PSTAT 175 Project - Revised

Ottia Houston *

2022-12-09

Introduction

We would like to evaluate change in disinfectant practices on burn victims and their infection time.

We got the data from 154 patients during and 18 month study. We are measuring the time (days) until staphylococcus infection.

In this study, the two disinfectant practices that were measured are routine bathing (the control) and body cleansing (the treatment). The control is the routine bathing which entailed surface decontamination with 10% povidone-iodine then a regular bathing with dial soap, while the treatment is a body cleansing that used 4% chlorhexidine gluconate.

The staphyloccus aureus, is a common pathogen often seen in places like the ICU. We are measuring the time until an infection was measured in days along with an indicator variable.

The covariates that were recorded was the type of burn (chemical, scald, flame, or electric), where the burn was located (head, buttocks, trunk, upper legs, lower legs, or respiratory), the total area burned (in percentages), gender (male or female), and race (white or not white). We also have recored two time-dependent covariates (excision or antibiotics) and their corresponding indicator variables.

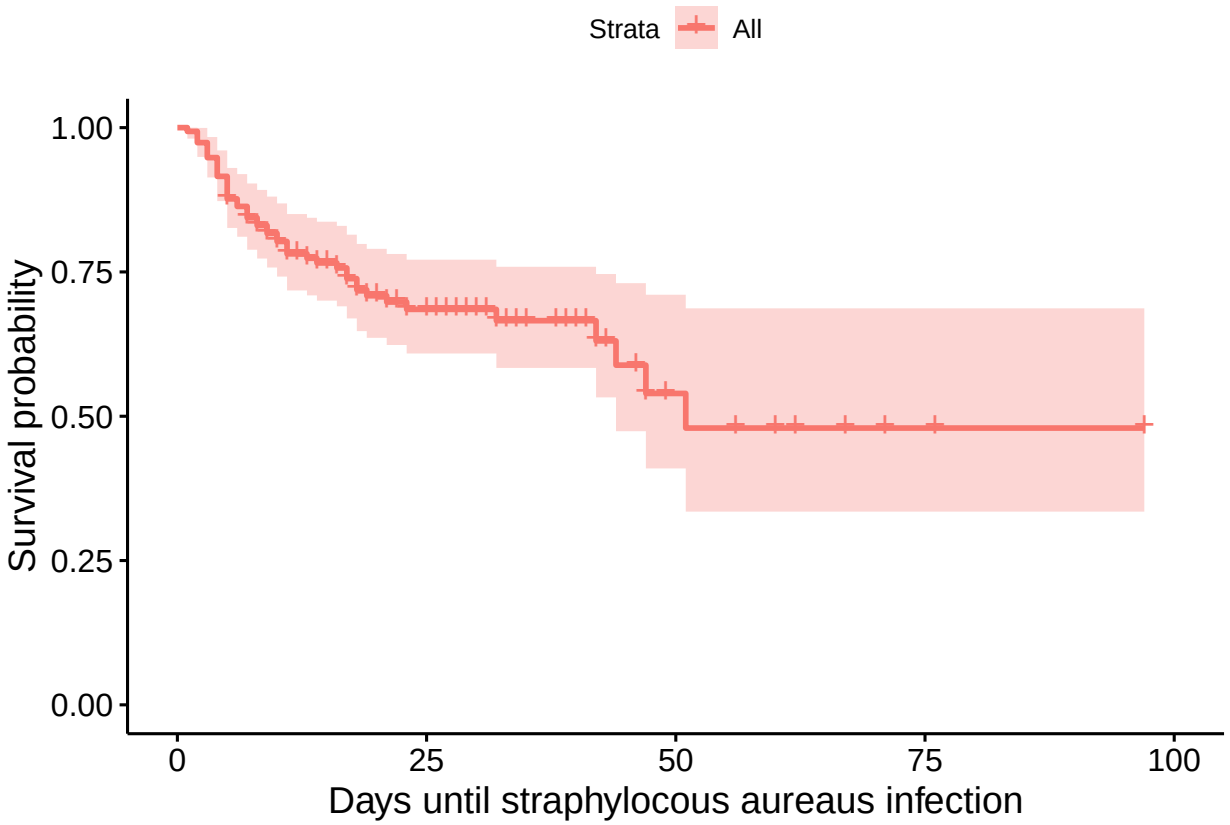
Whether or not antibiotics or excision is performed on the patient depends on a case by case basis but is usually implemented when a patient enters sepsis. Sepsis is a change in the patient that can be of concern for infection (Norbury, W. et al).

Overall, this study is going to be able to allow us to analyze if the change in disinfectant practices will be a benefit when it comes to burn victims infection rate.

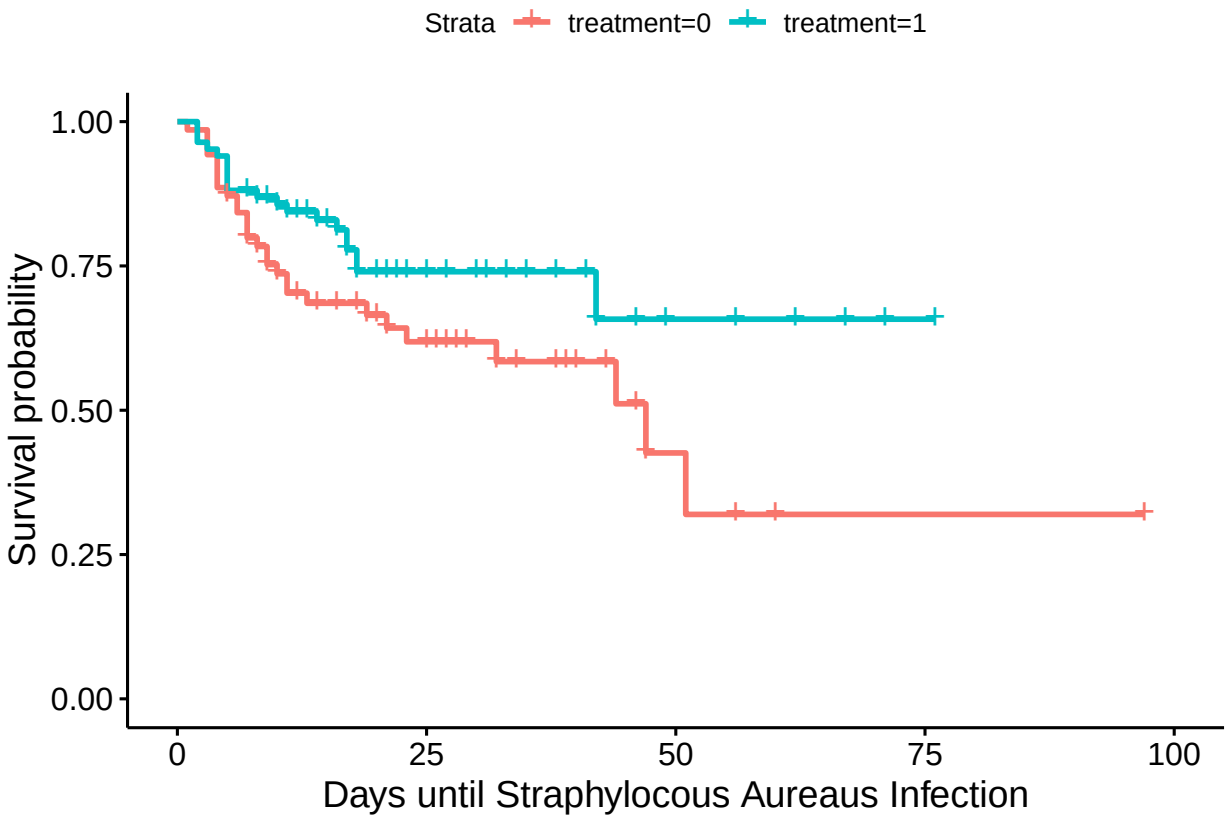
Model

First we are going to establish a model using the survival library by Therneau T (2022). Looking at a km plot we can see,

```
#KM of burn data
surviver.KM <- survfit(Surv(time_sa,sa_infection)~1, data=burn)
ggsurvplot(surviver.KM) + labs(x="Days until straphylocous aureaus infection")
```




```
#KM of the treatment
surviver.km <- survfit(Surv(time_sa, sa_infection) ~ treatment, data = burn)
ggsurvplot(surviver.km) + labs(x="Days until Straphylocous Aureaus Infection")
```



that treatment 1 is doing slightly better than treatment 0.

Trying to find the best model, we are using the stepAIC function from the MASS library and coxph from the survival library.

Using stepAIC forward selection, we determined that the best model is coxph(survVec ~ race + excision + pa_treatment + burnpct, data = burn). We saw that the AIC for the model is AIC=420.0 which was the best AIC in comparison to the rest. Also we can saw the LRT p-value to be less than $\alpha = 0.05$ as it's at about p=2.582e-05.

But looking at the p-value for the burnpct covariate, it was over $\alpha = 0.05$, so we decided to take it out. Leaving us with the model coxph(survVec ~ race + excision + pa_treatment, data = burn) with an AIC of 421.12 which is not too far away from the initial model AIC and the covariates p-values are all significant at the 0.05 level.

Now, looking at the Analysis of Deviance Table for our model we get,

```
survVec<- Surv(burn$time_sa, burn$sa_infection)
anova(coxph(survVec ~ race + excision + pa_treatment, data = burn))
```

```
## Analysis of Deviance Table
##  Cox model: response is survVec
## Terms added sequentially (first to last)
##
##              loglik  Chisq Df Pr(>|Chi|)
## NULL         -219.29
## race         -214.50 9.5656  1   0.001983 **
## excision     -211.53 5.9440  1   0.014767 *
```

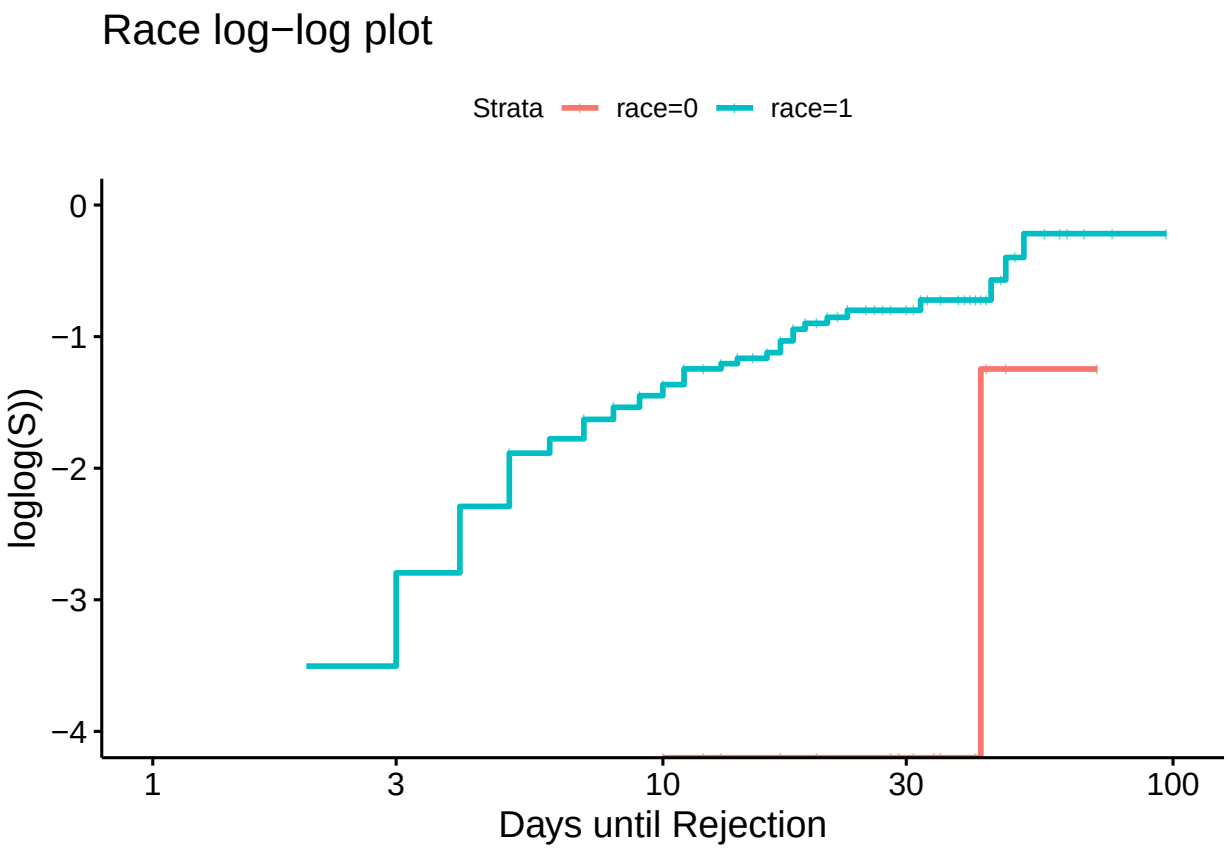
```
## pa_treatment -207.57 7.9232  1   0.004880 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

we can see the p-values are good to good enough at the 0.05 level.

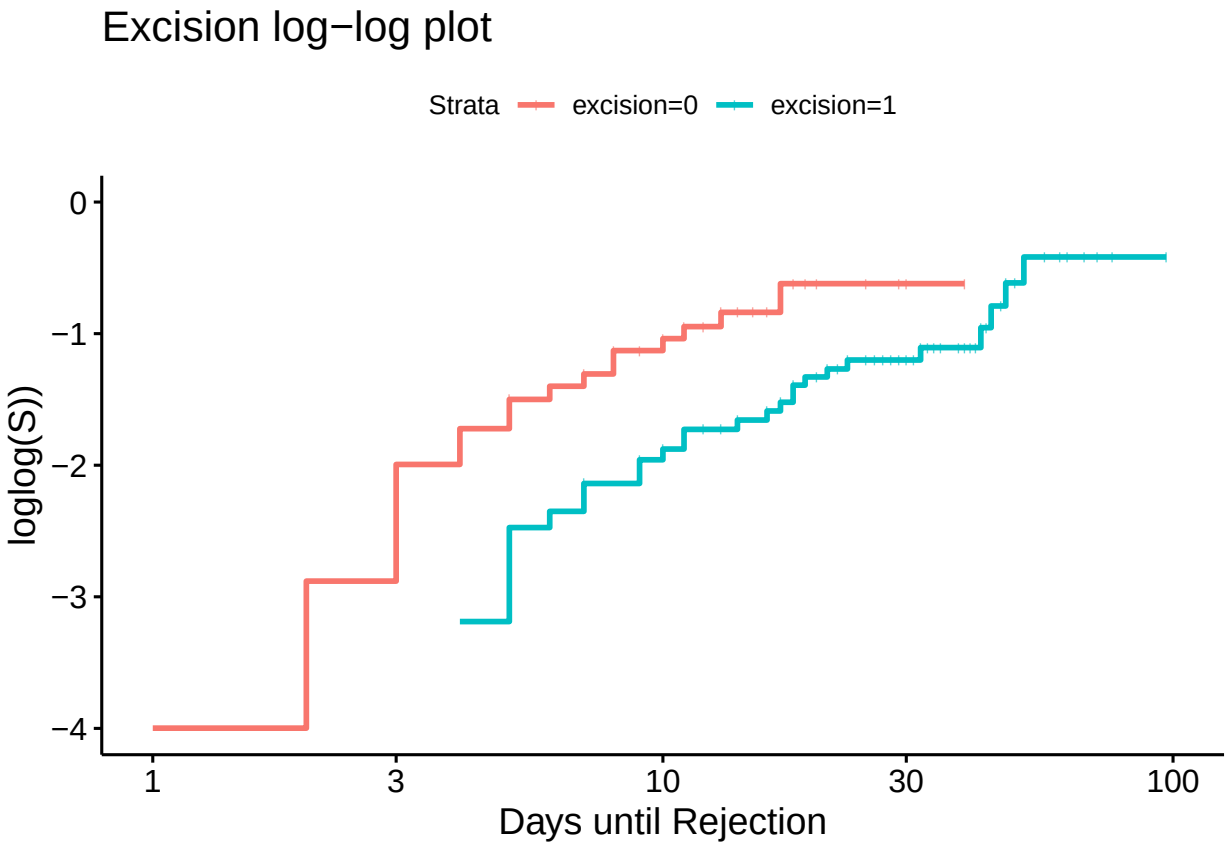
Checking our Assumptions

Now checking the assumptions by using a log-log plot we can see

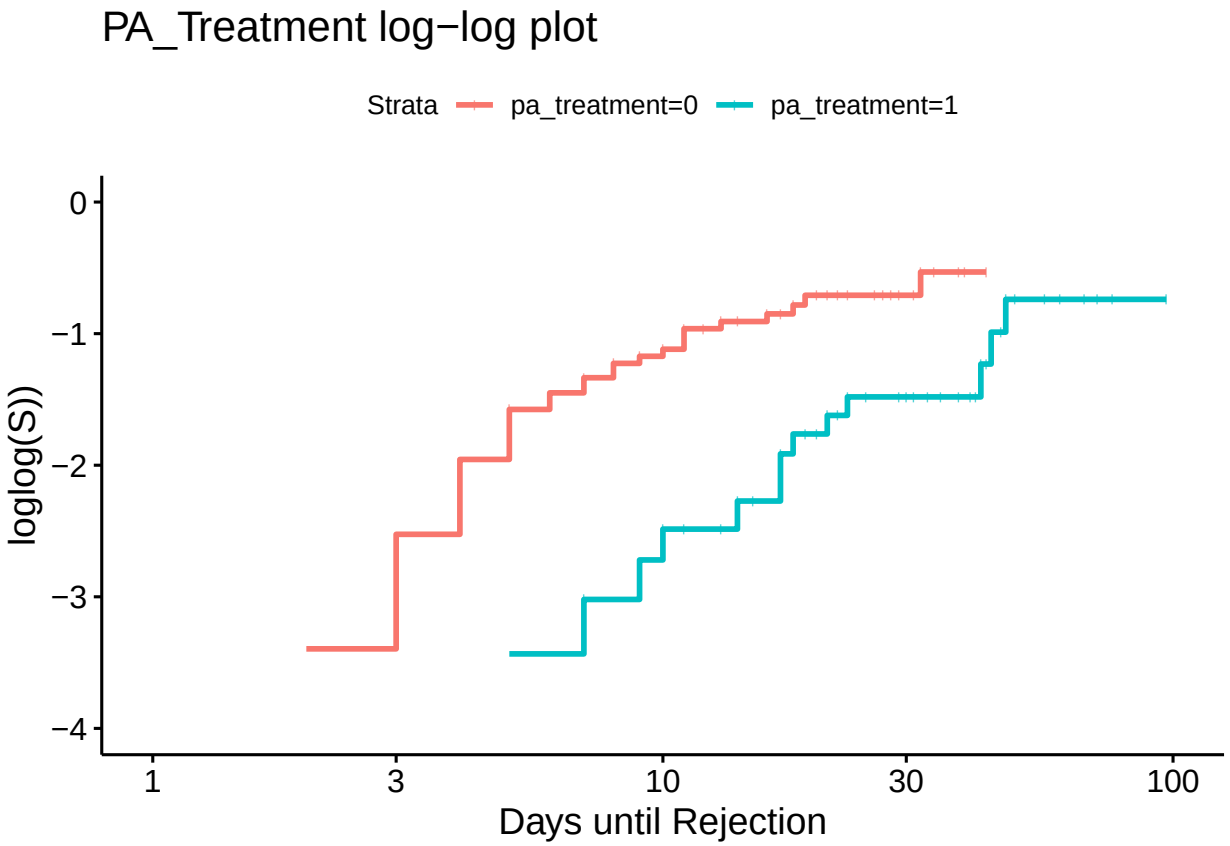
```
ggsurvplot(survfit(survVec ~ race, data = burn), fun="cloglog",
legend.labs=c(), censor.shape=124, censor.size=1,
ylim=c(-4,0)) +
labs(x="Days until Rejection", y="loglog(S))", title = "Race log-log plot")
```




```
ggsurvplot(survfit(survVec ~ excision, data = burn), fun="cloglog",
legend.labs=c(), censor.shape=124, censor.size=1,
ylim=c(-4,0)) +
labs(x="Days until Rejection", y="loglog(S))", title = "Excision log-log plot")
```




```
ggsurvplot(survfit(survVec ~ pa_treatment, data = burn), fun="cloglog",
legend.labs=c(), censor.shape=124, censor.size=1,
ylim=c(-4,0)) +
labs(x="Days until Rejection", y="loglog(S))", title = "PA_Treatment log-log plot")
```



that for the excision and pa_treatment it seem that they do not violate the coxph assumptions. They are parallel-ish to each other and they do not cross. But race is a bit worrying as it is not parallel at all. To look into it further we are going to look at cox.zph.

```
coxPH <- coxph(Surv(burn$time_sa, burn$sa_infection) ~ race + pa_treatment + excision , data = burn)
coxTest <- cox.zph(coxPH)
print(coxTest)
```

```
##              chisq df    p
## race          2.23  1 0.13
## pa_treatment  1.54  1 0.22
## excision      1.09  1 0.30
## GLOBAL        4.56  3 0.21
```

We can see that the p-value is higher than 0.05, so we fail to reject the null hypothesis. This means that are proportional hazards assumptions are met. So, we are justified in using the proportional hazards assumption in our modeling of the affect of race, pa_treatment, excision when inquiring about the rate of SA infection.

Conclusion

```
coxph(Surv(burn$time_sa, burn$sa_infection) ~ race + excision + pa_treatment , data = burn)
```

```
## Call:
```

```
## coxph(formula = Surv(burn$time_sa, burn$sa_infection) ~ race +
##     excision + pa_treatment, data = burn)
##
##                  coef exp(coef) se(coef)      z       p
## race          2.1425    8.5204   1.0114  2.118 0.03415
## excision     -0.6463    0.5240   0.3217 -2.009 0.04451
## pa_treatment -0.8990    0.4070   0.3338 -2.693 0.00708
##
## Likelihood ratio test=23.43  on 3 df, p=3.28e-05
## n= 154, number of events= 48
```

It seems that race when to race we can see that non-white people have a lower survival rate in terms of getting an SA infection. And with excision and pa_treatment, it seems like the treatment groups (getting an excision and getting a pa_treatment) has a higher survival rate when it comes to getting an SA infection.

For the hazard ratios we take exp(coeff) that was reported above. So we get 8.520713 for race, 0.5327013 for excision, and 0.4069764 for pa_treatment. Their respective 95% confidence interval rates are (7.4064, 9.6344), (.2023, .8457), and (.0732, .7408).

So going back to the initial question at hand. We wanted to see if the change in treatment affects the SA infection rate in burn victims. Our model that we produced showed that treatment was not significant. In conclusion, we can see that the treatment does not matter that much when it comes to SA infection in burn victims.

It seems that race, excision, and pa_treatment is best when it comes to determining SA infection.

Advanced Model

Looking at our simpler model and the description, we know that we have two time dependent covariates which are excision and pa_treatment. To account for these two covariates, we decided to use a (start, stop, status) time varying model. When split at t = 25. We'll also be applying episode to excision and pa_treatment since they are the covariates relying on time.

```
splitBurn <- survSplit(Surv(time_sa, sa_infection) ~ race + excision + pa_treatment, data = burn, cut =
```

```
coxph(Surv(tstart, tstop, sa_infection) ~ race + episode:excision + episode:pa_treatment, data = splitBu
```

```
## Call:
## coxph(formula = Surv(tstart, tstop, sa_infection) ~ race + episode:excision +
##     episode:pa_treatment, data = splitBurn)
##
##                          coef exp(coef) se(coef)      z       p
## race                   2.1322    8.4335   1.0115  2.108 0.03504
## episode:excision      -0.6343    0.5303   0.3126 -2.029 0.04241
## episode:pa_treatment -0.7354    0.4793   0.2764 -2.660 0.00781
##
## Likelihood ratio test=22.66  on 3 df, p=4.751e-05
## n= 203, number of events= 48
```

Looking at the covariates' p-values we can see that they are significant to the model. Peering further, the LRT is

```
anova(coxph(Surv(tstart, tstop, sa_infection) ~ race + episode:excision + episode:pa_treatment, data = s
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(tstart, tstop, sa_infection)
## Terms added sequentially (first to last)
##
##                        loglik  Chisq Df Pr(>|Chi|)
## NULL                  -219.29
## race                  -214.50 9.5656  1   0.001983 **
## episode:excision      -211.61 5.7796  1   0.016213 *
## episode:pa_treatment -207.96 7.3159  1   0.006835 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

And with the Analysis of Deviance Table, the p-values are less than 0.05 which means the covariates are significant at the 0.05 level.

Checking Assumptions: Advanced Model

Now we'll examine if there are any assumption violations via the cox.zph function.

```
survVec2 <- Surv(splitBurn$tstart, splitBurn$tstop, splitBurn$sa_infection)
cox.zph(coxph(Surv(tstart, tstop, sa_infection) ~ race + episode:excision + episode:pa_treatment, data =
```

```
##                       chisq df    p
## race                   2.40  1 0.12
## episode:excision       1.32  1 0.25
## episode:pa_treatment   2.14  1 0.14
## GLOBAL                 5.40  3 0.14
```

As we can see, the p-values for race, episode:excision, episode:pa_treatment, and globally are above 0.05 so we fail to reject the null so that means that they do not violate the coxph assumptions.

Conclusions: Advanced Model

```
coxph(Surv(tstart, tstop, sa_infection) ~ race + episode:excision + episode:pa_treatment, data = splitBu
```

```
## Call:
## coxph(formula = Surv(tstart, tstop, sa_infection) ~ race + episode:excision +
##     episode:pa_treatment, data = splitBurn)
##
##                          coef exp(coef) se(coef)      z       p
## race                   2.1322    8.4335   1.0115  2.108 0.03504
## episode:excision      -0.6343    0.5303   0.3126 -2.029 0.04241
## episode:pa_treatment -0.7354    0.4793   0.2764 -2.660 0.00781
##
## Likelihood ratio test=22.66  on 3 df, p=4.751e-05
## n= 203, number of events= 48
```

It seems that the covariates is still like the simpler model.When it comes to race we can see that non-white people have a lower survival rate and with excision and pa_treatment, it seems like the treatment groups (getting an excision and getting a pa_treatment) has a higher survival rate when it comes to getting an SA infection.

To get the hazard rates we take exp(coeff) that was reported previous in this section. So we get 8.4554 for race, 0.5303 for excision, and 0.4793 for pa_treatment. Their respective 95% confidence interval rates are (7.4219, 9.4449), (.2177, .8429), and (.2029, .7557).

*(Note: The revision was done on my own. I emailed the professor and he said I can submit the revision individually)

Citations

Norbury, W. et al. (2016) Infection in Burns, Surgical infections. U.S. National Library of Medicine. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4790211/ (Accessed: November 17, 2022).

Kassambara A, Kosinski M, Biecek P (2021). *survminer: Drawing Survival Curves using 'ggplot2'*. R package version 0.4.9, https://CRAN.R-project.org/package=survminer.

Klein and Moeschberger (1997) Survival Analysis Techniques for Censored and truncated data, Springer. Ichida et al. Stat. Med. 12 (1993): 301-310.

Klein Ob, Moeschberger, Yan mbJ (2012). *KMsurv: Data sets from Klein and Moeschberger (1997), Survival Analysis*. R package version 0.1-5, https://CRAN.R-project.org/package=KMsurv.

Therneau T (2022). *A Package for Survival Analysis in R*. R package version 3.3-1, https://CRAN.R-project.org/package=survival.

Terry M. Therneau, Patricia M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York. ISBN 0-387-98784-3.

Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, *4*(43), 1686. doi:10.21105/joss.01686 https://doi.org/10.21105/joss.01686.