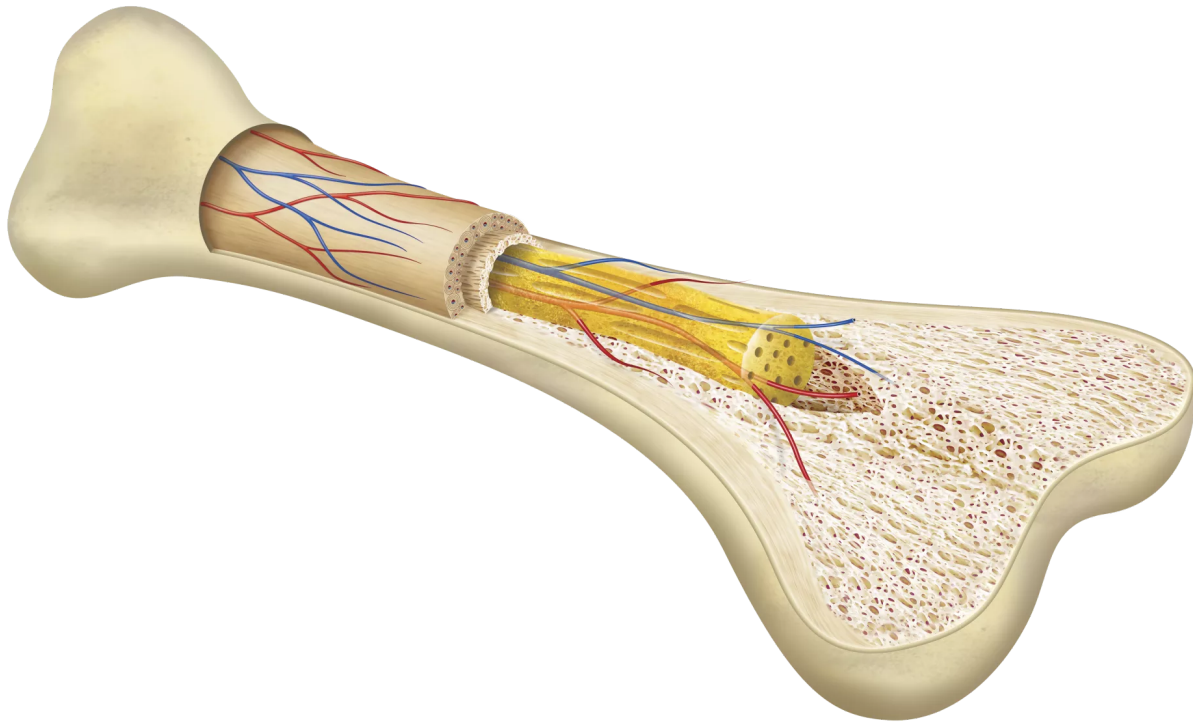


# **Bone Marrow Transplant in Children**



**STAT 495 Project**

**Survival Analysis Using KM Estimators, Log-Rank Tests, and Cox Proportional Hazards Model**

**Nicole Pham**

**May 3, 2022**

## Contents

<b>I. Introduction.....</b>	<b>2</b>
<b>II. Background.....</b>	<b>2</b>
<b>III. Data Description.....</b>	<b>2</b>
<b>IV. Results.....</b>	<b>2</b>
<b>V. Conclusion.....</b>	<b>3</b>
<b>VI. Appendix.....</b>	<b>4</b>
A. SAS Code.....	4
B. SAS Output.....	5
C. R Code.....	19
D. R Output.....	21

**VII. References.....**  
**.....29**

## I. Introduction

Bone marrow transplant is able to give terminally ill people a longer chance to live their lives. Anyone on the brink of death would take up the chance to live a little longer. After all, these people have nothing left to lose. This procedure could be lifesaving if the procedure goes well. Having lost a couple of family members to cancer, I cannot help but wonder if this procedure may have helped them.

## II. Background

This study follows pediatric patients with many hematologic diseases also known as blood disorders. Some patients have malignant disorders which are invasive and could be cancer causing while other patients have nonmalignant cases which are not as severe. Malignant disorders include different types of leukemia and myelodysplastic syndrome. Nonmalignant cases include different types of anemia and X-linked adrenoleukodystrophy.

## III. Data Description

The study was conducted in 2019 to find the most important factors that influence the success and failure of bone marrow transplantation in children. This data set contained 187 observations and 37 variables. However, there were some missing values. After cleaning up the data set, there were 142 observations left. This data set was obtained from UCI Learning Repository, which is a “collection of databases, domain theories, and data generators that are used by the machine learning community for the empirical analysis of machine learning algorithms.”

## IV. Results

A survival analysis was done on this data set using the Kaplan-Meier estimator and curve, the log-rank test, and the Cox proportional hazards model. With the log-rank test, being in the low risk group, not relapsing post treatment, not relapsing in general, and not developing extensive chronic graft vs. host disease are significant predictors of survival time at the 5% significance level. After modeling the Cox proportional hazards model, the fitted model is:

- $$S^{\wedge}(t) = [Sbar(t)] \exp(-0.09336(\text{bone marrow}) + 0.05447(\text{cd34gx10d6} - 12.609246479) - 0.02332(\text{cd3dcd34} - 5.0159072817) - 0.16746(\text{cd3dkgx10d8} - 4.8414084507) + 1.05643(\text{ALL disease}) + 1.63728(\text{chronic disease}) + 0.58095(\text{lymphoma disease}) + 1.54839(\text{nonmalignant disease}) - 0.04016(\text{antigen} + 0.063380282) + 0.09218(\text{alel} + 0.246478873) - 0.82448(\text{low risk}) - 0.49710(\text{no agvhdiiiv}) - 1.11888(\text{no extcgvhd}) + 2.13560(\text{relapse}) + 0.44568(\text{female}) + 0.07501(\text{donorage} - 32.931082373) - 0.86334(\text{donor age} \geq 35) - 0.50689(\text{gender}$$

matches) + 0.49341(donor A blood type) - 15.70981(donor AB blood type) +  
 0.29191(donor B blood type) - 0.20888(recipient A blood type) - 0.36243(recipient  
 AB blood type) + 0.05214(recipient B blood type) - 1.04493(rh is '-') -  
 0.13791(matched blood type) + 1.29634(hla 9/10) - 0.69751(hla 8/10) - 1.48325(1  
 antigen difference) - 2.22979(1 allele difference) - 0.25511(2 differences) +  
 0.02907(age - 9.4584507042) + 15.42561(age >= 10) - 0.01781(age 5-10) -  
 15.83636(age 10-20) + 0.02963(body mass - 33.927464789))

At the 5% significance level, chronic disease, extensive chronic GvHD, relapse, and having a  
 difference in 1 allele are significant predictors. The hazard for patients with chronic disease  
 is  $\exp(1.63728) * 100 = 514.11665\%$  of that for patients with acute myeloid  
 leukemia(AML). The hazard for patients that have not developed extensive chronic graft vs.  
 host disease is  $\exp(-1.11888) * 100 = 32.66454\%$  of that for patients that have developed  
 extensive chronic graft vs. host disease. The hazard for patients who have relapsed is  
 $\exp(2.13560) * 100 = 846.21223\%$  of that for patients that have not relapsed. The hazard  
 for patients that have a difference in 1 allele with their donor is  $\exp(-2.22979) * 100 =$   
 $10.75510\%$  of that for patients that have a match of allele and antigens with their donor.

At the 10% significance level, CD34+ cell dose per kg of recipient body weight at the  
 baseline, CD3+ cell dose per kg of recipient body weight at the baseline, nonmalignant  
 disease, low risk group, and '-' rh factor are significant predictors. As CD34+ cell dose per  
 kg of recipient body weight at the baseline increases by one unit, the hazard of dying from  
 bone marrow transplant increases by  $(\exp(0.05447) - 1) * 100 = 5.59808\%$ . As CD3+ cell  
 dose per kg of recipient body weight at the baseline increases by one unit, the hazard of  
 dying from bone marrow transplant changes by  $(\exp(-0.16746) - 1) * 100 = -15.41896\%$  or  
 decreases by 15.41896%. The hazard for patients who have nonmalignant disease is  
 $\exp(1.54839) * 100 = 470.38908\%$  of that for patients with acute myeloid leukemia(AML).  
 The hazard for patients that are in the low risk group is  $\exp(-0.82448) * 100 = 43.84629\%$   
 of that for patients that are in the high risk group. The hazard for patients with a '-' rh factor  
 is  $\exp(-1.04493) * 100 = 35.17164\%$  of that for patients with a '+' rh factor.

## V. Conclusion

The results show that having severe complications after the treatment and having any type  
 of relapse decrease the child's probability of surviving after the bone marrow transplant.  
 Increased dosage of CD3+ cells per kg of the patients body weight increases their chances  
 of survival as found in the study. Other factors that influence the success of the  
 transplantation are having matched or only one difference in allele and antigens, having a  
 negative rh factor, and being in the low risk group.

## VI. Appendix

## A. SAS Code

```

proc lifetest data=bonemarrowcleaned plots=(survival);
time survivalttime*survivalstatus(0); /*1 = dead, 0 = alive*/
strata extcgvhd; /* yes = 0, no = 1 */
format extcgvhd extcgvhdfmt.;
run;

proc phreg data=bonemarrowcleaned outest=betas;
class disease(ref = "AML")
      ageint(ref = "0") /* (0,5] = 0, (5, 10] = 1, (10, 20] = 2
*/
      abomatch(ref = "1") /* matched = 1, mismatched = 0 */
      hlamismatch(ref = "0") /* matched = 0, mismatched = 1 */
      riskgroup(ref = "1") /* high risk = 1, low risk = 0 */
      stemcellsource(ref = "1") /* peripheral blood = 1, bone
marrow = 0 */
      gender(ref = "1") /* male = 1, female = 0*/
      donorage35(ref = "0") /*0 = <35, 1 = >=35*/
      gendermatch(ref = "1") /* female to male = 1, other = 0 */
      donorabo(ref = "0") /* O = 0, A = 1, B = -1, AB = 2 */
      abo(ref = "0") /* O = 0, A = 1, B = -1, AB = 2 */
      rh(ref = "1") /* + = 1, - = 0 */
      hlamatch(ref = "0") /* 10/10 = 0, 9/10 = 1, 8/10 = 2, 7/10
= 3 (allele/antigens) */
      hlagri(ref = "0") /* hla matched = 0, diff is one antigen
= 1, diff is one allele = 2, diff is DRB1 cell = 3
                        2 differences = 4 */
      age10(ref = "0") /* age <10 = 0, age >=10 = 1 */
      relapse(ref = "0") /* no = 0, yes = 1*/
      agvhdiiiv(ref = "0") /* yes = 0, no = 1*/
      extcGvHD(ref = "0") /*yes = 0, no = 1*/;
model survivalttime*survivalstatus(0)= stemcellsource cd34gx10d6
cd3dcd34 cd3dkgx10d8
      disease diseasegroup antigen alel riskgroup
agvhdiiiv extcGvHD relapse gender donorage donorage35
      gendermatch donorabo abo rh abomatch hlamatch
hlamismatch
      hlagri age age10 ageint rbodymass;
baseline out=outdata survival=Sbar;
run;

proc print data=betas;
run;

```

```
proc print data=outdata;  
run;
```

## B. SAS Output

**The LIFETEST Procedure**

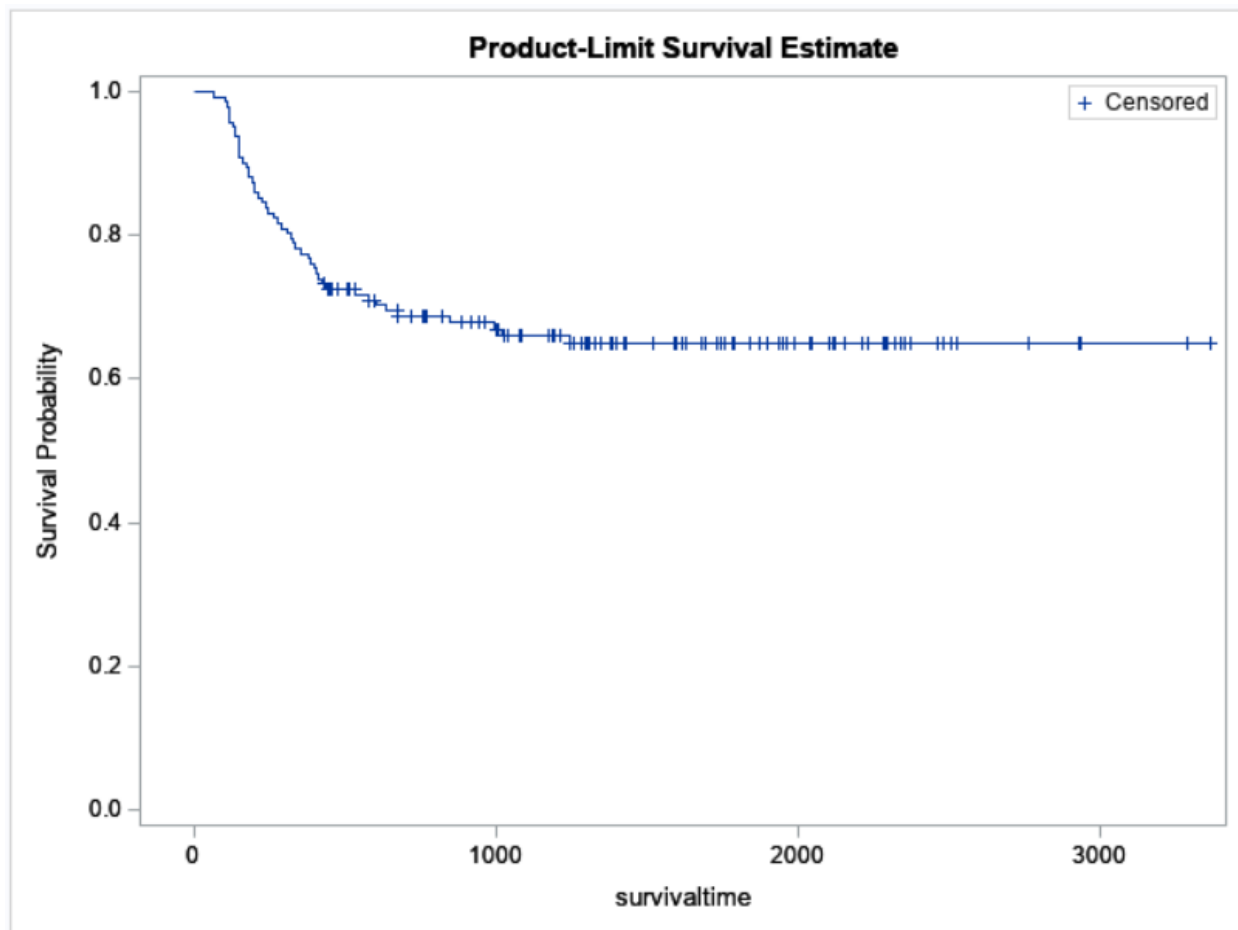
<b>Product-Limit Survival Estimates</b>					
<b>survivaltime</b>	<b>Survival</b>	<b>Failure</b>	<b>Survival Standard Error</b>	<b>Number Failed</b>	<b>Number Left</b>
0.00	1.0000	0	0	0	142
68.00	0.9930	0.00704	0.00702	1	141
102.00	0.9859	0.0141	0.00989	2	140
111.00	0.9789	0.0211	0.0121	3	139
113.00	0.9718	0.0282	0.0139	4	138
115.00	0.9648	0.0352	0.0155	5	137
117.00	0.9577	0.0423	0.0169	6	136
130.00	0.9507	0.0493	0.0182	7	135
136.00	0.9437	0.0563	0.0193	8	134
137.00	0.9366	0.0634	0.0204	9	133
147.00	0.9296	0.0704	0.0215	10	132
149.00	.	.	.	11	131
149.00	0.9155	0.0845	0.0233	12	130
150.00	0.9085	0.0915	0.0242	13	129
163.00	0.9014	0.0986	0.0250	14	128
174.00	0.8944	0.1056	0.0258	15	127
181.00	0.8873	0.1127	0.0265	16	126
182.00	0.8803	0.1197	0.0272	17	125
191.00	0.8732	0.1268	0.0279	18	124
200.00	0.8662	0.1338	0.0286	19	123
202.00	0.8592	0.1408	0.0292	20	122



214.00		0.8521	0.1479	0.0298	21	121
224.00		0.8451	0.1549	0.0304	22	120
236.00		0.8380	0.1620	0.0309	23	119
246.00		0.8310	0.1690	0.0314	24	118
261.00		0.8239	0.1761	0.0320	25	117
274.00		0.8169	0.1831	0.0325	26	116
290.00		0.8099	0.1901	0.0329	27	115
306.00		0.8028	0.1972	0.0334	28	114
321.00		0.7958	0.2042	0.0338	29	113
330.00		0.7887	0.2113	0.0343	30	112
335.00		0.7817	0.2183	0.0347	31	111
353.00		0.7746	0.2254	0.0351	32	110
382.00		0.7676	0.2324	0.0354	33	109
385.00		0.7606	0.2394	0.0358	34	108
397.00		0.7535	0.2465	0.0362	35	107
403.00		0.7465	0.2535	0.0365	36	106
413.00		0.7394	0.2606	0.0368	37	105
421.00		0.7324	0.2676	0.0372	38	104
433.00	*	.	.	.	38	103
435.00		0.7253	0.2747	0.0375	39	102
445.00	*	.	.	.	39	101
452.00	*	.	.	.	39	100
458.00	*	.	.	.	39	99
472.00	*	.	.	.	39	98

510.00	*	.	.	.	39	97
511.00	*	.	.	.	39	96
531.00	*	.	.	.	39	95
534.00		0.7176	0.2824	0.0378	40	94
576.00		0.7100	0.2900	0.0382	41	93
580.00	*	.	.	.	41	92
599.00	*	.	.	.	41	91
606.00		0.7022	0.2978	0.0386	42	90
634.00		0.6944	0.3056	0.0389	43	89
671.00	*	.	.	.	43	88
672.00		0.6865	0.3135	0.0393	44	87
676.00	*	.	.	.	44	86
718.00	*	.	.	.	44	85
754.00	*	.	.	.	44	84
766.00	*	.	.	.	44	83
767.00	*	.	.	.	44	82
824.00	*	.	.	.	44	81
849.00		0.6780	0.3220	0.0397	45	80
884.00	*	.	.	.	45	79
914.00	*	.	.	.	45	78
942.00	*	.	.	.	45	77
964.00	*	.	.	.	45	76
996.00		0.6691	0.3309	0.0402	46	75
999.00	*	.	.	.	46	74

1000.00	*	.	.	.	46	73
1006.00	*	.	.	.	46	72
1018.00		0.6598	0.3402	0.0407	47	71
1025.00	*	.	.	.	47	70
1041.00	*	.	.	.	47	69
1074.00	*	.	.	.	47	68
1084.00	*	.	.	.	47	67
1171.00	*	.	.	.	47	66
1185.00	*	.	.	.	47	65
1188.00	*	.	.	.	47	64
1195.00	*	.	.	.	47	63
1214.00	*	.	.	.	47	62
1243.00		0.6492	0.3508	0.0414	48	61
1243.00	*	.	.	.	48	60
1257.00	*	.	.	.	48	59
1280.00	*	.	.	.	48	58

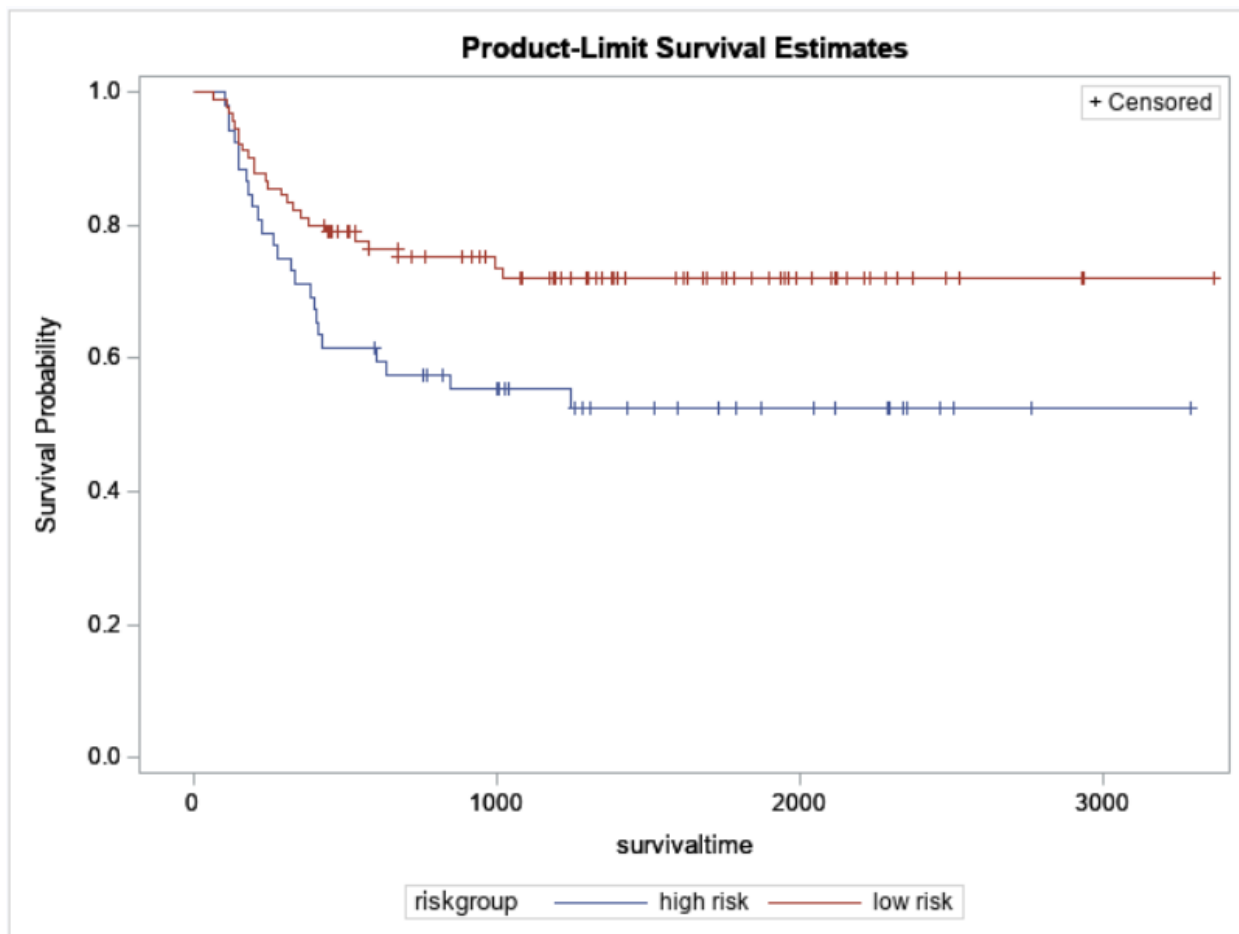


**Summary of the Number of Censored and Uncensored Values**

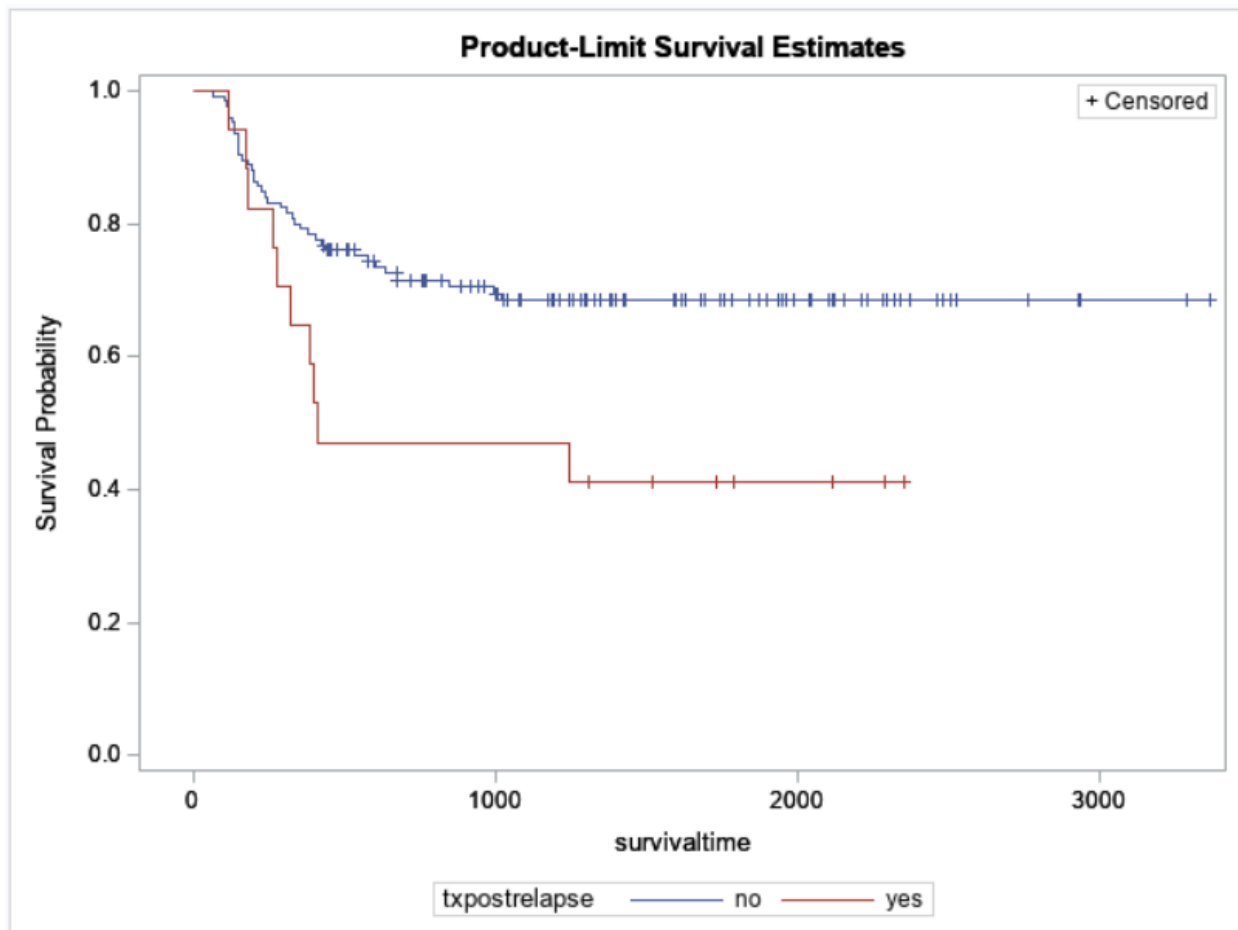
Total	Failed	Censored	Percent Censored
142	48	94	66.20

**Test of Equality over Strata**

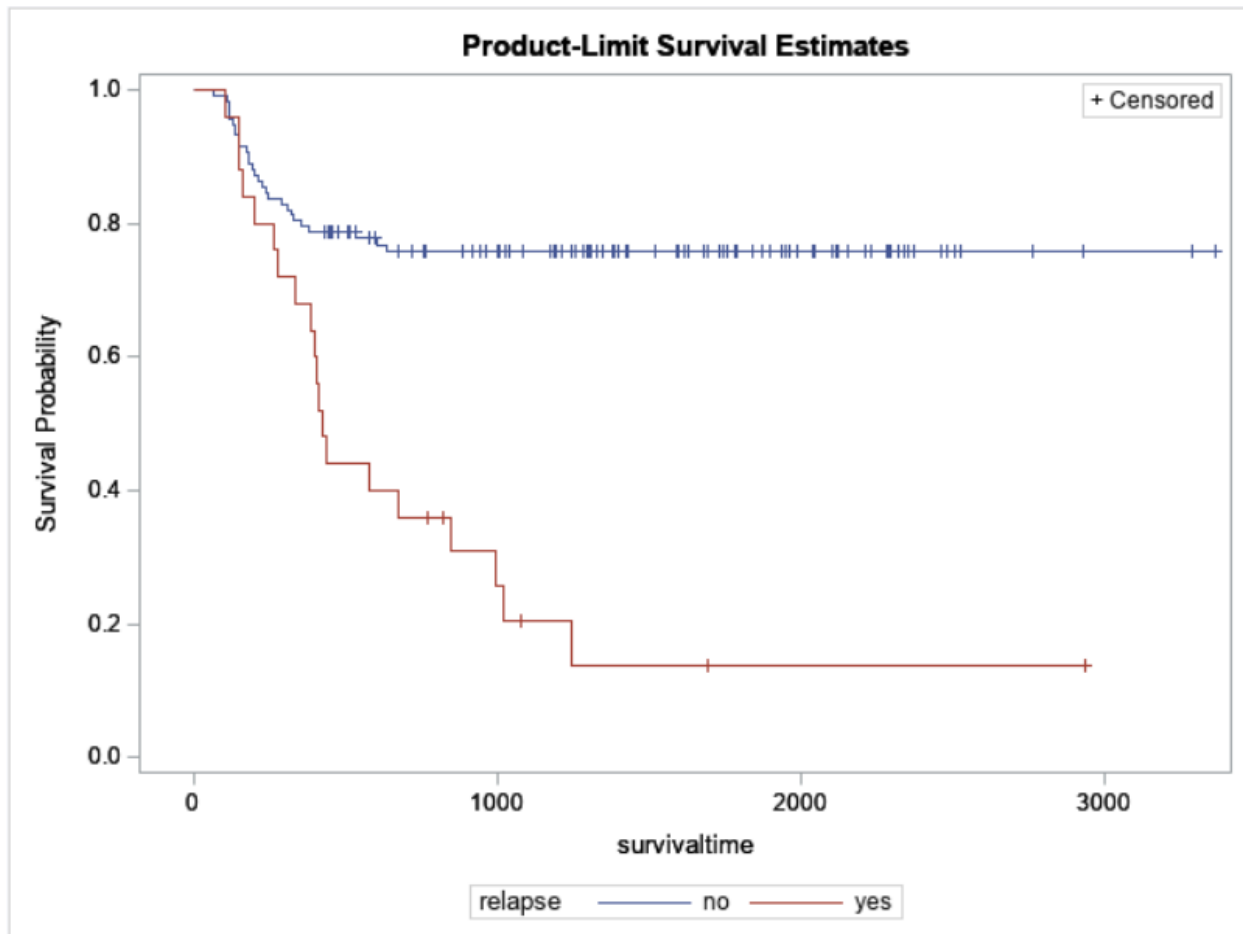
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.1834	1	0.0228
Wilcoxon	4.7290	1	0.0297
-2Log(LR)	5.4263	1	0.0198



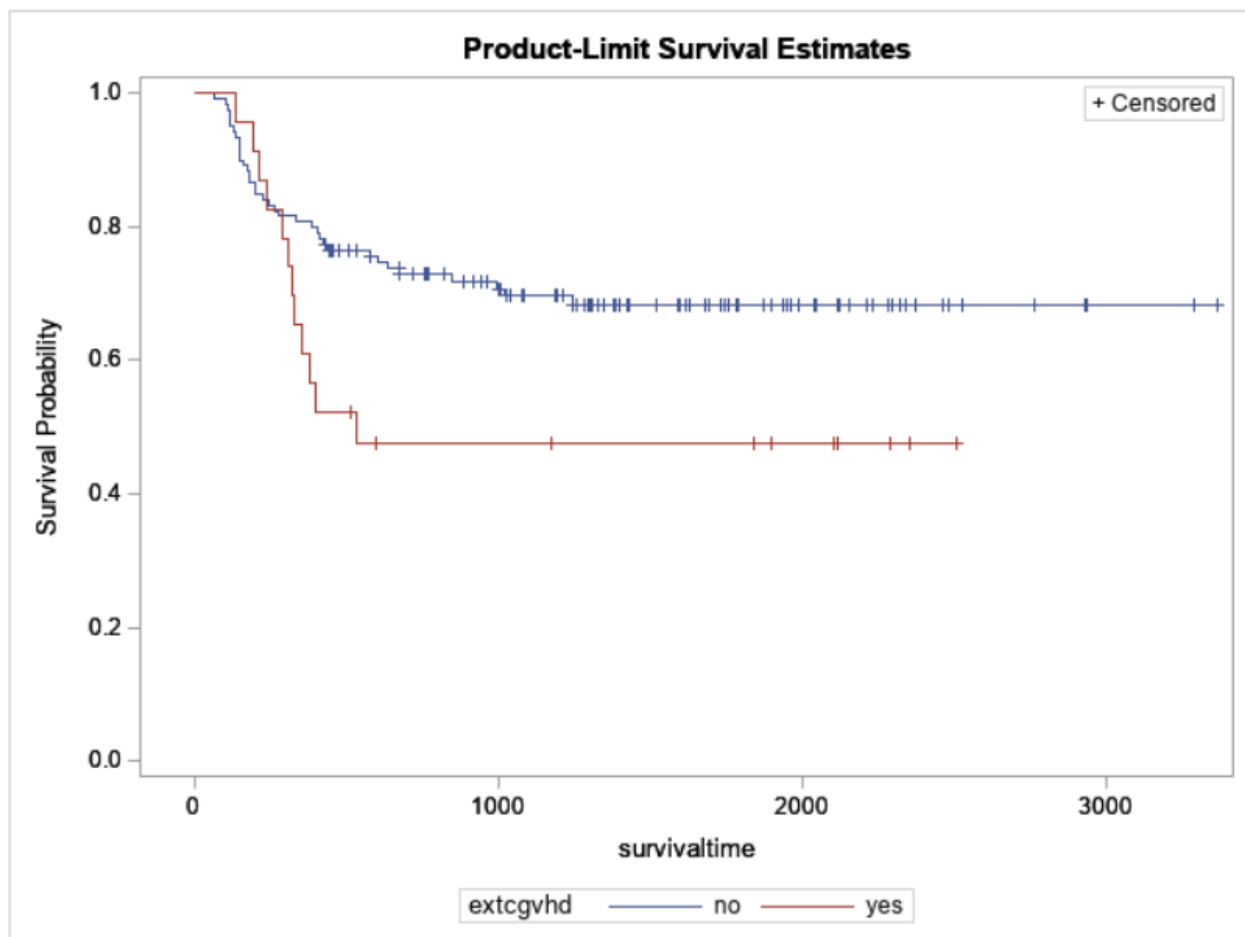
Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	5.0334	1	0.0249
Wilcoxon	4.2007	1	0.0404
-2Log(LR)	4.4766	1	0.0344



Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	26.9515	1	<.0001
Wilcoxon	18.0441	1	<.0001
-2Log(LR)	31.7265	1	<.0001



Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	4.0649	1	0.0438
Wilcoxon	3.5394	1	0.0599
-2Log(LR)	3.9442	1	0.0470





Analysis of Maximum Likelihood Estimates								
Parameter		DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	Label
stemcellsource	0	1	-0.09336	0.66949	0.0194	0.8891	0.911	stemcellsource 0
cd34gx10d6		1	0.05447	0.03305	2.7159	0.0994	1.056	
cd3dcd34		1	-0.02332	0.02443	0.9112	0.3398	0.977	
cd3dkgx10d8		1	-0.16746	0.09415	3.1637	0.0753	0.846	
disease	ALL	1	1.05643	0.78393	1.8160	0.1778	2.876	disease ALL
disease	chronic	1	1.63728	0.73425	4.9723	0.0258	5.141	disease chronic
disease	lymphoma	1	0.58095	1.06152	0.2995	0.5842	1.788	disease lymphoma
disease	nonmalignant	1	1.54839	0.91789	2.8456	0.0916	4.704	disease nonmalignant
diseasegroup		0	0	.	.	.	.	
antigen		1	-0.04016	0.59620	0.0045	0.9463	0.961	
alel		1	0.09218	0.36357	0.0643	0.7999	1.097	
riskgroup	0	1	-0.82448	0.48887	2.8443	0.0917	0.438	riskgroup 0
agvhdiiv	1	1	-0.49710	0.51422	0.9345	0.3337	0.608	agvhdiiv 1
extcgvhd	1	1	-1.11888	0.50328	4.9426	0.0262	0.327	extcgvhd 1
relapse	1	1	2.13560	0.53819	15.7456	<.0001	8.462	relapse 1
gender	0	1	0.44568	0.45850	0.9449	0.3310	1.562	gender 0
donorage		1	0.07501	0.05101	2.1623	0.1414	1.078	
donorage35	1	1	-0.86334	0.74378	1.3473	0.2457	0.422	donorage35 1
gendermatch	0	1	-0.50689	0.59964	0.7146	0.3979	0.602	gendermatch 0
donorabo	1	1	0.49341	0.45297	1.1865	0.2760	1.638	donorabo 1
donorabo	2	1	-15.70981	1126	0.0002	0.9889	0.000	donorabo 2
donorabo	-1	1	0.29191	0.51775	0.3179	0.5729	1.339	donorabo -1

abo	1	1	-0.20888	0.53769	0.1509	0.6977	0.811	abo 1
abo	2	1	-0.36243	0.85851	0.1782	0.6729	0.696	abo 2
abo	-1	1	0.05214	0.61807	0.0071	0.9328	1.054	abo -1
rh	0	1	-1.04493	0.61841	2.8551	0.0911	0.352	rh 0
abomatch	0	1	-0.13791	0.44330	0.0968	0.7557	0.871	abomatch 0
hلامatch	1	1	1.29634	1.08707	1.4221	0.2331	3.656	hلامatch 1
hلامatch	2	1	-0.69751	1.75948	0.1572	0.6918	0.498	hلامatch 2
hلامatch	3	0	0	.	.	.	.	hلامatch 3
hلامismatch	1	0	0	.	.	.	.	hلامismatch 1
hلامgri	1	1	-1.48325	0.95859	2.3942	0.1218	0.227	hلامgri 1
hلامgri	2	1	-2.22979	1.11836	3.9753	0.0462	0.108	hلامgri 2
hلامgri	3	0	0	.	.	.	.	hلامgri 3
hلامgri	4	1	-0.25511	1.45030	0.0309	0.8604	0.775	hلامgri 4
hلامgri	5	0	0	.	.	.	.	hلامgri 5
hلامgri	7	0	0	.	.	.	.	hلامgri 7
age		1	0.02907	0.14272	0.0415	0.8386	1.029	
age10	1	1	15.42561	4433	0.0000	0.9972	5003289	age10 1
ageint	1	1	-0.01781	0.72932	0.0006	0.9805	0.982	ageint 1
ageint	2	1	-15.83636	4433	0.0000	0.9971	0.000	ageint 2
rbodmass		1	0.02963	0.02800	1.1192	0.2901	1.030	

Obs	cd34gx10d6	cd3dcd34	cd3dkgx10d8	diseasegroup	antigen	alel	donorage	age	rbodmass	disease	ageint	abomatch	hلامismatch	riskgroup	stemcellsource	gender	donorage35
1	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
2	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
3	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
4	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
5	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
6	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
7	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
8	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
9	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
10	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
11	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
12	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
13	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
14	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
15	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
16	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
17	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
18	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
19	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
20	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0
21	12.609246479	5.0159072817	4.8414084507	0.823943662	-0.063380282	-0.246478873	32.931082373	9.4584507042	33.927464789	AML	0	1	0	1	1	1	0

gendermatch	donorabo	abo	rh	hلامatch	hlagri	age10	relapse	agvhdiiiv	extcgvhd	survivaltime	Sbar
1	0	0	1	0	0	0	0	0	0	0	1.00000
1	0	0	1	0	0	0	0	0	0	68	0.98580
1	0	0	1	0	0	0	0	0	0	102	0.97174
1	0	0	1	0	0	0	0	0	0	111	0.95671
1	0	0	1	0	0	0	0	0	0	113	0.94160
1	0	0	1	0	0	0	0	0	0	115	0.92610
1	0	0	1	0	0	0	0	0	0	117	0.91071
1	0	0	1	0	0	0	0	0	0	130	0.89504
1	0	0	1	0	0	0	0	0	0	136	0.87946
1	0	0	1	0	0	0	0	0	0	137	0.86408
1	0	0	1	0	0	0	0	0	0	147	0.84880
1	0	0	1	0	0	0	0	0	0	149	0.81798
1	0	0	1	0	0	0	0	0	0	150	0.80182
1	0	0	1	0	0	0	0	0	0	163	0.78582
1	0	0	1	0	0	0	0	0	0	174	0.77000
1	0	0	1	0	0	0	0	0	0	181	0.75445
1	0	0	1	0	0	0	0	0	0	182	0.73900
1	0	0	1	0	0	0	0	0	0	191	0.72385
1	0	0	1	0	0	0	0	0	0	200	0.70866
1	0	0	1	0	0	0	0	0	0	202	0.69366
1	0	0	1	0	0	0	0	0	0	214	0.67888
1	0	0	1	0	0	0	0	0	0	224	0.66373
1	0	0	1	0	0	0	0	0	0	236	0.64868

246	0.63392
261	0.61945
274	0.60446
290	0.58920
306	0.57426
321	0.55968
330	0.54546
335	0.53141
353	0.51730
382	0.50350
385	0.48949
397	0.47559
403	0.46101
413	0.44582
421	0.43020
435	0.41459
534	0.39887
576	0.38255
606	0.36509
634	0.34815
672	0.33192
849	0.31494
996	0.29752
1018	0.27889
1243	0.25831

### C. R Code

```

bonemarrow.data = read.csv("C:/Users/nicol/Documents/STAT/STAT
495/PROJECT/bonemarrowcleaned.csv",
                          header = TRUE, sep = ",")

library(survival)
#fitting overall KM survival curve
overall.surv<-survfit(Surv(survivaltime, survivalstatus==1) ~ 0,
                    data=bonemarrow.data, se.fit=FALSE)
summary(overall.surv)

plot(overall.surv, mark.time=TRUE, pch=4, col="blue",
     main="Overall KM Survival Curve", xlab="Days",
     ylab="Survival Distribution Function")

#KM curves and log-rank test by risk group
riskgroup.surv<-survfit(Surv(survivaltime, survivalstatus==1) ~
riskgroup,
                      data=bonemarrow.data, se.fit=FALSE)
survdiff(Surv(survivaltime, survivalstatus==1) ~ riskgroup,
         data=bonemarrow.data)

plot(riskgroup.surv, mark.time=TRUE, pch=4, col=c("blue","red"),
     main="KM Survival Curves by Risk Group", xlab="Days",
     ylab="Survival
Distribution Function")
legend("bottomleft", lty=1, col=c("blue","red"), legend=c("low
risk","high risk"), text.col=c("blue","red"))

#KM curves and log-rank test by tx post relapse
txpostrelapse.surv<-survfit(Surv(survivaltime, survivalstatus==1) ~
txpostrelapse,
                          data=bonemarrow.data, se.fit=FALSE)
survdiff(Surv(survivaltime, survivalstatus==1) ~ txpostrelapse,
         data=bonemarrow.data)

plot(txpostrelapse.surv, mark.time=TRUE, pch=4, col=c("blue","red"),
     main="KM Survival Curves by Treatment Post Relapse",
     xlab="Days", ylab="Survival
Distribution Function")
legend("bottomleft", lty=1, col=c("blue","red"),
     legend=c("no","yes"), text.col=c("blue", "red"))

#KM curves and log-rank test by relapse

```

```

relapse.surv<-survfit(Surv(survivaltime, survivalstatus==1) ~
relapse,
                    data=bonemarrow.data, se.fit=FALSE)
survdifff(Surv(survivaltime, survivalstatus==1) ~ relapse,
          data=bonemarrow.data)

plot(relapse.surv, mark.time=TRUE, pch=4, col=c("blue","red"),
     main="KM Survival Curves by Relapse", xlab="Days",
     ylab="Survival
Distribution Function")
legend("bottomleft", lty=1, col=c("blue","red"),
legend=c("no","yes"), text.col=c("blue", "red"))

#KM curves and log-rank test by extcgvhd
extcgvhd.surv<-survfit(Surv(survivaltime, survivalstatus==1) ~
extcgvhd,
                    data=bonemarrow.data, se.fit=FALSE)
survdifff(Surv(survivaltime, survivalstatus==1) ~ extcgvhd,
          data=bonemarrow.data)

plot(extcgvhd.surv, mark.time=TRUE, pch=4, col=c("blue","red"),
     main="KM Survival Curves by Development of extcGvHD",
     xlab="Days", ylab="Survival
Distribution Function")
legend("bottomleft", lty=1, col=c("blue","red"),
legend=c("yes","no"), text.col=c("blue", "red"))

disease.rel<- relevel(as.factor(bonemarrow.data$disease), ref =
"AML")
ageint.rel<- relevel(as.factor(bonemarrow.data$ageint), ref = "0")
abomatch.rel<- relevel(as.factor(bonemarrow.data$abomatch), ref =
"1")
hlamismatch.rel<- relevel(as.factor(bonemarrow.data$hlamismatch), ref
= "0")
riskgroup.rel<- relevel(as.factor(bonemarrow.data$riskgroup), ref =
"1")
stemcellsource.rel<-
relevel(as.factor(bonemarrow.data$stemcellsource), ref = "1")
gender.rel<- relevel(as.factor(bonemarrow.data$i.gender), ref = "1")
donorage35.rel<- relevel(as.factor(bonemarrow.data$donorage35), ref =
"0")
gendermatch.rel<- relevel(as.factor(bonemarrow.data$gendermatch), ref
= "1")
donorabo.rel<- relevel(as.factor(bonemarrow.data$donorabo), ref =
"0")

```

```

abo.rel<- relevel(as.factor(bonemarrow.data$abo), ref = "0")
rh.rel<- relevel(as.factor(bonemarrow.data$rh), ref = "1")
hلامatch.rel<- relevel(as.factor(bonemarrow.data$hلامatch), ref =
"0")
hلامgri.rel<- relevel(as.factor(bonemarrow.data$hلامgri), ref = "0")
age10.rel<- relevel(as.factor(bonemarrow.data$age10), ref = "0")
relapse.rel<- relevel(as.factor(bonemarrow.data$relapse), ref = "0")
agvhdiiiiiv.rel<- relevel(as.factor(bonemarrow.data$agvhdiiiiiv), ref =
"0")
extcgvhd.rel<- relevel(as.factor(bonemarrow.data$extcgvhd), ref =
"0")

#estimating the beta coefficients
cox.model<-coxph(Surv(survivaltime, survivalstatus==1) ~
stemcellsource.rel +
                    cd34gx10d6 + cd3dcd34 + cd3dkgx10d8 + disease.rel
+
                    diseasegroup + antigen + alel + riskgroup.rel +
agvhdiiiiiv.rel +
                    extcgvhd.rel + relapse.rel + gender.rel + donorage
+ donorage35.rel +
                    gendermatch.rel + donorabo.rel + abo.rel + rh.rel
+ abomatch.rel + hلامatch.rel +
                    hلامismatch.rel + hلامgri.rel + age + age10.rel +
ageint.rel + rbodymass,
                    data=bonemarrow.data)
summary(cox.model)

#estimating the baseline survival function
base.surv<-survfit(cox.model, se.fit=FALSE)
summary(base.surv)

mean(bonemarrow.data$cd34gx10d6)
mean(bonemarrow.data$cd3dcd34)
mean(bonemarrow.data$cd3dkgx10d8)
mean(bonemarrow.data$antigen)
mean(bonemarrow.data$alel)
mean(bonemarrow.data$donorage)
mean(bonemarrow.data$age)
mean(bonemarrow.data$rbodymass)

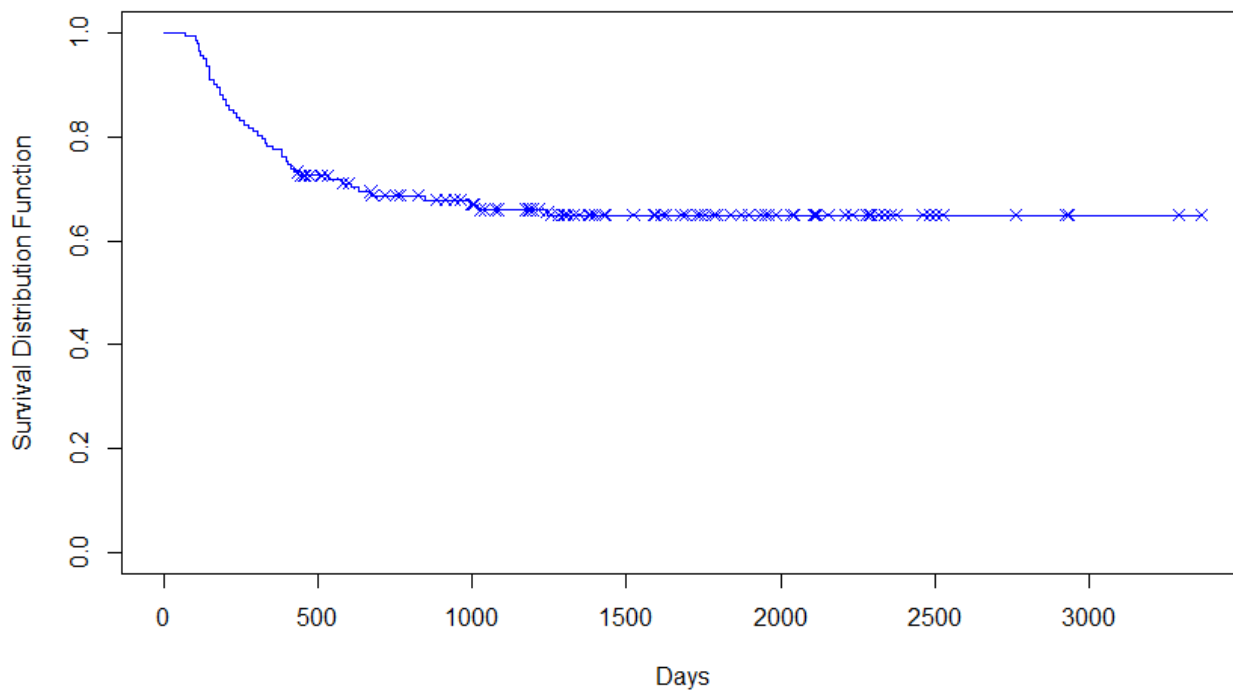
```

#### D. R Output

time	n.risk	n.event	survival
68	142	1	0.993
102	141	1	0.986
111	140	1	0.979
113	139	1	0.972
115	138	1	0.965
117	137	1	0.958
130	136	1	0.951
136	135	1	0.944
137	134	1	0.937
147	133	1	0.930
149	132	2	0.915
150	130	1	0.908
163	129	1	0.901
174	128	1	0.894
181	127	1	0.887
182	126	1	0.880
191	125	1	0.873
200	124	1	0.866
202	123	1	0.859
214	122	1	0.852
224	121	1	0.845
236	120	1	0.838
246	119	1	0.831
261	118	1	0.824
274	117	1	0.817
290	116	1	0.810
306	115	1	0.803
321	114	1	0.796
330	113	1	0.789
335	112	1	0.782
353	111	1	0.775
382	110	1	0.768
385	109	1	0.761
397	108	1	0.754
403	107	1	0.746
413	106	1	0.739
421	105	1	0.732
435	103	1	0.725
534	95	1	0.718
576	94	1	0.710
606	91	1	0.702
634	90	1	0.694
672	88	1	0.687
849	81	1	0.678
996	76	1	0.669
1018	72	1	0.660
1243	62	1	0.649

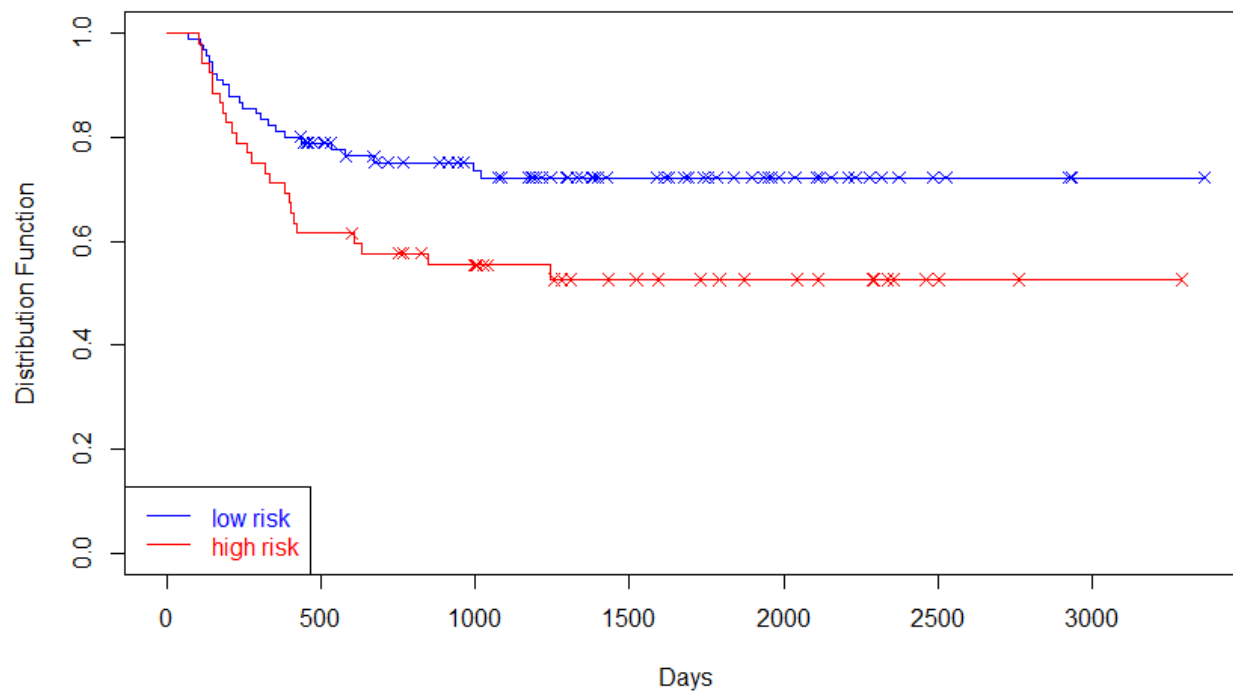


Overall KM Survival Curve



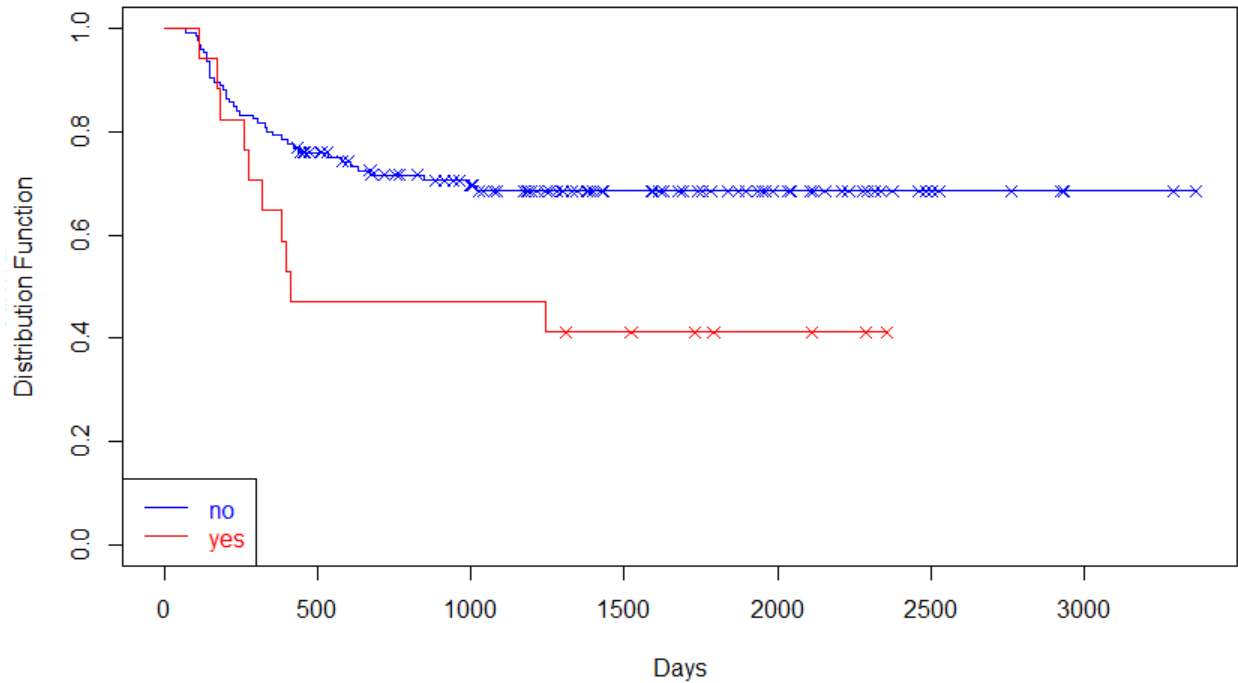
Chisq= 5.2 on 1 degrees of freedom, p= 0.02

KM Survival Curves by Risk Group



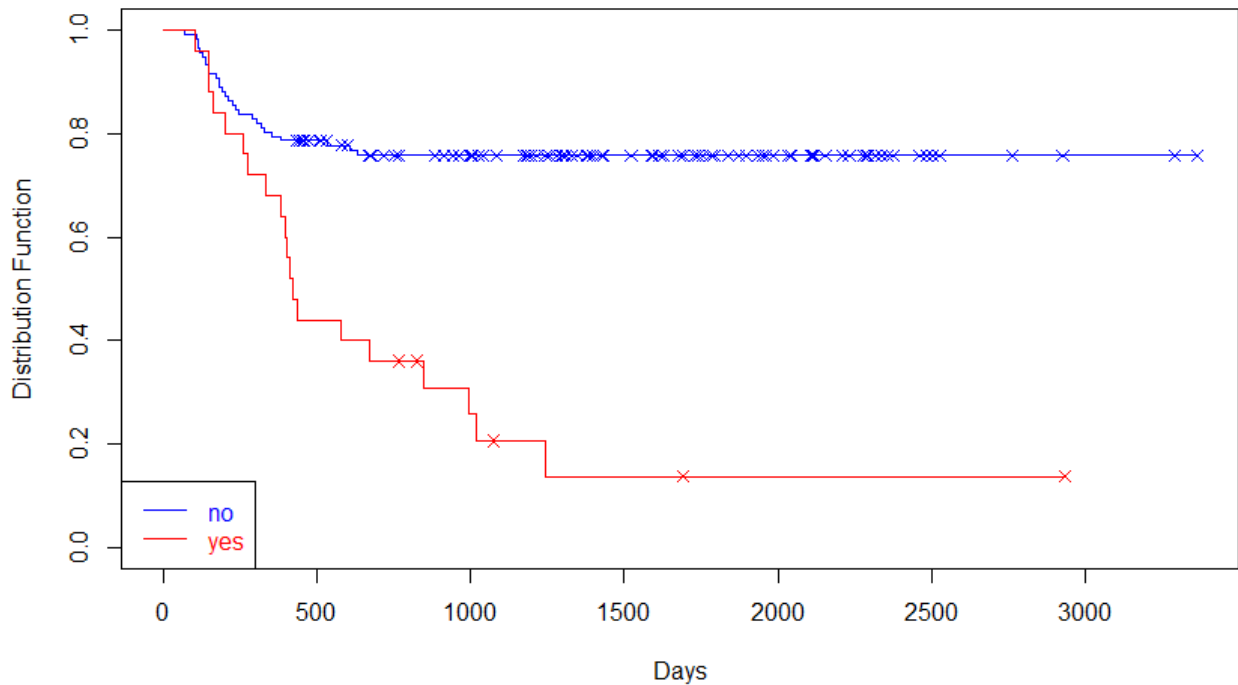
Chisq= 5 on 1 degrees of freedom, p= 0.02

**KM Survival Curves by Treatment Post Relapse**



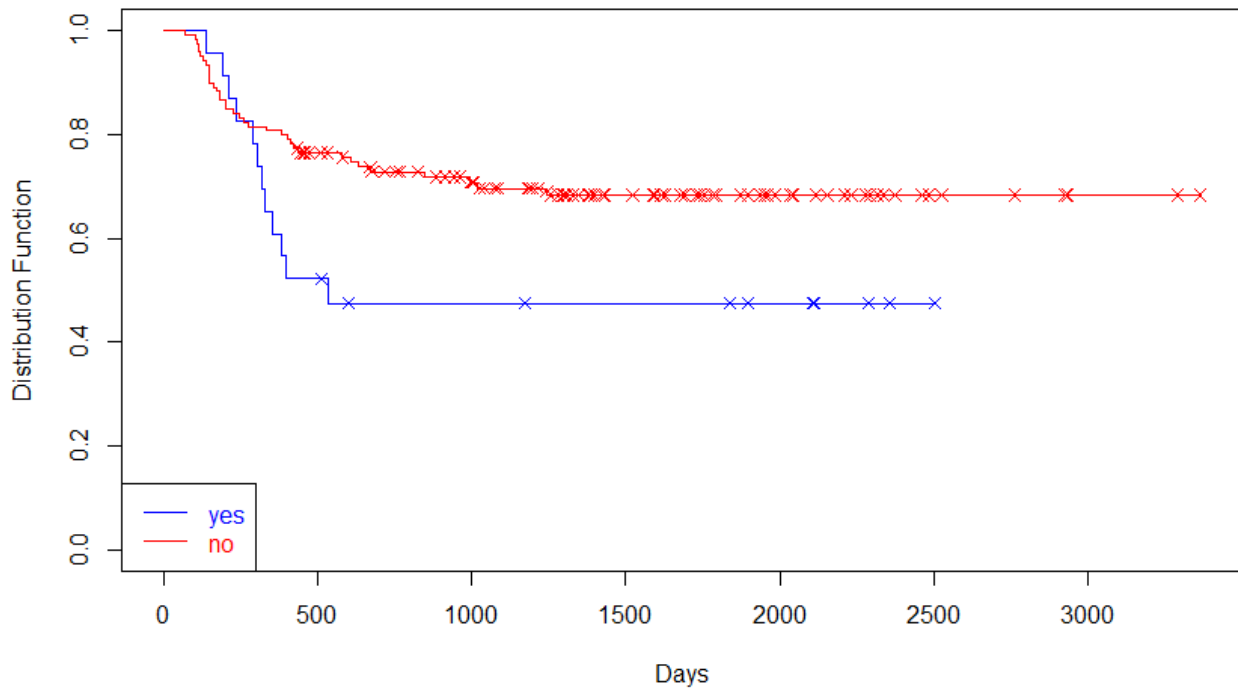
Chisq= 27 on 1 degrees of freedom,  $p= 2e-07$

**KM Survival Curves by Relapse**



Chisq= 4.1 on 1 degrees of freedom,  $p= 0.04$

### KM Survival Curves by Development of extcGvHD



coef	exp(coef)	se(coef)	z	Pr(> z )
stemcellsource.rel0	0.8881	-9.427e-02	9.100e-01	6.699e-01 -0.141
cd34gx10d6	0.0986 .	5.457e-02	1.056e+00	3.304e-02 1.652
cd3dcd34	0.3372	-2.348e-02	9.768e-01	2.447e-02 -0.960
cd3dkgx10d8	0.0749 .	-1.677e-01	8.456e-01	9.417e-02 -1.781
disease.relALL	0.1774	1.058e+00	2.882e+00	7.847e-01 1.349
disease.relchronic	0.0255 *	1.642e+00	5.165e+00	7.351e-01 2.234
disease.relymphoma	0.5854	5.793e-01	1.785e+00	1.062e+00 0.546
disease.relnonmalignant	0.0910 .	1.552e+00	4.720e+00	9.182e-01 1.690
diseasegroup	NA	NA	NA	0.000e+00 NA
antigen	0.9486	-3.850e-02	9.622e-01	5.968e-01 -0.065
alel	0.8028	9.083e-02	1.095e+00	3.636e-01 0.250

riskgroup.rel0 0.0913 .	-8.258e-01	4.379e-01	4.891e-01	-1.688
agvhdiiiiiv.rel1 0.3316	-4.995e-01	6.068e-01	5.146e-01	-0.971
extcgvhd.rel1 0.0264 *	-1.118e+00	3.269e-01	5.034e-01	-2.221
relapse.rel1 7.17e-05 ***	2.140e+00	8.500e+00	5.390e-01	3.971
gender.rel0 0.3286	4.484e-01	1.566e+00	4.589e-01	0.977
donorage 0.1399	7.535e-02	1.078e+00	5.105e-02	1.476
donorage35.rel1 0.2442	-8.665e-01	4.204e-01	7.441e-01	-1.164
gendermatch.rel0 0.4008	-5.040e-01	6.041e-01	5.999e-01	-0.840
donorabo.rel-1 0.5736	2.914e-01	1.338e+00	5.177e-01	0.563
donorabo.rel1 0.2780	4.917e-01	1.635e+00	4.533e-01	1.085
donorabo.rel2 0.9954	-1.772e+01	2.023e-08	3.061e+03	-0.006
abo.rel-1 0.9274	5.640e-02	1.058e+00	6.187e-01	0.091
abo.rel1 0.6976	-2.090e-01	8.114e-01	5.380e-01	-0.389
abo.rel2 0.6731	-3.625e-01	6.959e-01	8.592e-01	-0.422
rh.rel0 0.0904 .	-1.048e+00	3.507e-01	6.188e-01	-1.693
abomatch.rel0 0.7524	-1.399e-01	8.694e-01	4.434e-01	-0.315
hlamatch.rel1 0.2308	1.304e+00	3.682e+00	1.088e+00	1.198
hlamatch.rel2 0.6919	-6.975e-01	4.978e-01	1.760e+00	-0.396
hlamatch.rel3 NA	NA	NA	0.000e+00	NA
hlamismatch.rel1 NA	NA	NA	0.000e+00	NA
hlagri.rel1 0.1204	-1.491e+00	2.251e-01	9.600e-01	-1.553
hlagri.rel2 0.0454 *	-2.241e+00	1.063e-01	1.120e+00	-2.001

hlagri.rel3	NA	NA	0.000e+00	NA
NA				
hlagri.rel4	-2.580e-01	7.726e-01	1.451e+00	-0.178
0.8588				
hlagri.rel5	NA	NA	0.000e+00	NA
NA				
hlagri.rel7	NA	NA	0.000e+00	NA
NA				
age	2.916e-02	1.030e+00	1.427e-01	0.204
0.8381				
age10.rel1	1.743e+01	3.704e+07	1.206e+04	0.001
0.9988				
ageint.rel1	-2.072e-02	9.795e-01	7.298e-01	-0.028
0.9773				
ageint.rel2	-1.784e+01	1.791e-08	1.206e+04	-0.001
0.9988				
rbodymass	2.961e-02	1.030e+00	2.799e-02	1.058
0.2901				
---				

time	n.risk	n.event	survival
68	142	1	0.986
102	141	1	0.972
111	140	1	0.957
113	139	1	0.942
115	138	1	0.926
117	137	1	0.911
130	136	1	0.895
136	135	1	0.880
137	134	1	0.865
147	133	1	0.849
149	132	2	0.818
150	130	1	0.802
163	129	1	0.786
174	128	1	0.770
181	127	1	0.755
182	126	1	0.739
191	125	1	0.724
200	124	1	0.709
202	123	1	0.694
214	122	1	0.679
224	121	1	0.664
236	120	1	0.649
246	119	1	0.634
261	118	1	0.620

274	117	1	0.605
290	116	1	0.590
306	115	1	0.575
321	114	1	0.560
330	113	1	0.546
335	112	1	0.532
353	111	1	0.518
382	110	1	0.504
385	109	1	0.490
397	108	1	0.476
403	107	1	0.462
413	106	1	0.446
421	105	1	0.431
435	103	1	0.415
534	95	1	0.399
576	94	1	0.383
606	91	1	0.366
634	90	1	0.349
672	88	1	0.333
849	81	1	0.316
996	76	1	0.298
1018	72	1	0.280
1243	62	1	0.259

```

> mean(bonemarrow.data$cd34gx10d6)
[1] 12.60925
> mean(bonemarrow.data$cd3dcd34)
[1] 5.015907
> mean(bonemarrow.data$cd3dkgx10d8)
[1] 4.841408
> mean(bonemarrow.data$antigen)
[1] -0.06338028
> mean(bonemarrow.data$alel)
[1] -0.2464789
> mean(bonemarrow.data$donorage)
[1] 32.93108
> mean(bonemarrow.data$age)
[1] 9.458451
> mean(bonemarrow.data$rbodmass)
[1] 33.92746

```

## VII. References

1. <https://archive.ics.uci.edu/ml/datasets/Bone+marrow+transplant%3A+children>