P8108 Group 2 Survival Analysis Project

Yiming Zhao (yz3955) Wenshan Qu (wq2160) Tucker Morgan (tlm2152)
Junzhe Shao (js5959) Benjamin Goebel (bpg2118)

2022-12-18

Introduction

Breast cancer is a leading cause of cancer-related morbidity and mortality worldwide (Hortobagyi et al.), and the choice of treatment can have a significant impact on patient outcomes. Hormone therapy and chemotherapy are two common treatment options for breast cancer, but their effectiveness in improving survival may vary depending on the specific characteristics of the patient and the tumor and there is plenty of literature (Wilcken et al.) comparing their effectiveness.

In this study, we aim to evaluate the effectiveness of hormone therapy and chemotherapy in improving breast cancer survival for our target population using the data from the Rotterdam tumor bank. Specifically, we will examine the treatment effect and taking into account various patient and tumor characteristics that may be the confounders of the treatment.

In addition to evaluating the effectiveness of different treatments, we will also compare the performance of non-parametric modeling approaches, such as random forests, to that of semi-parametric approaches, such as the Cox proportional hazards model, in predicting breast cancer survival. Understanding the performance of these different modeling approaches can inform the selection of the most appropriate method for predicting breast cancer survival in different settings.

Exploratory Data Analysis

Table 1: Distribution of covariates between hormone therapy treatment groups

	0 (N=2643)	1 (N=339)	Total (N=2982)	p value
age				< 0.001
Mean	54.098	62.549	55.058	
meno				< 0.001
Mean	0.519	0.879	0.560	
size				< 0.001
<=20	1283 (48.5%)	104 (30.7%)	1387 (46.5%)	
20-50	1119 (42.3%)	172 (50.7%)	1291 (43.3%)	
>50	241 (9.1%)	63 (18.6%)	304 (10.2%)	
grade				< 0.001
Mean	2.722	2.826	2.734	
nodes				< 0.001
Mean	2.327	5.720	2.712	
pgr				< 0.001
Mean	168.706	108.233	161.831	
er				0.069
Mean	164.792	180.608	166.590	
chemo				< 0.001
0	2091 (79.1%)	311 (91.7%)	2402 (80.5%)	
1	552 (20.9%)	28 (8.3%)	580 (19.5%)	

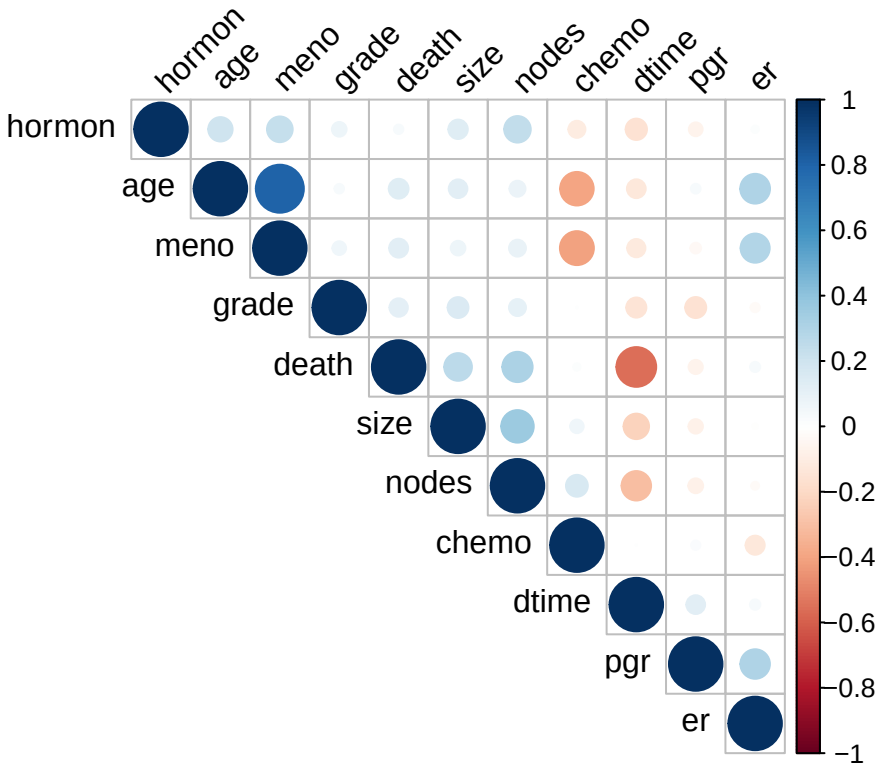
We worked on exploratory data analysis of candidate prognostic variables in the breast cancer datasets. Comparing demographic information of patients who accepted hormone therapy to those who did not, those show significant differences in age at surgery, tumor size, tumor grade, number of positive lymph nodes, menopausal status, progesterone receptors, and estrogen receptors. So two groups (received hormone therapy vs did not receive hormone therapy) are clinically different. Covariates were unbalanced.

Table 2: Distribution of covariates between chemotherapy treatment groups

	0 (N=2402)	1 (N=580)	Total (N=2982)	p value
age				< 0.001
Mean	57.560	44.698	55.058	
meno				< 0.001
Mean	0.658	0.153	0.560	
size				0.002
<=20	1148 (47.8%)	239 (41.2%)	1387 (46.5%)	
20-50	1028 (42.8%)	263 (45.3%)	1291 (43.3%)	
>50	226 (9.4%)	78 (13.4%)	304 (10.2%)	
grade				0.964
Mean	2.734	2.734	2.734	
nodes				< 0.001
Mean	2.353	4.198	2.712	
pgr				0.002
Mean	157.556	179.536	161.831	
er				< 0.001
Mean	183.599	96.148	166.590	
hormon				< 0.001
0	2091 (87.1%)	552 (95.2%)	2643 (88.6%)	
1	311 (12.9%)	28 (4.8%)	339 (11.4%)	

We got similar results by comparing patients who accepted chemotherapy to patients who did not. Comparing demographic information of patients who received hormone therapy to those who did not, those show significant differences in age at surgery, tumor size, number of positive lymph nodes, menopausal status, progesterone receptors and estrogen receptors.Two groups(received chemotherapy vs did not receive chemotherapy) are clinically different. Covariates were unbalanced.

Figure 1: Correlation of variables in Rotterdam data set



As a correlation analysis was conducted, the correlation matrix indicates that pre-menopausal women and young women are moderately correlated to taking chemotherapy. A highly positive correlation between menopausal status and age, and two moderately negative correlations between age and chemotherapy, menopausal status and chemotherapy correspond to clinical practice.

Methods

The dataset of interest for this analysis comes from the Rotterdam tumor bank, including data from 2982 breast cancer patients. Follow up time for patients varied from just 1 month to as long as 231 months. Several prognostic variables are recorded including year of surgery, age at surgery, menopausal status (pre- or post-), tumor size (mm), differentiation grade, number of positive lymph nodes, progesterone receptors (fmol/l), estrogen receptors (fmol/l), and indicators for hormonal treatment and chemotherapy treatment. The outcome considered in this analysis was patient death. The censoring mechanism is right censoring, which we assume to be non-informative.

Log-rank test

To investigate our research questions, we first use a non-parametric log-rank test, which makes no distributional assumptions about our data. The log-rank test compares differences between expected and observed events at each event time point, k, to derive the following test statistics:

$$L = \sum_{i=1}^{k} (d_{0i} - e_{0i});$$

$$var(L) = \sum_{i=1}^{k} \frac{n_{0i} n_{1i} d_i (n_i - d_i)}{n_i^2 (n_i - 1)};$$

where $L/\sqrt{var(L)} \sim N(0,1)$. A significant result from the log-rank test indicates a statistically significant difference between the two groups. Notably, the log-rank test does not allow us to adjust for covariates or prognostic variables.

Regularized Cox Proportional Hazard Model

As a further step in this analysis, we consider the Cox Proportional Hazard (Cox PH) model, which allows us to model the hazard ratio based on covariates to understand their impact on the survival function. The Cox PH typically takes the form:

$$h(t|Z = z) = h_0(t) e^{\beta' z}.$$

In this application, we use the Least Absolute Shrinkage and Selection Operator (LASSO), an application of the ℓ_1 norm regularization penalty. In the Cox PH framework, this penalty term takes the form of:

$$\lambda \sum |\beta_i|$$

where λ represents our penalty coefficient. This penalty helps to avoid over-fitting of our data in an effort to improve the generalizability of our predictive model in different settings. The algorithm used here in `glmnet` uses the Breslow approximation to handle ties. For more details on the derivation of this term and the algorithm used to fit the penalized Cox PH model (Simon et al.).

Random Survival Forest

The survival tree and the corresponding random survival forest (RSF) are highly favorable non-parametric methods when studying survival data. Generally, for a single survival tree, it will assign subjects to groups based on certain splitting rules regarding their covariates, and the subjects in each group will share a similar survival behavior. Trees are known for their instability, in the sense that small perturbations can induce a large change in their predictive function. Hence, random forests is an ingenious solution to this problem by reducing the variance of a single tree and enlarging the class of models. To obtain a prediction at a given x, the Nelson-Aalen estimates of the cumulative hazard functions at each node are averaged (Bou-Hamad, Imad, et al.). This model, being non-parametric, makes no additional assumptions. However, we continue to assume that the censoring is non-informative.

Brier score

We compare the Cox proportional-hazards model with the random survival forest by calculating the Brier score for each model on the test data set. The formula for the Brier score is as follows.

$$BS = \frac{1}{n} \sum_{i=1}^{n} (p_i - o_i)^2$$

The Brier score is used to evaluate the accuracy of probabilistic predictions from a model; its value ranges from 0 to 1 with 0 being perfect and 1 being the opposite. We split the data into training and test data sets (80-20 split). We fit the models on the training data set, and we calculate the Brier score using the test data set. For each observation in the test data set, predictions are made at the observed time of censoring or event (Brier 1-3).

Results

Log-rank test

To answer the research questions of whether hormone treatment and chemotherapy are effective to breast cancer, the initial step is to test whether the survival probability functions of treatment group and control group are identical with log-rank test. In this report, we conducted 3 log-rank tests.

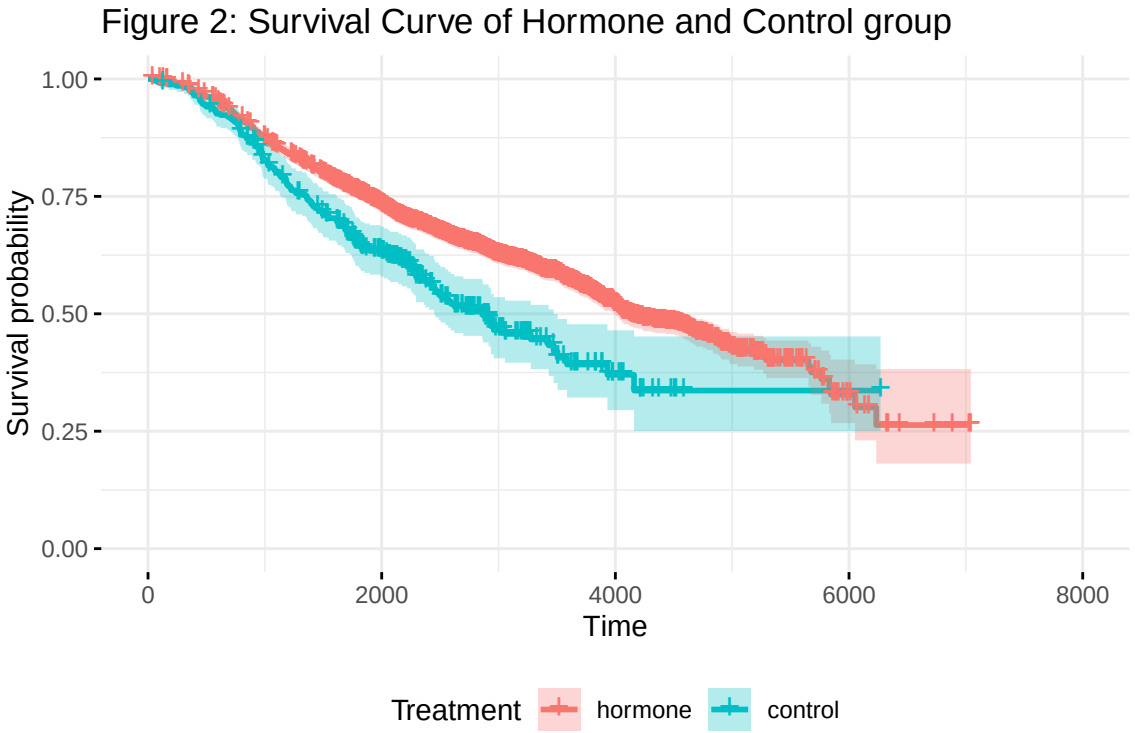
Hormone Therapy

First, we test whether hormone therapy is effective. The null hypothesis of this test is $H_0 : S_1(t) = S_0(t)$, where $S_1(t)$ is the survival function of hormone treatment group, $S_0(t)$ is the survival function of control group, and the alternative hypothesis is $H_1 : S_1(t) \neq S_0(t)$.

Table 3: Log Rank Test of Hormone Therapy

hormon	N	obs	exp
0	2643	1113	1161.6299
1	339	159	110.3701

The test statistic is 23.7, and the corresponding p-value is $1.133^{-6} \ll 0.05$, thus we reject the null and conclude that we are 95% confident that $S_1(t) \neq S_0(t)$.



And based on the survival plot above, we can further conclude that the hormone treatment is effective to breast cancer on average.

Chemotherapy

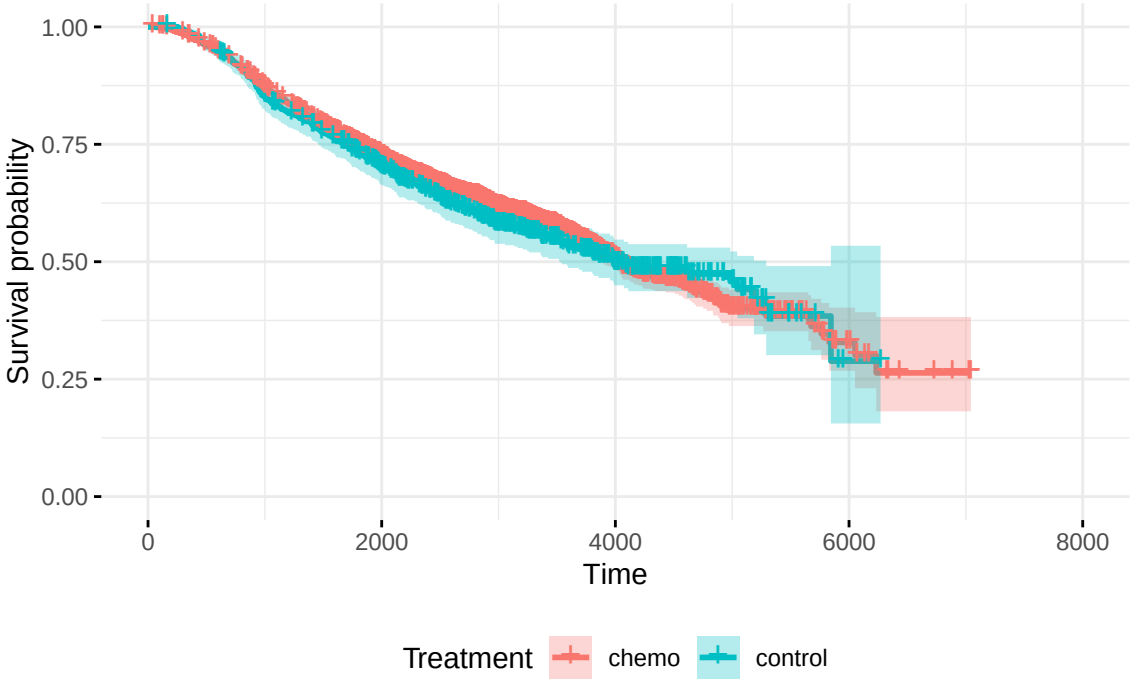
Then, we test whether chemotherapy is effective. The null hypothesis of this test is $H_0 : S_1(t) = S_0(t)$, similarly, $S_1(t)$ is the survival function of chemo treatment group, $S_0(t)$ is the survival function of control group, and the alternative hypothesis is $H_1 : S_1(t) \neq S_0(t)$.

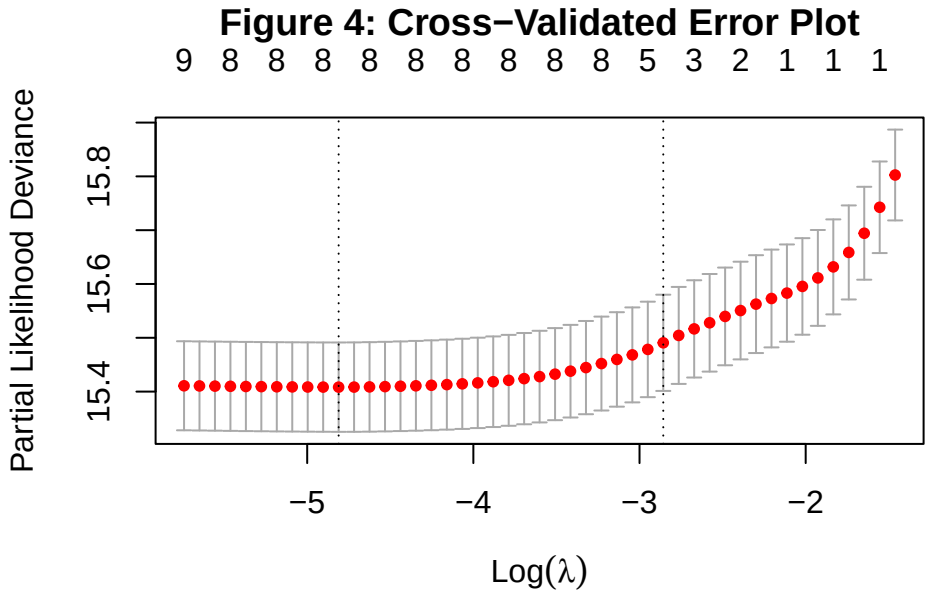
Table 4: Log Rank Test of Chemotherapy

chemo	N	obs	exp
0	2402	1014	1023.9321
1	580	258	248.0679

The test statistic is 0.495, and the corresponding p-value is $0.48 > 0.05$, thus we fail to reject the null and conclude that we are 95% confident that $S_1(t) = S_0(t)$. In other words, the chemotherapy is not effective to the treatment of breast cancer.

Figure 3: Survival Curve of Chemo and Control group





In Figure 4, we see the results of our cross-validation with the partial likelihood deviance, a measure of error, on the y-axis, and the natural log of the penalty term on the x-axis. The far-left dashed line represents the $\log(\lambda)$ value associated with the model that produced the minimum cross-validation error. The far-right dashed line represents the $\log(\lambda)$ value associated with the "1se" model that results in a cross-validation error one standard error from the minimum. We choose the "1se" model in this application because it increases the penalization on the model, enforces parsimony, and helps to reduce overfitting of our model on the training data set.

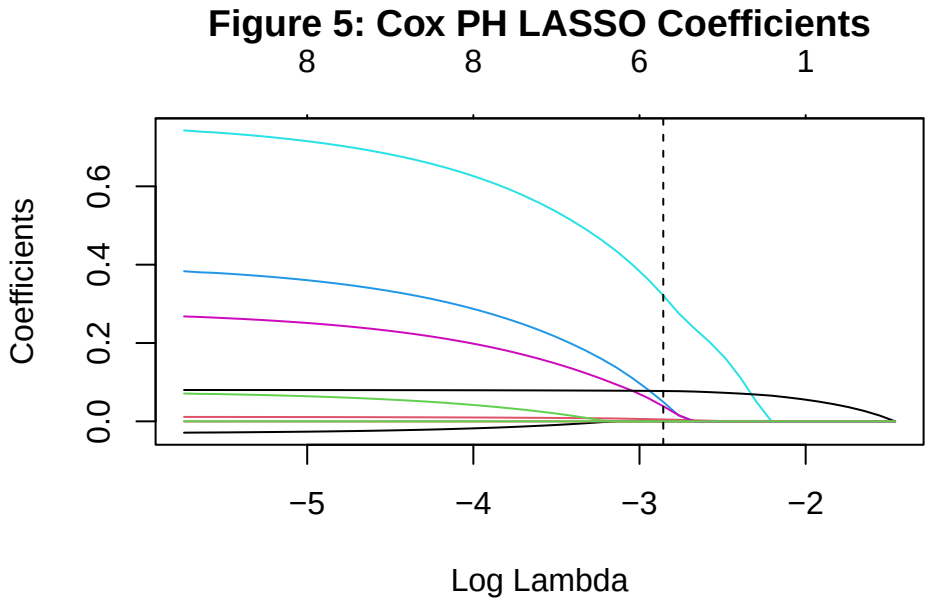


Figure 5 shows the values of each coefficient in the model as the penalty term increases. The dashed line shows the "1se" penalty term, and so the coefficient values at this point can be seen in the model. The cross-validated modeling results are shown below.

Table 5: Cox Proportion Hazard LASSO Coefficients (1se)

	coef	exp_coef
year	0.0000	1.0000
age	0.0048	1.0048
meno	0.0000	1.0000
size20-50	0.0495	1.0508
size>50	0.3216	1.3793
grade	0.0386	1.0393
nodes	0.0774	1.0805
pgr	0.0000	1.0000
er	0.0000	1.0000
hormon1	0.0000	1.0000
chemo1	0.0000	1.0000

Here, the regularization procedure removes `meno`, `er`, `hormon` and `chemo`. However, we are still interested in assessing the treatment effects of hormone therapy and chemotherapy, so we will add these back to fit our final model.

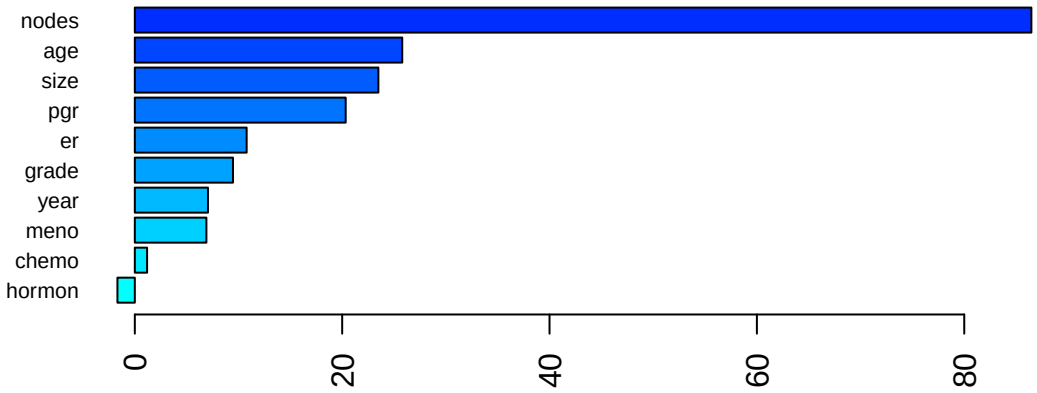
Table 6: 95% Confidence Intervals of Final Cox PH Coefficients

	2.5 %	97.5 %
age	1.0096113	1.0209412
size20-50	1.3250933	1.7610328
size>50	1.7728357	2.6497651
grade	1.1207981	1.5335347
nodes	1.0717882	1.0964130
pgr	0.9993588	0.9998718
hormon1	0.7818093	1.1517154
chemo1	0.8503259	1.2090508

And we find significant effects for age at surgery, size of tumor, differentiation grade, number of positive lymph nodes, and pgr. We find non-significant effects for each of the two treatments of interest.

The main assumption of this model is that the hazards are proportional between subgroups. We can check this assumption by calculating and plotting Schoenfeld residuals, which can be found in the appendix. From the plots we see violations in number of positive lymph nodes and minor violations in age, tumor size, differentiation grade, and progesterone receptors where 0 does not lie in the Schoenfeld residual 95% confidence interval.

Figure 6: Variable Importance in the Random Survival Forest



With `ranger` package, we trained the random survival forest (RSF) with training dataset used for survival prediction. As a non-parametric method, there are no parameters in RSF that could be interpreted. The ultimate goal of RSF is to predict the survival probability function of a given data point based on its covariate vector. Compared to semi-parametric Cox-PH model which forces the outcome and the covariates to have a special connection, the RSF makes prediction based on the survival time of training data points that shares similar propensity with the given input data point.

Figure 7: Survival Prediction for Patient 7

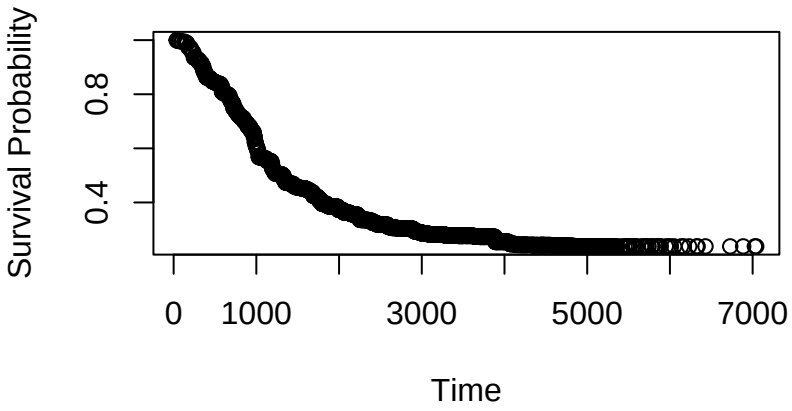


Table 7: Predicted Median Survival Time of Patient 7

	time	survival
382	1319	0.4994471
383	1322	0.4994471
384	1323	0.4990825
385	1325	0.4990825
386	1327	0.4990101
387	1331	0.4849616

Table 8: Observed Data for Patient 7

	year	age	meno	size	grade	nodes	pgr	er	hormon	chemo	dtime	death
2463	1992	69	1	20-50	2	8	5	6	1	0	1869	0

Since the "truth" of test data point (a single survival time) and the prediction we made here (a survival probability function) are not comparable, here we show the prediction result of the 7th test data point (pid = 58). The survival curve has been shown above, and the estimated median survival time is 1217 days.

Brier Score

First, we calculated the Brier score for the Cox proportional-hazards model by:

$$BS = \frac{1}{n}\sum_{i=1}^{n}(p_i - o_i)^2;$$

and found the Brier score for the Cox proportional-hazards model to be 0.329. Second, we found a Brier score for the random survival forest model of 0.319. The two models have very similar Brier scores. While we did not perform any statistical tests, given how close the two values are, it is fair to conclude that the Cox proportional-hazards model and the random survival forest are comparable in terms of predicting the outcome. However, if the goal of an analysis is inference, the semi-parametric Cox proportional-hazards model would be preferred to the non-parametric random survival forest.

An alternative metric that could be used instead of the Brier score is the area under the receiver operating characteristic curve.

Post-hoc Analysis

After testing the effect of hormone and chemotherapy separately, we were interested in whether there is interaction (or confounding) between the two treatments. Hence, we reassigned the data into 4 new treatment groups: 1) receive both hormone and chemotherapy; 2) only receive hormone therapy; 3) only receive chemotherapy; and 4) not receiving both of the therapies.

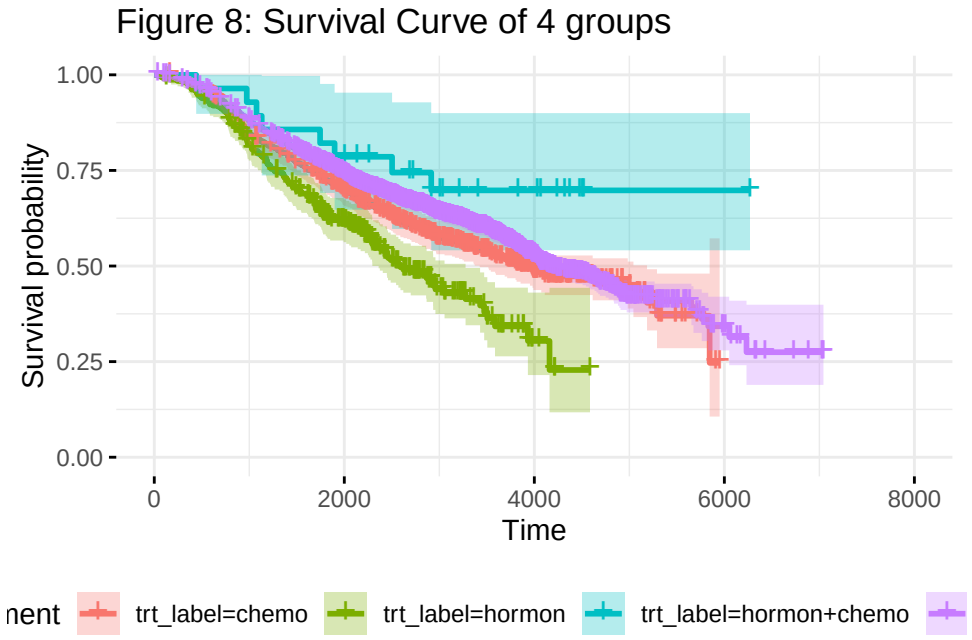
Then we tested the survival function of these 4 groups with log-rank tests. Here, the null hypothesis is $H_0 : S_1(t) = S_2(t) = S_3(t) = S_4(t)$, corresponding to the survival function of 4 groups; and the alternative hypothesis is: the survival probability function of 4 groups are not identical.

Table 9: Log Rank Test of Therapy Combinations

trt_label	N	obs	exp
chemo	552	250	233.74154
hormon	311	151	96.04381
hormon+chemo	28	8	14.32631
none	2091	863	927.88833

The test statistics is 40.4, with p-value $8.84^{-9} \ll 0.05$, then we reject the null and conclude that the survival function of 4 groups are not identical. But this result is not very meaningful, because we are more interested in the way they are different.

Consequently, we refer to the survival plot:



Based on the plot, it seems like: Hormone therapy or Chemotherapy itself cannot improve the survival of breast cancer, while the combination therapy of Hormone + Chemo can improve the survival.

Although it is a very interesting finding, but we should question whether this conclusion is valid. According to the summary data shown below, the sample size of hormone+chemo treatment group is only 28, which accounts for 8% of people who receive hormone therapy, and accounts for 1% of the whole dataset. As a result, the survival curve predicted for this combo treatment group can be unconvincing and biased comparing to other treatment groups, which makes the conclusion drawn above to be invalid.

Table 10: Frequency of Treatments in Sample

Var1	Freq
chemo	552
hormon	311
hormon+chemo	28

Var1	Freq
none	2091

Although the results drawn from the third plot are not reliable, it is still a discovery that should be highlighted by researchers, since it is a common phenomenon biologically that two drugs can perform treatment effect only when taken together. More clinical trials and statistically analyses are suggested to further explore this issue.

Discussion

Conclusions

Hormone therapy and chemotherapy appear to have non-significant treatment effects on outcomes after adjusting for covariates. Tumor size and differentiation grade had the largest magnitude effects on hazard and were significant after adjusting for the covariates chosen using the LASSO regularization method in the Cox Proportional Hazard model. The Cox PH and random survival forest models performed similarly in predicting the outcome. If questions of inference are of interest, we would recommend using the Cox PH model here due to limitations of inference using the random survival forest. However, the non-parametric nature of the random survival forest model may make it more generalizable and more widely applicable to new data.

Limitations and Future Work

Some limitations are that there was a small sample size for the treated, especially when it comes to treatment subgroup of receiving both chemo and hormone therapy (only 1% of samples). As discussed in the exploratory data analysis, the treatment groups varied significantly on many variables. When looking at our inference and prediction methods, we found violations of the proportional hazard assumptions, which are important for a valid Cox PH model. In future work, we would recommend using a variable to stratify the Cox PH model, particularly number of positive lymph nodes. Alternatively, a time-varying coefficient could be added to the model. The other model of interest, the random survival forest, introduces some amount of bias when estimating individual survival curves. Lastly, in this analysis we ignored remission, however this could provide valuable insight in a future study.

References

Bou-Hamad, Imad, et al. "A Review of Survival Trees." Statistics Surveys, vol. 5, no. none, 2011, https://doi.org/10.1214/09-ss047.

Brier, Glenn W. "Verification of Forecasts Expressed in Terms of Probability." Monthly Weather Review, vol. 78, no. 1, 1950, pp. 1–3., https://doi.org/10.1175/1520-0493(1950)078%3C0001:-vofeit>2.0.co;2.

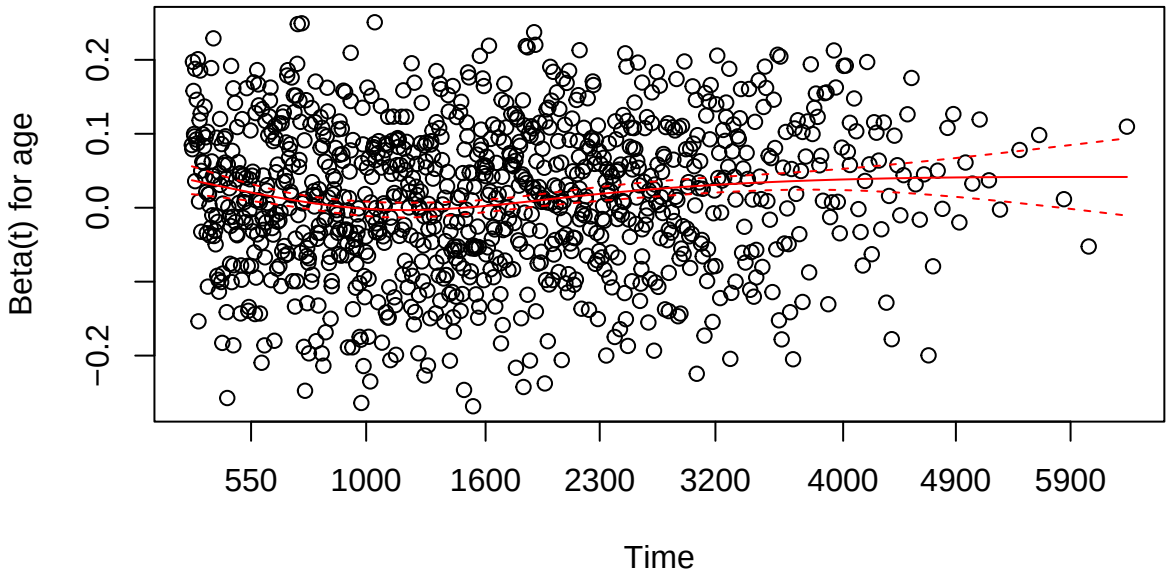
Hortobagyi, Gabriel N., et al. "The Global Breast Cancer Burden: Variations in Epidemiology and Survival. " Clinical Breast Cancer, vol. 6, no. 5, Dec. 2005, pp. 391–401. ScienceDirect, https://doi.org/10.-3816/CBC.2005.n.043.

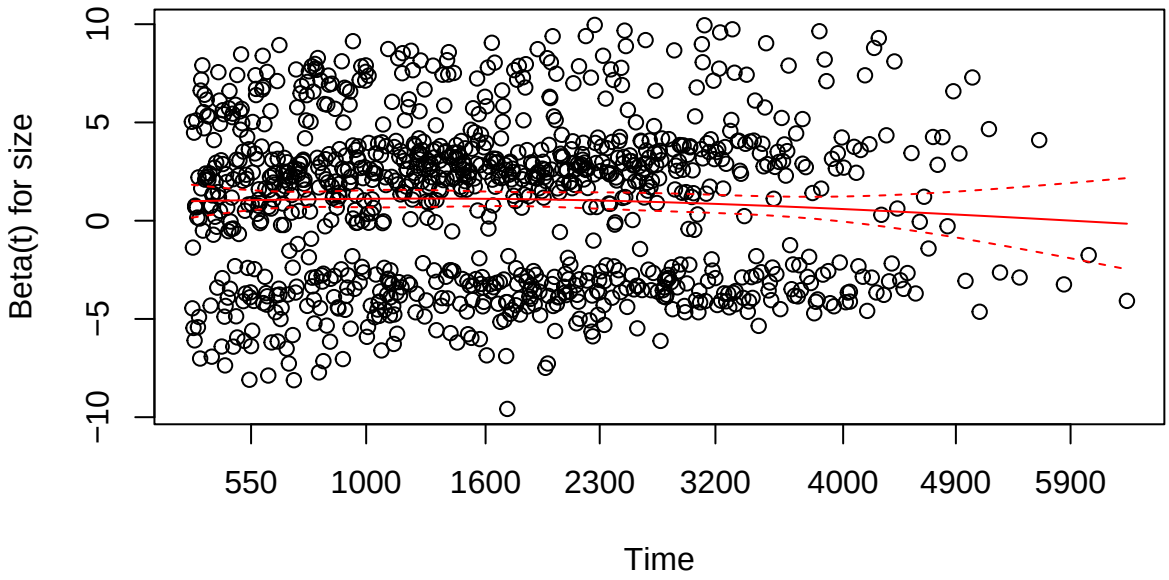
Simon, Noah et al. "Regularization Paths for Cox's Proportional Hazards Model via Coordinate Descent." Journal of statistical software vol. 39,5 (2011): 1-13. doi:10.18637/jss.v039.i05/

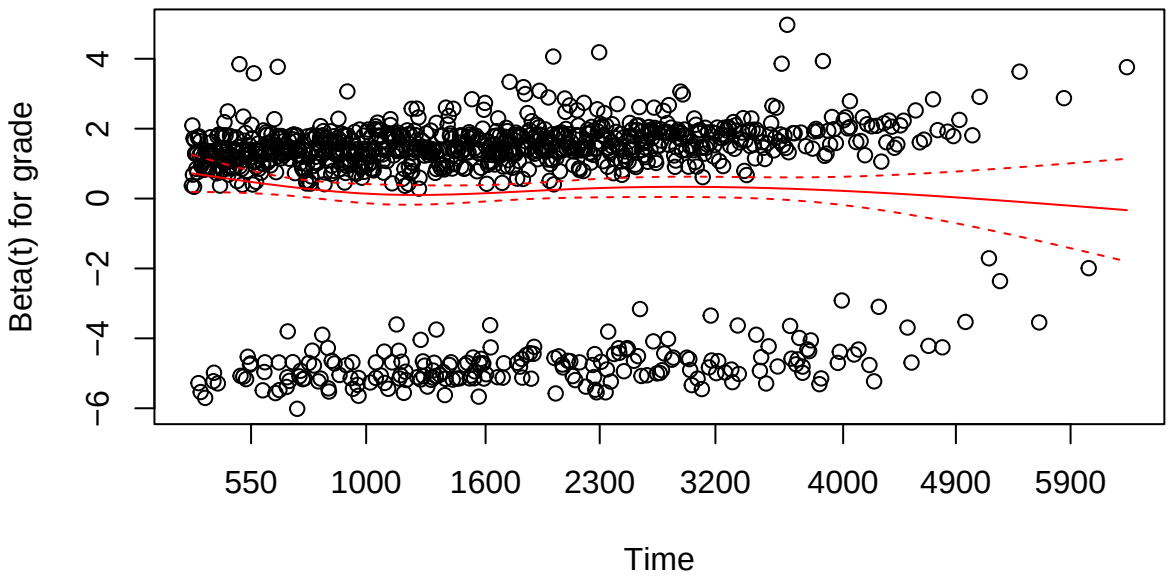
Wilcken, Nicholas, et al. "Chemotherapy Alone versus Endocrine Therapy Alone for Metastatic Breast Cancer." The Cochrane Database of Systematic Reviews, vol. 2003, no. 2, Apr. 2003, p. CD002747. PubMed Central, https://doi.org/10.1002/14651858.CD002747.

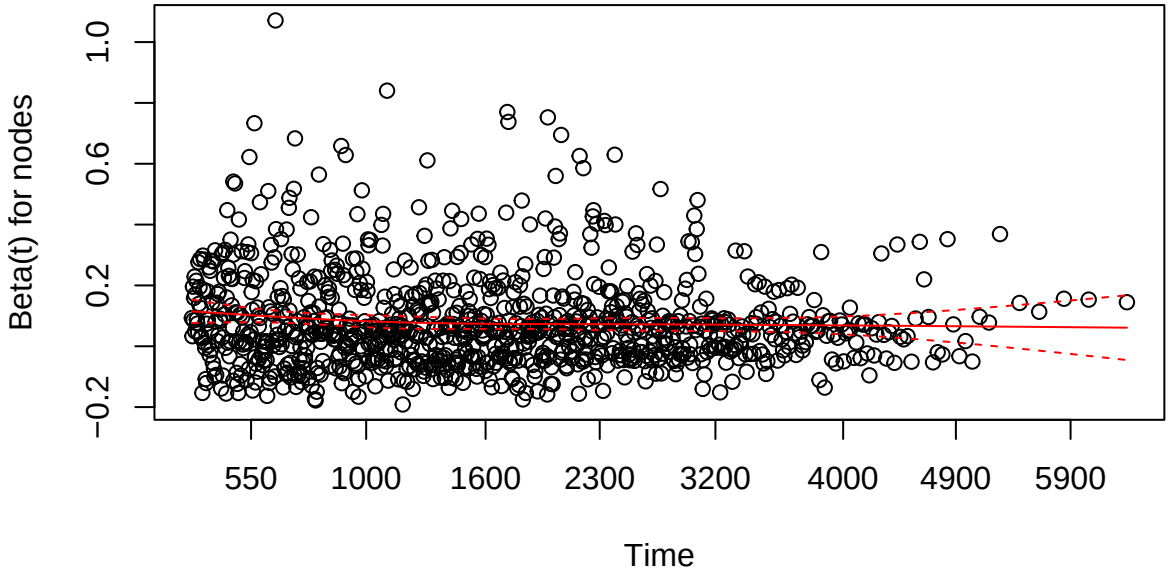
Appendix

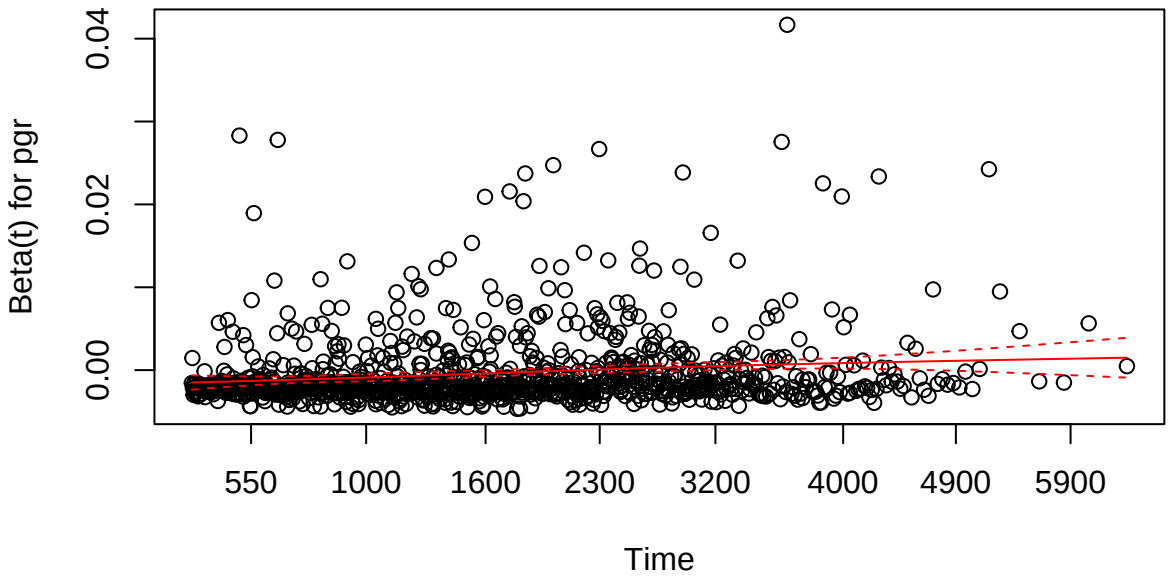
Figures 9 - 15

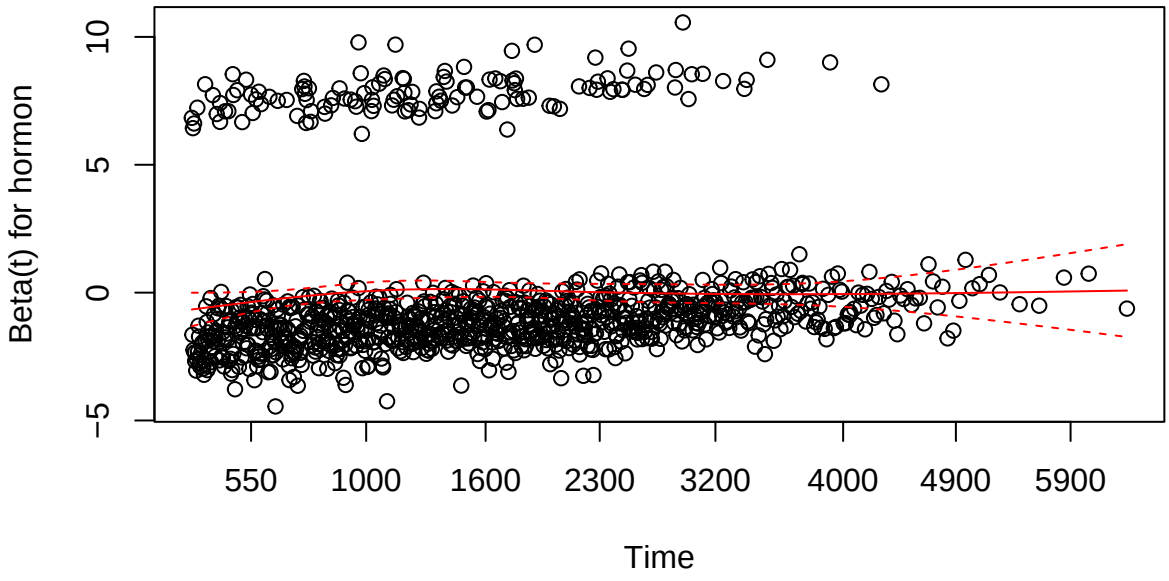


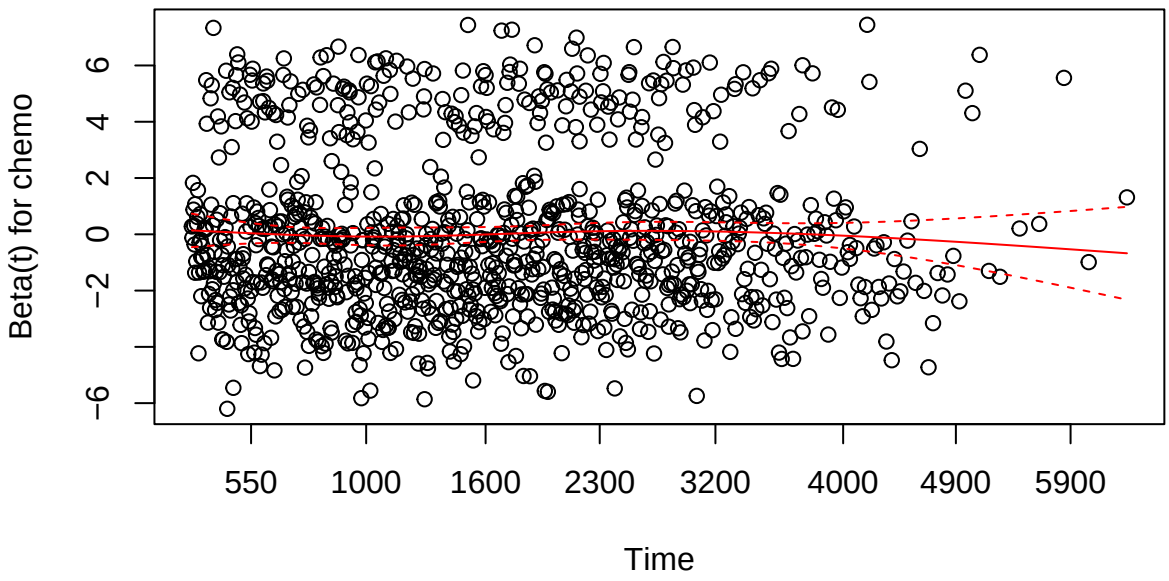



Beta(t) for grade
4
2
0
−2
−4
−6
550
1000
1600
2300
3200
4000
4900
5900
Time


Beta(t) for nodes
1.0
0.6
0.2
−0.2
550
1000
1600
2300
3200
4000
4900
5900
Time


Beta(t) for pgr
0.00
0.02
0.04
550
1000
1600
2300
3200
4000
4900
5900
Time


Beta(t) for hormon
−5
0
5
10
550
1000
1600
2300
3200
4000
4900
5900
Time



From these plots, we see violations in number of positive lymph nodes and minor violations in age, tumor size, differentiation grade, and progesterone receptors where 0 does not lie in the Schoenfeld residual 95% confidence interval.