ST 625 Project Report

Hana Sagor, Funfay Jen

15/Mar/2021

Introduction

An international multi-center placebo-controlled trial was undertaken to assess the efficacy of long-term use of CyA as the sole therapy. Three hundred forty-nine patients with PBC were randomized to receive CyA ($3mg \cdot kg^{-1}.day^{-1}$), or placebo with follow-up for 6 years. The end point was death or liver transplantation. The variables we consider in this analysis are • **ptno**: patient identification • **unit**: hospital (1: Hvidovre 2: London 3: Copenhagen 4: Barcelona 5: Munich 6: Lyon) • **tment**: treatment (0: placebo, 1: CyA) • **sex**: (1: males, 0: females) • **age**: years • **stage**: histological stage (1, 2, 3, 4) • **gibleed**: previous gastrointestinal bleeding (1: yes, 0: no) • **crea**: creatinine (micromoles/L) • **alb**: albumin (g/L) • **bili**: bilirubin (micromoles/L) • **alkph**: alkaline phosphatase (IU/L) • **asptr**: aspartate transaminase (IU/L) • **weight**: body weight (kg) • **days**: observation time (days) • **status**: status at exit (0: censored, 1: liver transplantation, 2 : dead)

Methods and Result

The goal of this analysis is to assess the effect of treatment. This involves three steps: model selection, model diagnostics, and estimation/inference. The first step involves three sub-steps: transformation of variables, determination of functional forms, and variable selection. On top of that, imputation is required for the data. The next subsections discuss each of these aspects.

Imputation

There are NA values throughout the data set. One important decision is whether to omit patients altogether if they have even one NA value, or to do some imputations. Some back-of-the-envelope calculations reveal that with only 1.6% NA values, we would be deleting 21.2% of the patients if imputations are not done. This is probably not a good idea. We have tried the simplest mean and median imputations as well as more sophisticated imputations with the aid of an R package. We use the latter for the analysis, but will compare results when appropriate.

Model Selection

Transformation of Variables

We first inspect univariate distributions of the variables to identify and patterns and whether to perform transformations.

Skewed data are serum bilirubin (micromoles/L), aspartate transaminase and alkaline phosphatase (IU/L), on which a log transformation is applied. We also transformed creatinine, which was not necessary. But this did not affect the final model selection since this variable turned out to be insignificant.

Determination of Functional Forms

We inspect the shape of the Martingale residuals for each continuous covariate separately. It takes some back-and-forth to try out quadratic and cubic terms for each covariate and to see how the significant covariates change. We finally arrive at candidate models with up to cubic transforms on the *age* and *Lbili* variables (more on the model selection process later).

Variable Selection

We also want to have at least some crude sense of what variables are important. We do this by carrying out the stepwise variable selection procedure. After dropping a non-significant variable (*unit*6) from the suggested final model given by the stepwise selection procedure, we arrive at a candidate model that has the following variables: *Lbili*, *age*, *alb*, *tment*, *sex* and *weight*. This model is identical to Model 2 in the original paper by M. Lombard et al. The rest of this report seeks to compare our results with the paper, while further assessing the suitability of this model as compared to other candidates.

We compare various candidate models and find that the model suggested by stepwise() and the paper has the second lowest AIC value only after the model with quadratic and cubic terms of *age*. Using simple grand mean/median imputation or omitting the patients with NAs lead us to the same ranking of AICs.

We use LRT to compare the top 2 models and do not find enough evidence to reject the simpler model.

Therefore, for parsimony, we go with model2 in this analysis. This also allows us to compare with the paper.

Model Diagnostics

We check the model assumptions once we decide on the suitable model.

The linear functional forms of the continuous covariates seem reasonable.

We perform both formal tests and graphical inspections for assessing the assumption. It seems that the PH assumption is reasonable for all of the covariates except for *age* in our chosen model. We decide to keep this model since the global test passed.

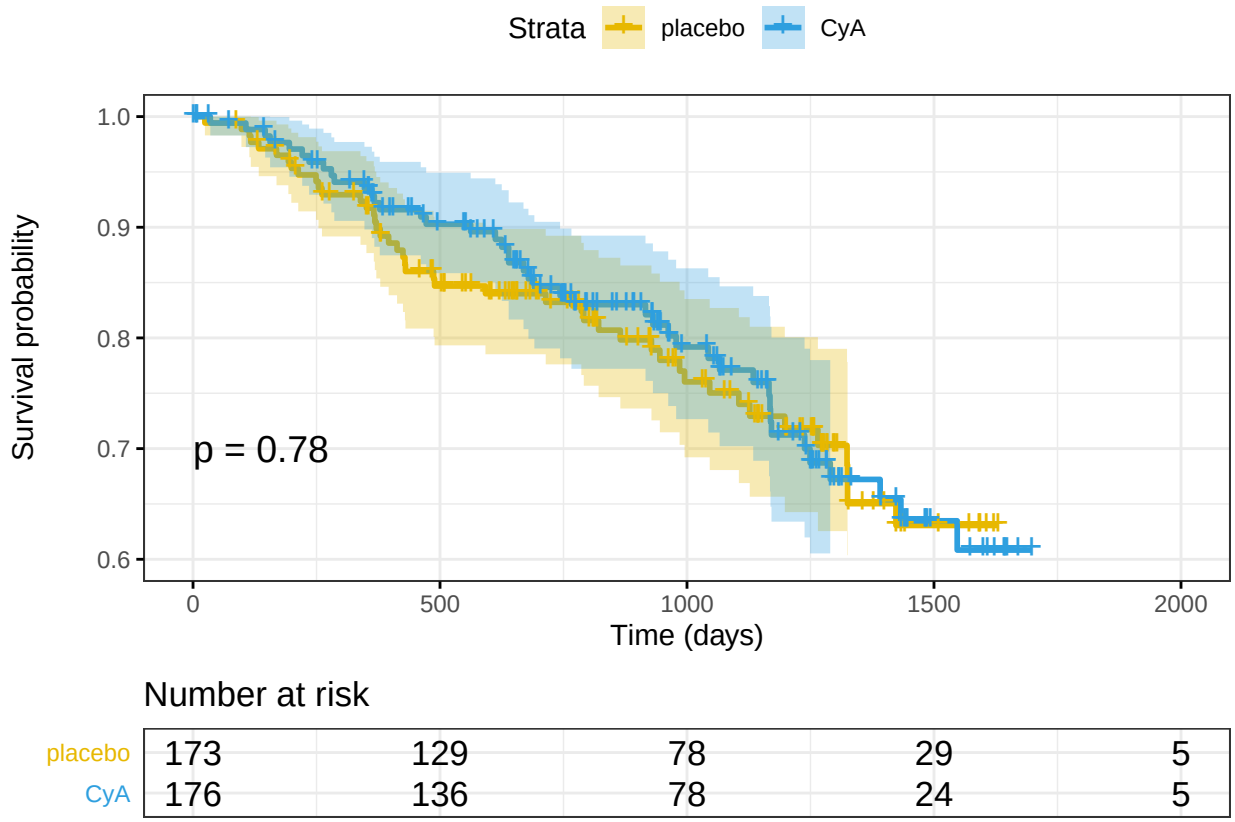
The Schoenfeld residuals indicate the PH assumption might not be met for the *Lbili* variable. Again, we still move forward with this model since overall the PH assumption seems to be satisfied.

Estimation and Inference

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |-------------------------|
##
##
## Total Observations in Table:  349
##
##
##              | Event
##    Treatment |         0 |         1 | Row Total |
## -------------|-----------|-----------|-----------|
##            0 |       127 |        46 |       173 |
##              |      0.73 |      0.27 |      0.50 |
## -------------|-----------|-----------|-----------|
##            1 |       132 |        44 |       176 |
```

```
##              |      0.75 |      0.25 |      0.50 |
## -------------|-----------|-----------|-----------|
## Column Total |       259 |        90 |       349 |
## -------------|-----------|-----------|-----------|
##
##
```

Among n=173 randomized to placebo, there were 46 deaths (27%) Among n=176 randomized to CyA, there were 44 deaths (25%) It appears that the difference between placebo and CyA groups is not significant without accounting for covariates.

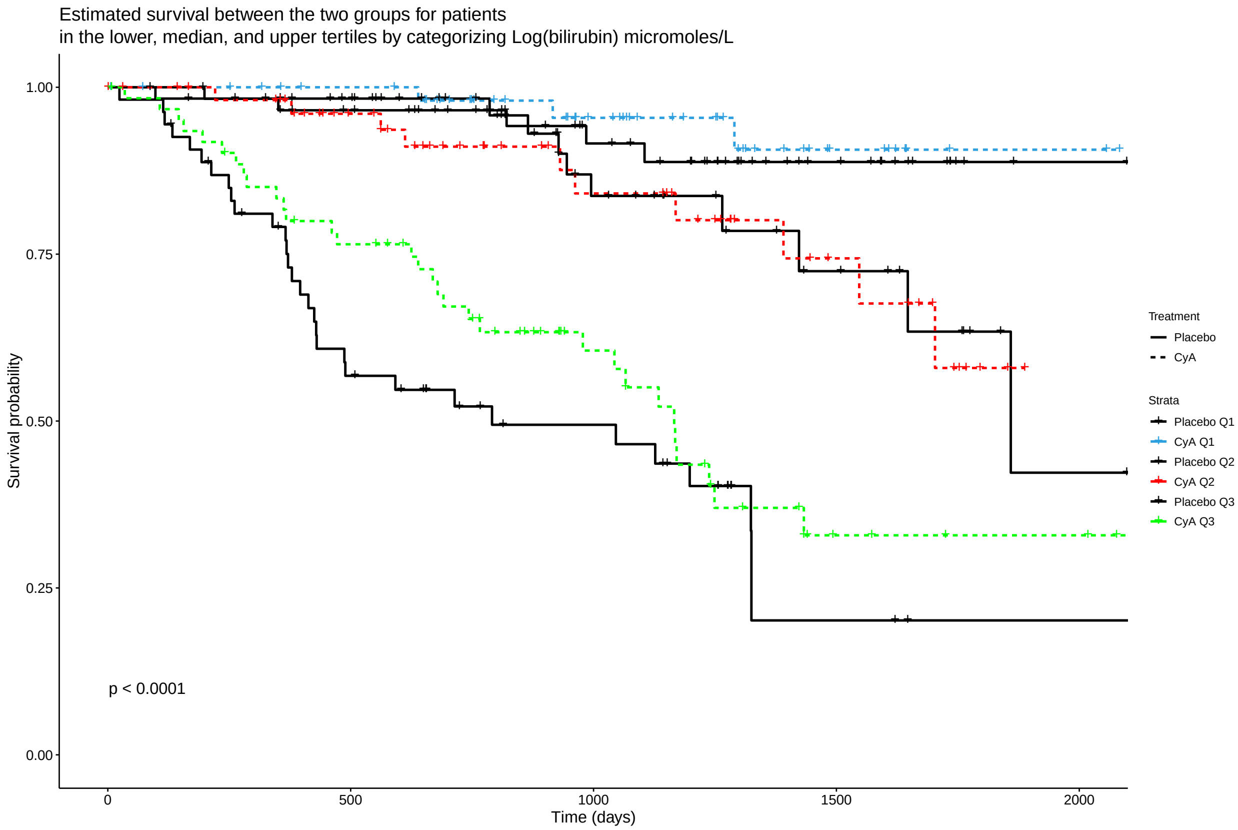


The KM plots by treatment also confirms the first guess from the previous table, showing significant overlap in the confidence bands. The log-rank test (p-value annotated on the KM plots) also fails to detect any difference in survival between the placebo and treatment groups.

Note that our results are close to the paper.

These univariate analyses may be misleading because of imbalance between the groups at entry with respect to important prognostic factors, whereas the multivariate analysis takes these differences in covariates into account, as discussed shortly.

We also reproduce Figure 3 in the paper by categorizing the *Lbili* levels into tertiles (this is exactly the same as categorizing *bili*, since the log function is monotone).



This does not exactly match the paper, and we are not sure if we used the same variable for categorizing patients in to tertiles. (Other variables have also been explored, but they do not lead to plots that match the paper.) However, we do start to see the potential of accounting for more variables: here, after *Lbili* is accounted for, there seems to be a visible advantage of the treatment (CyA) group over the placebo group for the third tertile of *Lbili*, and even perhaps in the first tertile. This idea is further explored with the Cox PH regression models. The other thing to take away from this plot is that the higher the *Lbili*, the poorer the survival or the higher the hazard. This indicates that the coefficient on *Lbili* is positive, which will be confirmed in the Cox model.

```
##
## ============================================================
##                              Dependent variable:
##                      ---------------------------------------
##                                     days
##                             (1)                 (2)
## ------------------------------------------------------------
## age                      0.029**             -0.974**
##                          (0.012)              (0.487)
##
## age2                                          0.020**
##                                               (0.010)
##
## age3                                         -0.0001**
##                                              (0.0001)
##
## sex1                    0.863***             1.038***
##                          (0.310)              (0.323)
##
```

```
## weight                        -0.026*            -0.029**
##                               (0.014)             (0.014)
##
## alb                           -0.064***          -0.061***
##                               (0.023)             (0.023)
##
## Lbili                         1.032***           1.051***
##                               (0.113)             (0.114)
##
## tment1                        -0.455**           -0.462**
##                               (0.221)             (0.221)
##
## ------------------------------------------------------------
## Observations                    349                 349
## R2                             0.332               0.341
## Max. Possible R2               0.934               0.934
## Log Likelihood               -403.737            -401.396
## Wald Test            134.810*** (df = 6) 135.430*** (df = 8)
## LR Test              140.752*** (df = 6) 145.433*** (df = 8)
## Score (Logrank) Test 161.637*** (df = 6) 165.248*** (df = 8)
## ============================================================
## Note:                              *p<0.1; **p<0.05; ***p<0.01
```

The above table lists the top two models side by side. We have already decided in favor of the simpler model based on the LRT, but we can also try justifying the choice from the coefficients. They are largely similar, except for the coefficients for *age*. We observed that the coefficients for quadratic and cubic terms of *age* are, while statistically significant, of low magnitudes. This hints at overfitting issues.

We also note that our chosen model gives coefficient estimates and standard errors that are very close to model 2 in the paper.

Based on the chosen parsimonious model, we conclude that, the hazard ratio of the treatment (CyA) group to the placebo group is 0.6343, with a confidence interval of [0.4116, 0.9774], accounting for other covariates. This shows that the treatment is statistically effective in reducing the hazard and improving survival of patients with the liver disease PBC.

Discussion

There are limitations of this analysis. First, we might have not appropriately accounted for correlation among the patients in the same hospital (unit). Second, the PH assumption needs to be further verified. Nonetheless, the fact that our results closely match those in the original paper gives us some confidence that our conclusions are at least on the right track.