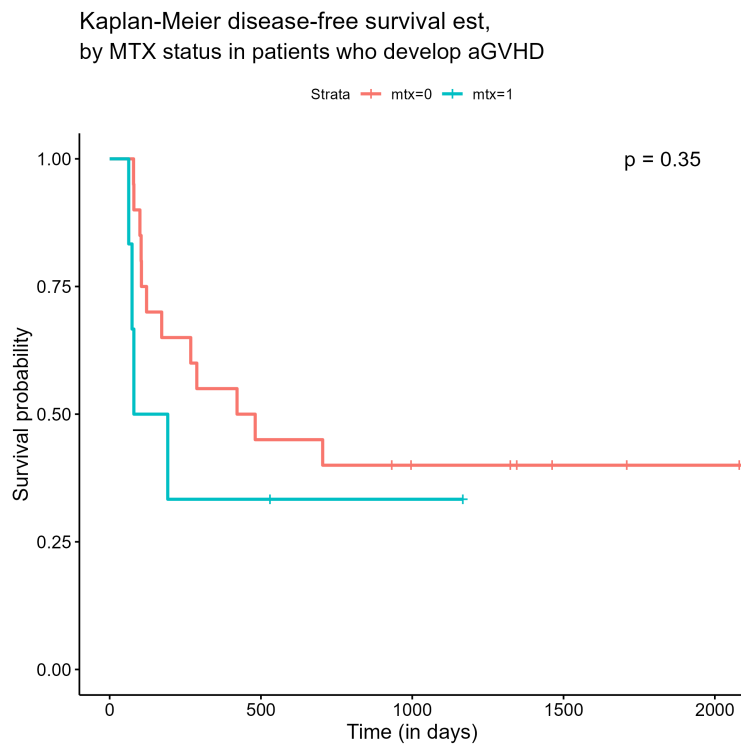Figure 17: Kaplan-Meier Estimate, Disease Free Survival Time

Figure 18: Kaplan-Meier Estimate, Disease Free Survival Time