Survival Analysis for Chlorhexidine Gluconate Protocols of *Staphyloccus aureus* Infection Treatment

Midterm Project

Richard Paul Yim

8 November 2022

Contents

1 Introduction

It's well understood that the primary source of infections of burn wounds by severely burned patients are the source of extended hospital stays. Within hospital burn units, of major concern are nosocomial infections, or infections occurring within a clinical environment, which can happen during surgical procedures removing wound eschar. The following survival analysis of burn patients is resulting from a study conducted from 1983 to 1985 introducing a protocol utilizing an antimicrobial detergent with 4 percent *chlorhexidine gluconate* (Hibiclens) substituting standard total body bathing with bar soap (Dial), for which the total body bathing with bar soap was started with decontamination with *povidoine-iodine* (Betadine).

This report is organized as follows: Section 2 introduces and discusses the data studied; Section 3 produces various model designs for survival analysis of the burn patients with both time independent and dependent covariates; Section 4 performs formal and relevant statistical analysis across designed models; Section 5 concludes the analysis with explicit interpretations; Section 6 and 7 concludes the report with discussion of the results and interpretations and future considerations for burn-care management.

2 Data

This section discusses and tables the original dataset and its covariates as well as a pivoted dataset with consideration for time-dependent covariates. Elementary statistical descriptions are also presented.

2.1 Original Dataset and Covariates

The burn dataset studied is the culmination of medical charts, microbiology laboratory records and abstracted data recorded and accounted for between January 1983 and December 1985. Burns for patients were categorized as either moderate or major by the American Burn Association (ABA) categorization. Additionally, the data primarily focuses on one group of infectious bacteria, *Staphylococcus aureus*. The dataset consists of 154 patients. Table 1 describes each variable followed by a recoded dataset used in the analysis seen in Table 2. (This dataset has 154 patient observations and 18 variables.)

2.2 Time-Dependent Covariate Dataset

The data as formatted above is useful for observing relapse/censoring with respect to time-independent covariates, but we also include a variation of the data pivoting on time-dependent covariates where we can later design an accurate statistical model that correctly and precisely incorporates time-dependent covariates. Table 3 shows the data dictionary for the additional time-dependent covariates for which the resulting dataset may possibly include multiple representations of the same patients for different event time indicators (e.g., patient n may have one data point corresponding no surgical excision nor antibiotic treatment, and another data point representing surgical excision perform but no treatment). (The resulting dimensions of this dataset is 288 event observations rows and 23 columns, including the newly added event indicators.)

Table 1: Original dataset Encodings.

Var	Definition
Obs	Observation number
Z1	Treatment: 0-routine bathing 1-Body cleansing
Z2	Gender (0=male 1=female)
Z3	Race: 0=nonwhite 1=white
Z4	Percentage of total surface area burned
Z5	Burn site indicator: head 1=yes, 0=no
Z6	Burn site indicator: buttock 1=yes, 0=no
Z7	Burn site indicator: trunk 1=yes, 0=no
Z8	Burn site indicator: upper leg 1=yes, 0=no
Z9	Burn site indicator: lower leg 1=yes, 0=no
Z10	Burn site indicator: respiratory tract 1=yes, 0=no
Z11	Type of burn: 1=chemical, 2=scald, 3=electric, 4=flame
T1	Time to excision or on study time
D1	Excision indicator: 1=yes 0=no
T2	Time to prophylactic antibiotic treatment or on study time
D2	Prophylactic antibiotic treatment: 1=yes 0=no
T3	Time to straphylocous aureaus infection or on study time
D3	Straphylocous aureaus infection: 1=yes 0=no

Table 2: Dataset recoded as factors/categories.

Var	New Var	Definition and Factor Levels
Obs		Observation number
T1		Time to excision or on study time
D1		Excision indicator: 1=yes 0=no
T2		Time to prophylactic antibiotic treatment or on study time
D2		Prophylactic antibiotic treatment: 1=yes 0=no
T3		Time to straphylocous aureaus infection or on study time
D3		Straphylocous aureaus infection: 1=yes 0=no
Z1	Treatment	Routine/Cleansing
Z2	Gender	Male/Female
Z3	Race	Nonwhite/White
Z4	PercentBurned	Percentage of total surface area burned
Z5	SiteHead	NotBurned/Burned
Z6	SiteButtock	NotBurned/Burned
Z7	SiteTrunk	NotBurned/Burned
Z8	SiteUpperLeg	NotBurned/Burned
Z9	SiteLowerLeg	NotBurned/Burned
Z10	SiteRespTract	NotBurned/Burned
Z11	BurnType	Chemical/Scald/Electric/Flame

Table 3: Additional Time-Dependent Covariates after Pivoting.

New Var	Definition and Factor Levels
tstart	On study time: 1=yes 0=no
tstop	Time to infection or on study time: 1=yes 0=no
tse	Time to excision or on study time: 1=yes 0=no
tpa	Time to prophylactic antibiotic treatment or on study time: 1=yes 0=no
status	Straphylocous aureous infection, on study time and stop time: 1=yes 0=no

2.3 Descriptive Statistics

Below we report various statistics presented in the original dataset. Table 4 presents means and medians of continuously distributed covariates such as percentage of total surface area burned, and Table 5 presents sample percentages for categorical covariates (e.g., percentage nonwhite/white, burn type distribution, etc.).

Table 4: Statistical descriptions of continuously distributed covariates.

Statistical Description	Value
Body Cleansing to Bathing Rate	0.5445
Female to Male Ratio	0.2207
Burn Rate on Head	0.4545
Burn Rate on Buttock	0.2273
Burn Rate on Trunk	0.8442
Burn Rate on Upper Leg	0.4091
Burn Rate on Lower Leg	0.3052
Burn Rate on Respiratory Tract	0.2922078
Surgical Excision Rate	0.6429
Prophylactic Antibiotic Rate	0.4091
Straphylocous Aureaus Infection Rate	0.3116883
Percentage Chemical Burn	5.84%
Percentage Scald Burn	11.69%
Percentage Electric Burn	7.14%
Percentage Flame Burn	75.32%

Table 5: Statistical descriptions of categorically distributed covariates.

Statistical Description	Value
Mean percentage of total surface area burned	24.69%
Mean time to surgical excision or on study time	12.11 days
Mean time to prophylactic antibiotic or on study time	16.59 days
Mean time to infection or on study time	21.80 days
Median time to surgical excision or on study time	11 days
Median time to prophylactic antibiotic or on study time	12 days
Median time to infection or on study time	17 days

It's interesting to note that on average the sample surface area burned is given as nearly 25% of a patient's body being burned. The distribution of time to surgical excision or on study time is right skewed; for antibiotic treatment or on study time it is even more right skewed; and for time to infection it is also right skewed. This commonality of right skewness for these various events has the connection that most of these events would necessarily occur sooner than later after being admitted to the hospital as an inpatient.

3 Survival Analysis and Regression

Survival analysis generally studies time-to-event processes, in particular, time-to-failure processes. Within the context of this report, "failure" is defined as occurrence of infection, or failure of infection control/prevention. This section finds that the hazards between the group that undergoes conventional bathing and the group that is given antibiotic treatment are proportional. As a follow-up, we model infection occurences between both groups using the Cox proportional hazards model over time-independent covariates, and then time-dependent covariates, concluding the section with a final model statistically corresponding to infection control/management efficacy between both modes of infection prevention.

3.1 Proportional Hazards and Survival

Before constructing a model to understand the effects of the covariates to the outcome of infection occurrence some assumptions must be verified. We verify these assumptions through four figures enumerated below as well as their interpretations.

1. Figure 1 shows Kaplan-Meier estimates of the relationship between infection occurrence (failure) between the two treatment types, routine bathing and prophylactic antibiotics. These curves effectively model from the data the probabilities of a patient in a group avoiding the complication of infection.
2. Figure 2 shows cumulative hazards computed from the survival estimates shown in Figure @ref{fig:km-label}. The cumulative hazards it is important to be able to handle cumulative hazards of two groups. The cumulative hazards effectively model the growing risk of infection for a given patient within a treatment group from start of observation of that patient.
3. Figure 3 shows the hazard ratio curve computed by taking the ratio of the cumulative hazards of the two groups from Figure 2. This figure is very important because it validates the use of the Cox proportional hazards model indicating that over time the variation of hazards between the two treatment groups remains constant and proportional, which will justify the use of a logistic regression model of survival, or infection control. A further use of the hazard ratio is that we may say that at any particular time, say at day 80, patients undergoing routine bathing are at more than two times as much risk for infection compared to the group receiving antibiotics.
4. Figure 4 is another variation for validating the proportional hazards assumption vital to using Cox proportional hazards models. The fairly parallel and nonintersecting curves of the complementary log-log plots indicates that the hazards between groups receiving routine bathing versus total body cleansing with antibiotic treatment is proportional.

Beyond these visual heuristics for verifying proportional hazards, we also see that enforcing a strict `p-value` of 0.05, the two Kaplan-Meier survival curve estimates are not significantly different, further evidencing proportional hazards. This analysis is continued with modeling in the next subsection.

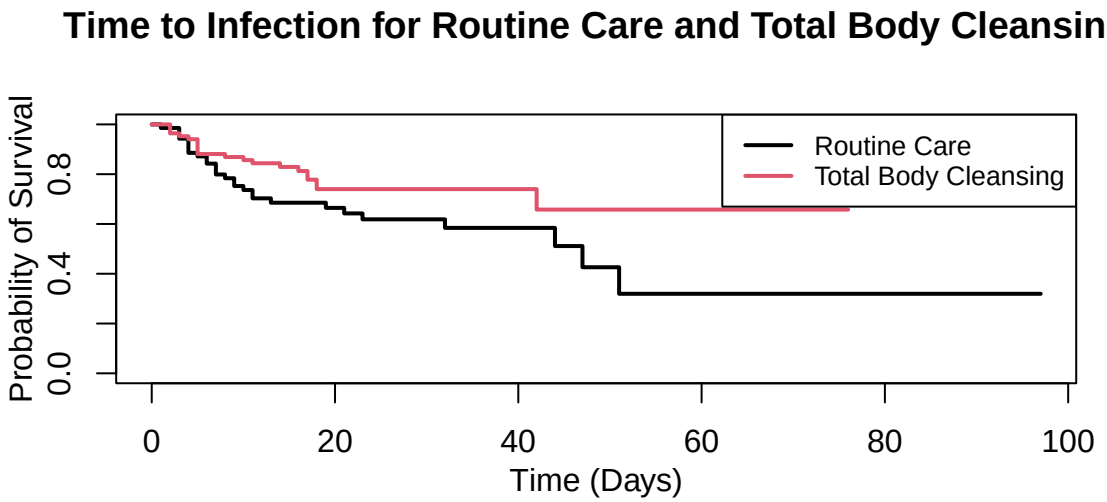

Figure 1: Kaplan-Meier estimates survival functions between two groups. The survival probability appears to be greater for the patients that receive antibiotic treatment.

3.2 Time-Independent Covariates

Out of all the covariates, it was commonly found that the covariates corresponding to burn site location were not statistically significant. Furthermore, the corresponding metric used to measure efficacy of the Cox model, the Akaike information criterion (AIC), found that the inclusion of these covariates resulted in greater AIC scores—greater AIC scores indicate higher model complexity and consequently less model effectiveness.

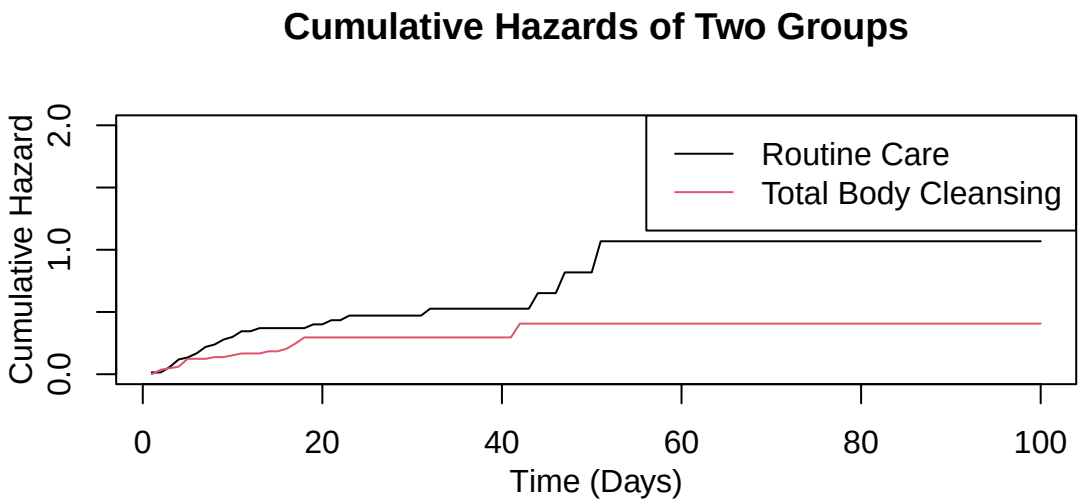

Figure 2: Cumulative hazards plot between two groups. The hazards appear to be roughly proportional between these two groups.

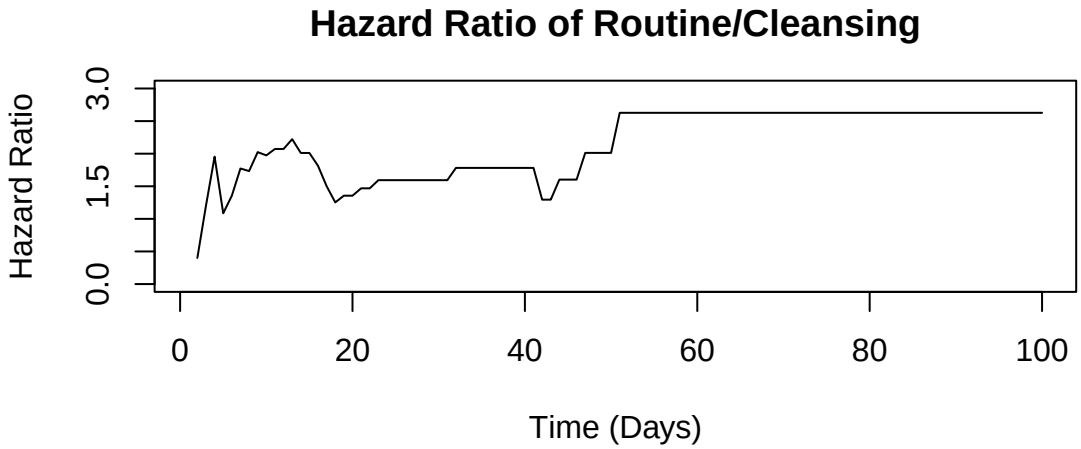

Figure 3: The hazard ratio appears to level off at around t=50, and appears to be roughly proportional.

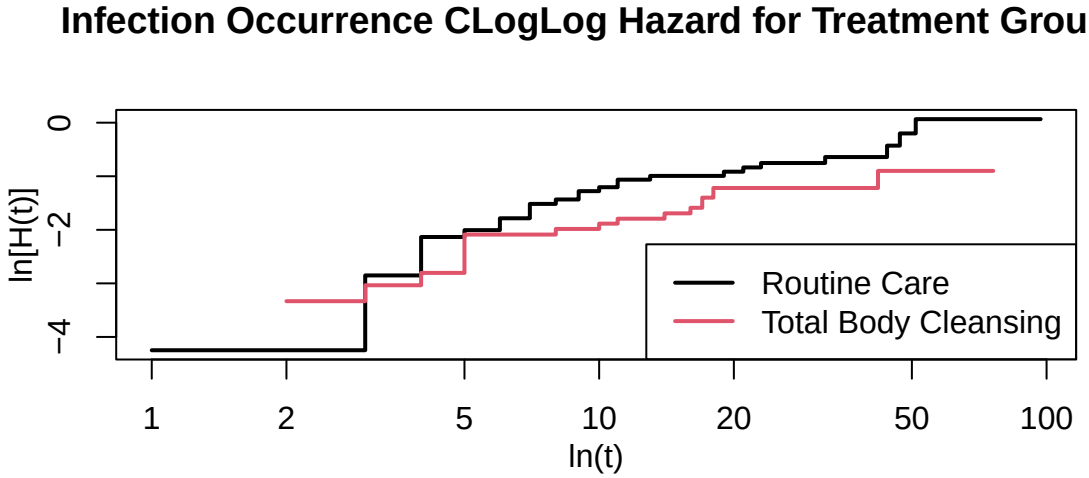

Figure 4: Complementary log-log plots of hazard functions.

(A general principle in building statistical models is the principle of parsimony: natural phenomena are best explained in as few elements as possible.)

The final time-independent covariate model that was constructed was a Cox model involving covariates of treatment type, race and burn type. In particular, treatment with antibiotic cleansing, white race and electric burn type were all found to be statistically significant covariates in the Cox model at `p-values` of `0.0436`, `0.0260` and `0.0454`, respectively. It is further worth noting that through the model selection process, gender and percent burned were found to not be statistically significant covariates.

An additional note to be made is that in the final model, we simply keep the nonsignificant factors associated with burn type: `Chemical`, `Flame` and `Scald`. Through a test via ANOVA, we find that the model including these other burn types different compared to the model including only the `Electric` burn type at statistically significant level with a `p-value` of `0.2228`.

In particular among the time-independent covariates of the Cox model as designed, the following conclusions are to be made:

1. All other variables held constant, patients going through the antibiotic cleansing have are 0.5483 times as likely to experience an infection compared to those who only undergo routine bathing. This appears to indicate that the prophylactic antibiotic if very effective at preventing infections.
2. All other variables held constant, patients that are white are 9.8182 times more likely to experience infection compared to nonwhites.
3. All other variables held constant, patients are 8.7745 times more likely to experience infection that those with other burn types.

(The above likelihoods are interpretations resulting from the exponentiated Cox regression coefficients.) With these selected covariates, the next natural step is to consider this Cox model with the addition of time-dependent covariates.

3.3 Time-Dependent Covariates

Adding new time-dependent covariates to the model from the previous subsection (`Race`, `BurnType` and `Treatment`), we find that we get most statistical significance among as many covariates as possible with just the covariates `Race`, `Treatment` and `tse` (time to surgical excision or on study time). (For `Race`, `Treatment` and `tse`, `p-values` are `0.1026`, `0.0301` and `0.0378`, respectively.) Again, it was found that building a time-dependent covariate Cox model with burn site covariates and burn percentage did not at all affect statistical significance of the other primary covariates in a statistically meaningful way.

In the final model considered with the inclusion of a time-dependent covariate, it is important to note that `treatment` was not found to be statistically significant at an α of `0.05`. However, it is included in the Cox model for two reasons: 1. it does not differ from the model excluding at a statistically significant level (`p-value=0.09971`) and 2. it's important to contextualize the main covariate of interest, which is the cleansing with the prophylactic antibiotic since the purpose of the study is to analyze and conclude its efficacy in infection control/prevention.

4 Residual Analysis

With the two Cox models—the time-independent and time-dependent covariate models—it is worth validating the proportional hazards assumptions as well as checking for any outlier in the data using various residual analysis techniques.

4.1 Time-Independent Model (Leverage Analysis)

The corresponding time-independent covariate model appears to satisfy the proportional hazards assumption vital to being able to correctly use a Cox model. In particular, Figure 5 shows that the Schoenfeld residuals are distributed in a way such that they are uniformly and randomly distributed with no strong pattern. It was found that race and burn type are statistically significant, and this is well reflected in the corresponding

Schoenfeld residual plots. On the contrary, there appears to be a slight lack of randomness and risk concordance appears to be present for the treatment type covariate, which agrees with the statistical insignificance outcome at an α of 0.05. Yet, the corresponding AICs of the model including this covariate to that of one not including it were found to not be different at a statistically significant level, and again within the practical significance of the study, the inclusion of treatment type is merited.

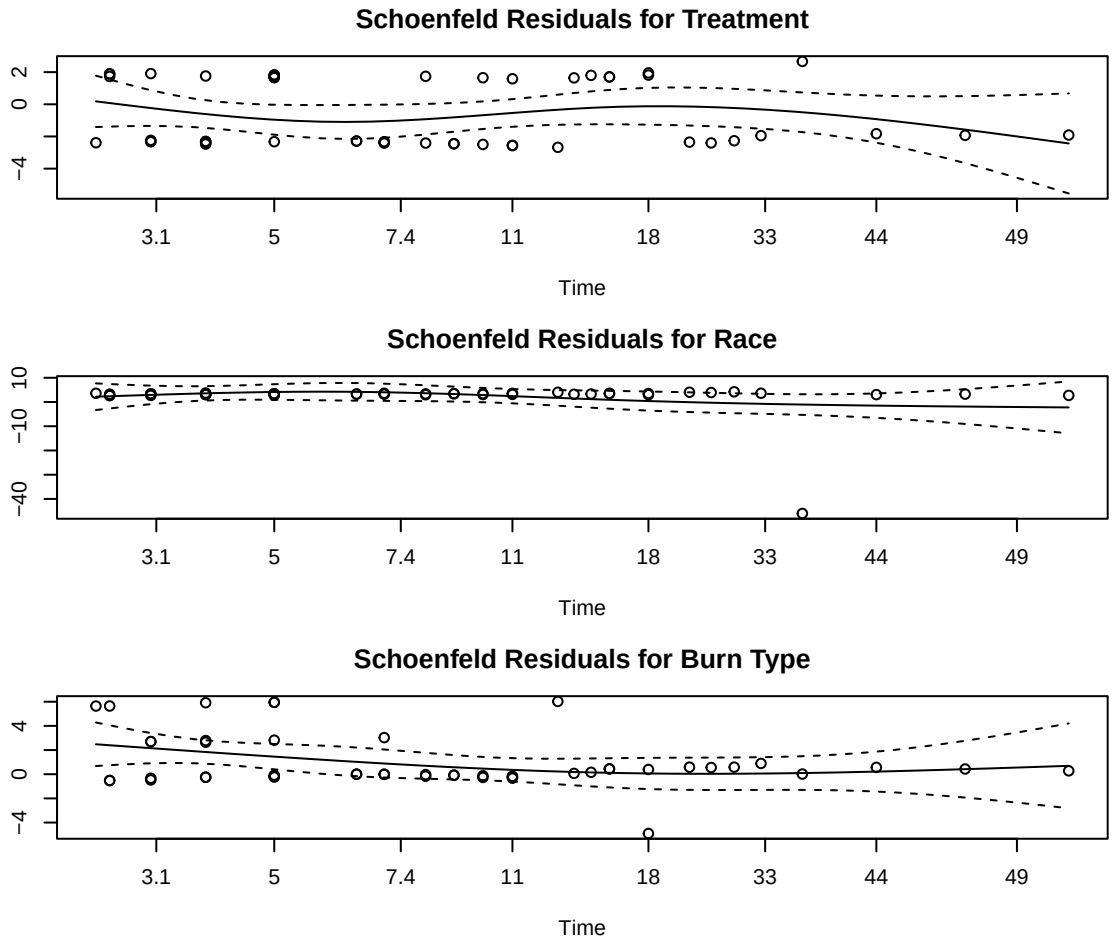

Figure 5: Schoenfeld residual plots show fairly controlled residuals. The fitted curves to the residual indicate no strong linearities or patterns. There appears to be some lack risk independence with respect to treatment type.

In addition to studying the Schoenfeld residuals, the corresponding cumulative hazard of Cox-Snell residuals seen in Figure 6 shows the overall goodness-of-fit of the Cox model that has been constructed. A strong linear 1:1 association of the cumulative hazard of the Cox-Snell residuals with intercept 0 indicates sufficient nonparametric fit from the Cox model.

Figures 7 shows Martingale and deviance residual plots. Figures 8 shows the corresponding dfbeta plots across each covariate. The Martingale plots serve as a visual heuristic for deciding if the linear model is poorly fit by any particular covariate, and in this case, there no strong lack of fit since the linear trend appears to be relative flat in the distribution of the Martingale residuals. The deviance residuals are used to identify outliers in the model by utilizing the Martingale residuals—in particular, observations with large deviance residuals are not well predicted by the model.

Additional ways to measure outliers are with dfbeta values effectively modeling the change in the coefficient across a covariate if the observation were dropped. The dfbeta values are computed across each covariate and

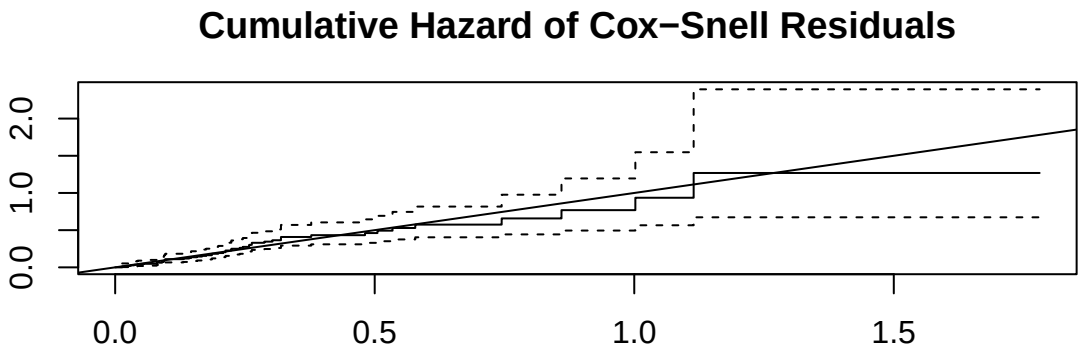

Figure 6: Cumulative Hazard plot of Cox-Snell residuals indicating sufficient goodness-of-fit.

greater dfbeta values indicate greater effects in change of the coefficients for a particular observation. Across these residual methods, we list the corresponding observations that are seen to be outliers:

1. Martingale residuals: 79, 153, 22, 41, 32, 67
2. Deviance residuals: 75, 116, 115, 58, 79, 153
3. Treatment influence: 79, 153, 116, 93, 32, 139
4. Race influence: 64, 4, 30, 39, 67, 116
5. Clinic influence: 48, 25, 36, 50, 67, 90

Among these residuals, the most important observations as measured by the largest residuals and frequency across each residual metric are 79, 153, 32, 67 and 116. We will discuss these outliers in the conclusion.

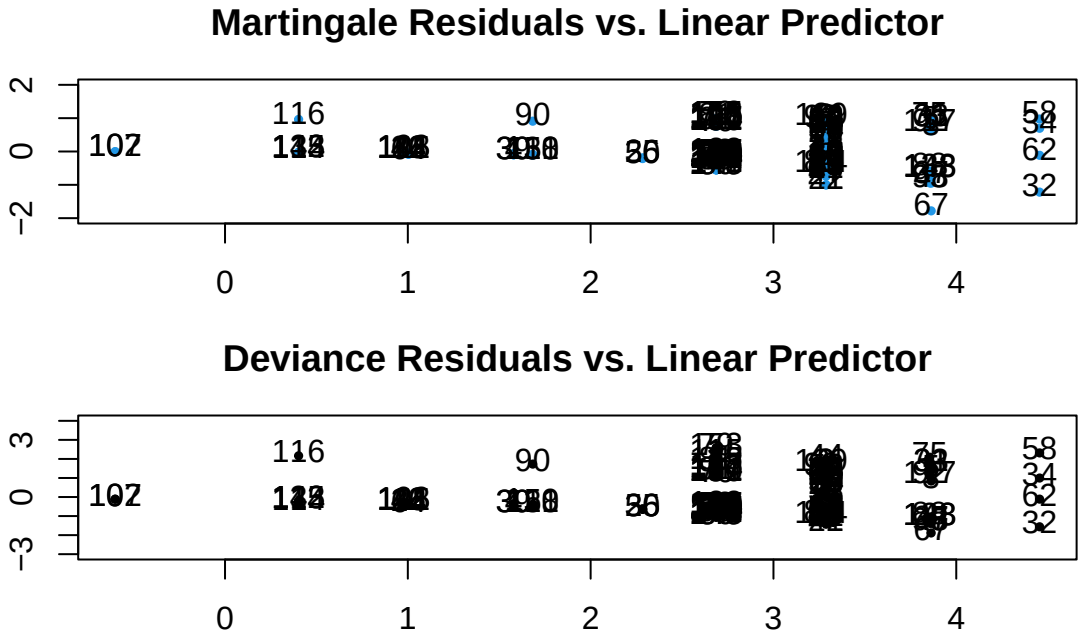

Figure 7: Martingale residuals versus linear predictor and deviance residuals versus linear predictor. Roughly linear in appearance with slope zero.

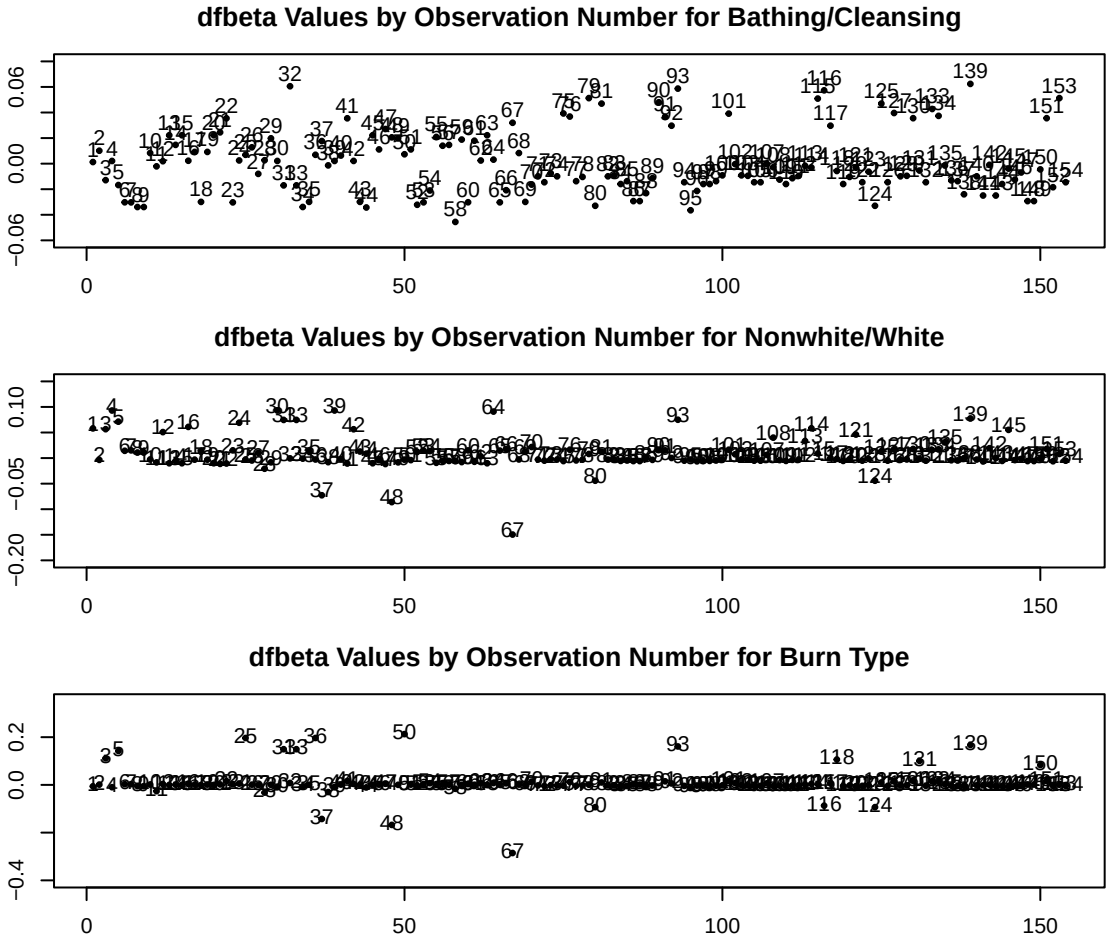

Figure 8: dfbeta values of each observation for treatment type, race and burn type.

4.2 Time-Dependent Model (Leverage Analysis)

The same type of analysis used for the time-independent model will applied exactly for the model constructed with time-dependent covariates. Figure 9 shows that the Schoenfeld residuals are distributed in a way such that they are uniformly and randomly distributed with no strong pattern. Again, race is found to be statitistically significant, and treatment not. However, we have a newly significant time-dependent covariate, `tse` or time to surgical excision. These statistical significances are well reflected in the corresponding Schoenfeld residual plots. Once again, there appears to be a slight lack of randomness and risk concordance appears to be present for the treatment type covariate, which agrees with the statistical insignificance outcome at an α of 0.05. In the context of practical significance with respect to this study we must include treatment as a covariate in this model as well.

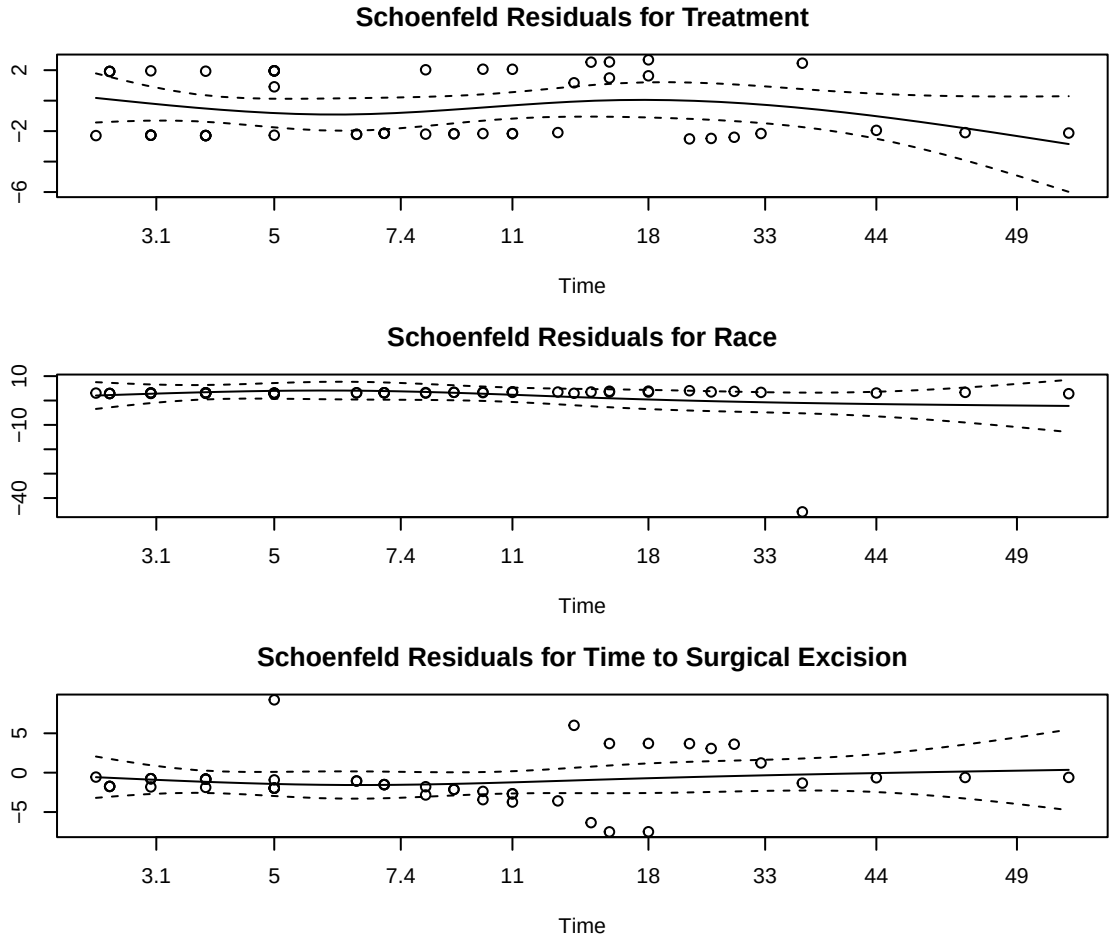

Figure 9: Schoenfeld residual plots show fairly controlled residuals. The fitted curves to the residual indicate no strong linearities or patterns. There appears to be some lack risk independence with respect to treatment type.

In Figure 10 we have a plot of the cumulative hazards of the Cox-Snell residuals with a fit of the cumulative hazard being very close to that of the line with slope 1 and intercept 0; this indicates fairly good fit w.r.t to the time-dependent covariate in the model and the other two time-independent covariates, race and burn type.

Figure 11 shows Martingale and deviance residual plots. Figure 12 shows the corresponding dfbeta plots across each covariate. Across these residual methods, we list the corresponding observations that are seen to be outliers:

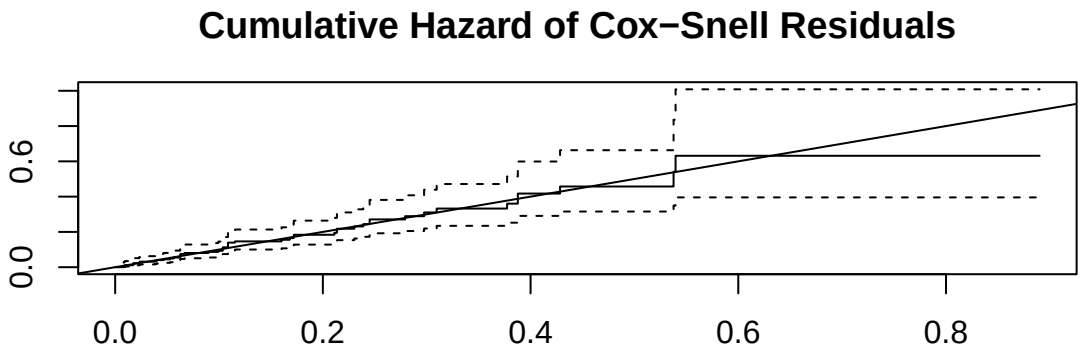

Figure 10: Cumulative Hazard plot of Cox-Snell residuals indicating sufficient goodness-of-fit.

1. Martingale residuals: 146, 149, 165, 239, 190, 218, 14, 162
2. Deviance residuals: 60, 32, 146, 149, 165, 239, 190, 218
3. Treatment influence: 58, 63, 75, 19, 238, 132, 146, 149
4. Race influence: 150, 14, 62, 170, 233, 207, 177, 12
5. Clinic influence: 49, 144, 153, 251, 78, 113, 242, 253

Among these residuals, the most important observations as measured by the largest residuals and frequency across each residual metric are 14, 146, 149 and 165. We will discuss these outliers in the conclusion.

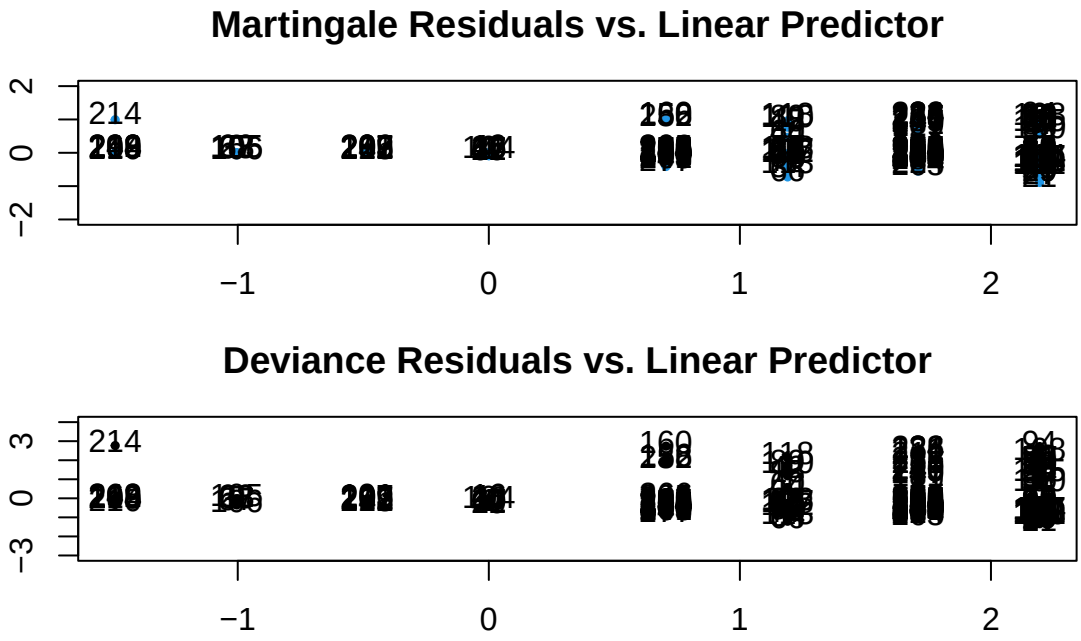

Figure 11: Martingale residuals versus linear predictor and deviance residuals versus linear predictor. Roughly linear in appearance with slope zero.

5 Conclusion

This report accomplished studying the the effects of treatments of a prophylactic antiobiotic compared to routine bathing in controlling infection occurrence. Both time-independent and time-dependent covariates

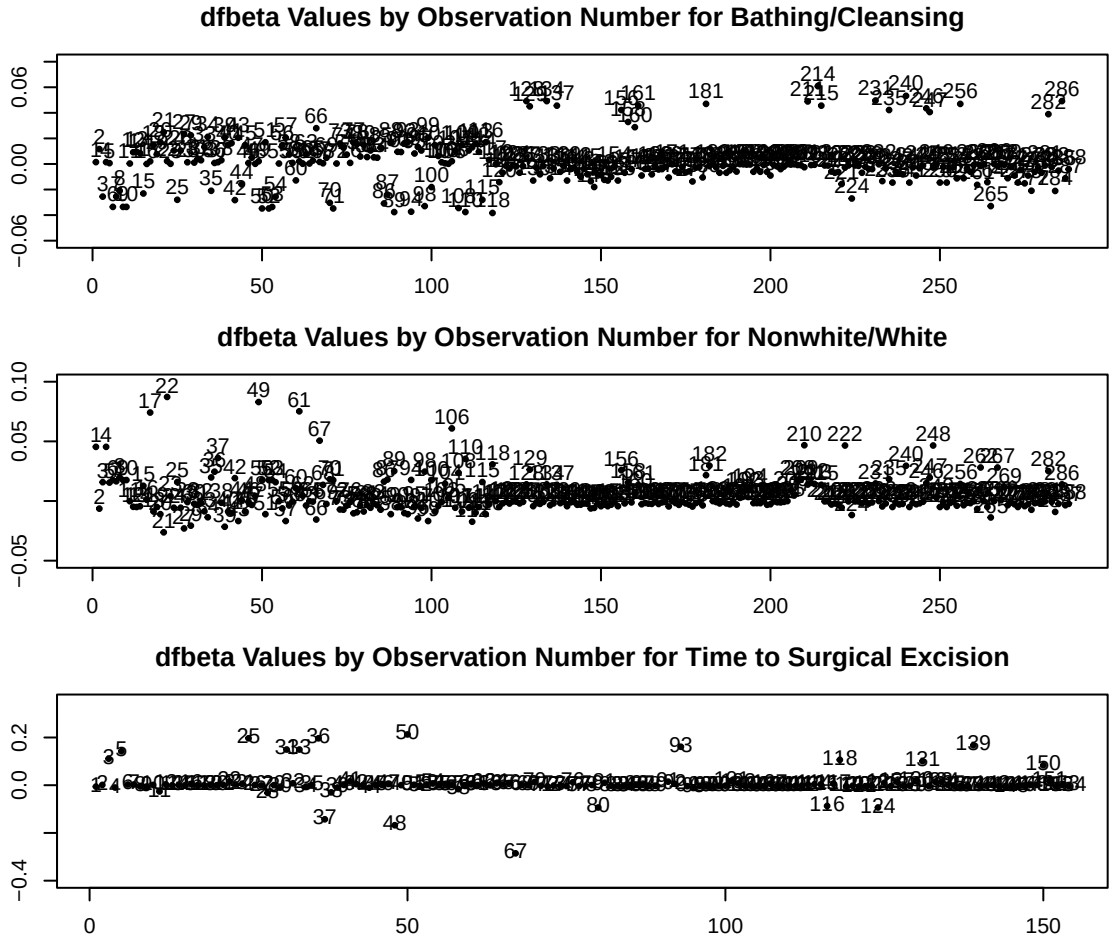

Figure 12: dfbeta values of each observation for treatment type, race and time to surgical excision (or on study time).

were used to build two different models. Table 6 shows the two models that were constructed and studied along with the covariates used for those models. Furthermore, we study various types of residuals across these models and note observations that appear to be outliers in these models.

Table 6: Models studied.

Model	Covariates Used
`Time-Independent`	`Treatment` + `Race` + `Burn Type`
`Time-Dependent`	`Treatment` + `Race` + `Time to Surgical Excision`

5.1 Discussion

First, we consider the outlier observations from the time-independent model:

- Patient 32: This person is shown to not have experienced any events of surgical excision, infection or prophylactic treatment. But this individual is interesting because he experienced electric burning which has the most statistical significance. This patient experienced no infection while receiving only routine bathing.
- Patient 67: This went through surgical excision and also prophylactic antibiotic treatment, but there time of being observed in the study for over 60 days, which is interesting since the survival probabilities decrease over time, yet this patient receiving only routine bathing managed to avoid infection.
- Patient 79: This person received antibiotic treatment, but was infected very soon at two days, and received surgical excision. Furthermore, this person was a woman who is generally underrepresented in this sample.
- Patient 116: This person is nonwhite, received cleansing, but experienced infection late into the study, furthermore, this person had a flame burn type. This might appear to be an outlier as a discordant point due to the burn type as well as the long time until infection.
- Patient 153: This individual received no treatment with only routine bathing, and no surgical excision with only a 10 percent burn rate under a flame, but they still experienced infection. This may appear to be an outlier because infection is likely to occur really only under situations of infection and the like where infections are more often nosocomial.

We now consider the outlier observations for the time-dependent model:

- Patient 14: This patient seems to be an outlier considering they survived very long without experiencing infection even after having surgical excision and cleansing.
- Patient 146: This patient does not experience an infection, but received cleansing and surgical excision with a very high burn percentage of 50%. Even though cleansing was not found to be statistically significant this patient's survival outcome suggests efficacy with the prophylactic antibiotic treatment.
- Patient 149: This patient is very similar to Patient 146 with 50% burn percentage, no infection occurrence, with antibiotic treatment.
- Patient 165: This patient appear to be an outlier since they have very low burn percentage, yet they experienced electric burns, which are found to be statistically significant with a 54% infection rate in this sample, but did not experience infection.

Between both models, these outliers are more often observations that would have been expected to experience infection but did not with reasons ranging from burn types such as electric experiencing high infection rates to a patient becoming infected very soon after surgical excision.

From analyzing these outliers the following conclusions may be made:

1. Routine bathing by itself still proves to be effective enough.
2. Patients can still get infected very early after surgical excision even with antibiotic treatment.
3. Despite high burn risks for electric burn types, or very high burn percentage, antibiotic cleansing proves to be effective despite greater risks for such patients.

An important thing to note is that the time-dependent covariate model determines treatment cleansing to be statistically significant with 45.17% less likely to experience infection. In the context of being able to control

nosocomial infections antibiotic treatment appears to be generally effective. However, when we account for the subtlety of sequential events occurring in the hospital antibiotic treatment appears to not be statistically significant, but the likelihood appears to still be lower at 61.61%. Furthermore, receiving surgical excision decreases likelihood of infection with a likelihood of infection of 36.64% compared receiving surgical excision at a later time at a statistically significant level—a natural interpretation here is that along with cleansing, it's important that the surgical excision is provided earlier when a patient is admitted for burn treatment.

5.2 Other Considerations

Antibiotic treatment is important for controlling nosocomial infections. Additionally, receiving surgical excision sooner than later is more important with respect to following up standard surgical procedures with antibiotic treatment. Interestingly, it should be emphasized from the outlier data that the actual surgical procedure for removing wound eschar should be investigated. Operating sterility and hygiene could be a very big contributor and should perhaps be also investigated with respect to nosocomial infections. Furthermore, the status of patient immediately admitted may have an impact on infection occurrence. Summarily, we can consider the following questions in the future:

1. Is operating sterility maximized? Antibiotic treatment and routine cleansing still only serve to prevent development of infections, yet they may start in operating room.
2. What is the nature of the patients infection beyond burn type and burn percentage? Burn degree is a potentially important classification that can determine infection occurrence. It can be easily suspected that higher degree burns give more opportunity for infection occurrence due to longer healing time. Burn degree furthermore may call for more extreme surgical excision and causally relate to greater risk for infections occurring potentially in the operating room, or because of the required extent of procedure for higher degree burns.
3. Is there an optimal treatment protocol with both routine bathing and cleansing? This dataset excludes the nuance in application and degree of application of the antibiotic as well as the consistency and way of performing routine bathing—there is certainly variability in both routing bathing and antibiotic application that is not accounted for from this data, especially since it would obvious that bathing and treating damaged skin is sensitive and cause pain.

With these considerations, it would be worth investigating the above points, and still we can readily conclude that prophylactic antibiotic treatment is fairly effective in preventing nosocomial infection at both a statistical and practical level.

6 Appendix

6.1 References

1. https://dmrocke.ucdavis.edu/Class/BST222.2022.Fall/BST222-Fall-2022.html
2. https://dmrocke.ucdavis.edu/Class/BST222.2022.Fall/Ichida.pdf

6.2 Code Chunks

```
knitr::opts_chunk$set(echo = TRUE)
# ============================================================================== #
# Loading Libraries and Reading Data
# ============================================================================== #
require(KMsurv)
require(survival)
data(burn)
# ============================================================================== #
# Recoded Data
# ============================================================================== #
burn1 <- burn
burn1 <- data.frame(burn1,Treatment=factor(burn1$Z1,labels=c("Routine","Cleansing")))
burn1 <- data.frame(burn1,Gender=factor(burn1$Z2,labels=c("Male","Female")))
burn1 <- data.frame(burn1,Race=factor(burn1$Z3,labels=c("Nonwhite","White")))
burn1 <- data.frame(burn1,PercentBurned=burn1$Z4)
burn1 <- data.frame(burn1,SiteHead=factor(burn1$Z5,labels=c("NotBurned","Burned")))
burn1 <- data.frame(burn1,SiteButtock=factor(burn1$Z6,labels=c("NotBurned","Burned")))
burn1 <- data.frame(burn1,SiteTrunk=factor(burn1$Z7,labels=c("NotBurned","Burned")))
burn1 <- data.frame(burn1,SiteUpperLeg=factor(burn1$Z8,labels=c("NotBurned","Burned")))
burn1 <- data.frame(burn1,SiteLowerLeg=factor(burn1$Z9,labels=c("NotBurned","Burned")))
burn1 <- data.frame(burn1,SiteRespTract=factor(burn1$Z10,labels=c("NotBurned","Burned")))
burn1 <- data.frame(burn1,BurnType=factor(burn1$Z11,labels=c("Chemical","Scald",
                                                             "Electric","Flame")))
burn.recode <- burn1[,c(1,13:29)]
# categorical PercentBurned
qs <- quantile(burn.recode$PercentBurned)
burn.recode$PercentBurnedGroup <- cut(burn.recode$PercentBurned,
                              breaks=c(qs[1]-1, qs[2], qs[3], qs[4], qs[5]+1),
                              labels=c("Q1", "Q2", "Q3", "Q4"))
burn.recode$PercentBurnedGroup
# ============================================================================== #
# Time-Varying Covariates
# ============================================================================== #
# set the final observation time, t3 = Time To infection or end Of study
burn.tdc <- burn.recode
burn.tdc <- tmerge(burn.tdc,burn.tdc,id=Obs,tstop=T3)
# set time until surgical excision tse = time To surgical excision
burn.tdc <- tmerge(burn.tdc,burn.recode,id=Obs,tse=tdc(T1))
# set time until prophylactic treatment tpa = time to prophylactic antiobiotic
burn.tdc <- tmerge(burn.tdc,burn.recode,id=Obs,tpa=tdc(T2))

#status only = 1 if at end of T3 and not censored
# D3 = Disease Free Survival Indicator 1-relapsed with infection, 0-Alive and Disease Free
status <- as.integer(with(burn.tdc,(tstop==T3 & D3)))
```

```
#put together
burn.tdc <- data.frame(burn.tdc,status)
# ============================================================================ #
# Sample Statistics and Descriptions
# ============================================================================ #
apply(burn[c('Z1','Z2','Z3','Z5','Z6','Z7','Z8','Z9','Z10')], 2,mean)
apply(burn[c('D1','D2','D3')], 2,mean)
apply(burn[c('T1','T2','T3')],2,mean)
apply(burn[c('T1','T2','T3')],2,median)
# ============================================================================ #
# KM curves over treatment
# ============================================================================ #
par(mar = c(3, 3, 3, 2))
burn.surv <- Surv(time=burn.recode$T3, event=burn.recode$D3)
plot(survfit(burn.surv~Treatment,data=burn.recode),col=1:2,lwd=2)
title("Time to Infection for Routine Care and Total Body Cleansing",
      ylab="Probability of Survival",xlab="Time (Days)", line=2)
legend("topright",c("Routine Care","Total Body Cleansing"),col=1:2,lwd=2,cex=0.8)
# ============================================================================ #
# cumulative hazards
# ============================================================================ #
par(mar = c(3, 3, 3, 2))
# Nelson-Aalen curve estimates
NAcurves <- survfit(burn.surv~Treatment, type="fleming-harrington",data=burn.recode)
timevec <- 1:100 #vector of time points
sf1 <- stepfun(NAcurves[1]$time,c(1,NAcurves[1]$surv))
sf2 <- stepfun(NAcurves[2]$time,c(1,NAcurves[2]$surv))
#now we can find the cumulative hazards
cumhaz1 <- -log(sf1(timevec))
cumhaz2 <- -log(sf2(timevec))
plot(timevec,cumhaz1,,ylab="",xlab="",type="l",ylim=c(0,2), col=1)
lines(timevec,cumhaz2,type="l",ylim=c(0,6), col=2)
legend("topright",c("Routine Care", "Total Body Cleansing"),col=1:2,lwd=1)
title("Cumulative Hazards of Two Groups",xlab="Time (Days)",ylab="Cumulative Hazard", line=2)
# ============================================================================ #
# Hazard Ratios
# ============================================================================ #
par(mar = c(4, 4, 3, 2))
plot(timevec, cumhaz1/cumhaz2, type="l", ylim=c(0,3), ylab="Hazard Ratio", xlab="Time (Days)")
title("Hazard Ratio of Routine/Cleansing")
# ============================================================================ #
# checking survdiff
# ============================================================================ #
burn.survdiff <- survdiff(burn.surv~Treatment,data=burn.recode)
print(burn.survdiff)
print(summary(burn.survdiff))
# ============================================================================ #
# complementary log-log
# ============================================================================ #
par(mar = c(3, 3, 3, 2))
plot(NAcurves,col=1:2,lwd=2,fun="cloglog")
title("Infection Occurrence CLogLog Hazard for Treatment Groups", ylab = "ln[H(t)]",
      xlab = "ln(t)", line=2)
```

```
legend("bottomright",c("Routine Care","Total Body Cleansing"),col=1:2,lwd=2)
# ============================================================================ #
# model checking time-independent
# ============================================================================ #
# basic Cox model
burn.cox <- coxph(burn.surv~Treatment,data=burn.recode)
# summary of coefficients
summary(burn.cox)
# model checking
burn1.cox <- coxph(burn.surv~Treatment+Gender+Race+PercentBurned+BurnType
                   +SiteHead+SiteButtock+SiteTrunk+SiteUpperLeg+SiteLowerLeg
                   +SiteRespTract, data=burn.recode)
summary(burn1.cox)
# AIC
drop1(burn1.cox, test="Chisq")
# ============================================================================ #
# final model build, time independent
# ============================================================================ #
# refined model
burn2.cox <- coxph(burn.surv~Treatment+Race+Gender+BurnType+PercentBurned, data=burn.recode)
burn3.cox <- coxph(burn.surv~Treatment+Race+BurnType, data=burn.recode)
summary(burn2.cox)
summary(burn3.cox)

drop1(burn2.cox, test="Chisq")
drop1(burn3.cox, test="Chisq")

burn3b <- coxph(burn.surv~Treatment+Race+factor(BurnType=="Electric"), data=burn.recode)
AIC(burn3.cox)
AIC(burn3b)
anova(burn3.cox,burn3b,test="Chisq")
# ============================================================================ #
# time-dependent covariate model
# ============================================================================ #
# tdc surv object
burn.tdc.surv <- Surv(time=burn.tdc$tstart,time2=burn.tdc$tstop,event=burn.tdc$status,type="counting")
# initial model
burn.tdc.cox <- coxph(burn.tdc.surv ~ Treatment+Race+BurnType+tse+tpa, data=burn.tdc)
summary(burn.tdc.cox)
# refined model
burn.tdc.coxRef <- coxph(burn.tdc.surv ~ Treatment+Race+tse, data=burn.tdc)
summary(burn.tdc.coxRef)
# dummy model
burn.tdc.coxRefd <- coxph(burn.tdc.surv ~ Race+tse, data=burn.tdc)
anova(burn.tdc.coxRef, burn.tdc.coxRefd)
# ============================================================================ #
# building zph residual objects
# ============================================================================ #
# zph objects (burn3.cox, burn.tdc.CoxRef)
burn3.zph <- cox.zph(burn3.cox)
burntdc.zph <- cox.zph(burn.tdc.coxRef)
# ============================================================================ #
# schoenfeld residual plots
```

```r
# ============================================================================ #
par(mfrow=c(3,1), mar = c(4, 3, 3, 1))
plot(burn3.zph[1], main = "Schoenfeld Residuals for Treatment")
plot(burn3.zph[2], main = "Schoenfeld Residuals for Race")
plot(burn3.zph[3], main = "Schoenfeld Residuals for Burn Type")
# ============================================================================ #
# Cox-Snell residual plots
# ============================================================================ #
par(mfrow=c(1,1), mar = c(2, 3, 3, 1))
#fit martingale for full model
burn.mart <- residuals(burn3.cox,type="martingale")
#find cox-snell residuals: martingales subtracted from event indicator
burn.cs <- burn.recode$D3 - burn.mart
#cumulative hazard of CS residuals
surv.csr <- survfit(Surv(burn.cs,burn.recode$D3)~1,type="fleming-harrington")
# plotting cumulative hazard of cs residuals
plot(surv.csr,fun="cumhaz")
abline(0,1)
title("Cumulative Hazard of Cox-Snell Residuals")
# ============================================================================ #
# other residual metrics
# ============================================================================ #
#fit residuals: martingale, deviance, and df beta
burn.mart <- residuals(burn3.cox,type="martingale")
burn.dev <- residuals( burn3.cox,type="deviance")
burn.dfb <- residuals( burn3.cox,type="dfbeta")
#find linear predictor
burn3.preds <- predict(burn3.cox)
n <- length(sort(abs(burn3.preds)))
# ============================================================================ #
# plotting ordered important observations
# ============================================================================ #
print(sort(abs(burn.mart))[(n-5):n])
print(sort(abs(burn.dev))[(n-5):n])
print(sort(abs(burn.dfb[,1]))[(n-5):n])
print(sort(abs(burn.dfb[,2]))[(n-5):n])
print(sort(abs(burn.dfb[,3]))[(n-5):n])
unusuals.burn3 <- c(79, 153, 32, 67, 116)
burn.recode[unusuals.burn3,]
# ============================================================================ #
# plot deviance residuals, martingale residuals
# ============================================================================ #
par(mfrow=c(2,1), mar = c(2, 3, 3, 1))
plot(burn3.preds,burn.mart,xlab="Linear Predictor",
ylab="Martingale Residual", ylim = c(-2,2), pch = 19, cex = 0.5, col=12)
text(burn3.preds,burn.mart+0.2, labels = rownames(burn.recode))
title("Martingale Residuals vs. Linear Predictor")
plot(burn3.preds,burn.dev,xlab="Linear Predictor",ylab="Deviance Residual",
ylim = c(-3,4), pch = 19, cex = 0.5)
text(burn3.preds,burn.dev+0.23, labels = rownames(burn.recode))
title("Deviance Residuals vs. Linear Predictor")
# ============================================================================ #
# plotting dfbeta values for time-independent covariates
```

```
# ========================================================================== #
par(mfrow=c(3,1), mar = c(2, 3, 3, 1))
plot(burn.dfb[,1],xlab="Observation Number",ylab="dfbeta for Treatment",
ylim=c(-.06,.08), pch = 19, cex = 0.5)
text(burn.dfb[,1]+.01, labels = rownames(burn.recode))
title("dfbeta Values by Observation Number for Bathing/Cleansing")
plot(burn.dfb[,2],xlab="Observation Number",ylab="dfbeta for Race",
ylim=c(-.2,.15), pch = 19, cex = 0.5)
text(burn.dfb[,2]+.01, labels = rownames(burn.recode))
title("dfbeta Values by Observation Number for Nonwhite/White")
plot(burn.dfb[,3],xlab="Observation Number",ylab="dfbeta for Burn Type",
ylim=c(-.4,.35), pch = 19, cex = 0.5)
text(burn.dfb[,3]+.01, labels = rownames(burn.recode))
title("dfbeta Values by Observation Number for Burn Type")
# ========================================================================== #
# burn tdc residual object
# ========================================================================== #
# zph objects (burn3.cox, burn.tdc.CoxRef)
burntdc.zph <- cox.zph(burn.tdc.coxRef)
# ========================================================================== #
# plotting schoenfeld residuals for tdc model
# ========================================================================== #
par(mfrow=c(3,1), mar = c(4, 3, 3, 1))
plot(burntdc.zph[1], main = "Schoenfeld Residuals for Treatment")
plot(burntdc.zph[2], main = "Schoenfeld Residuals for Race")
plot(burntdc.zph[3], main = "Schoenfeld Residuals for Time to Surgical Excision")
# ========================================================================== #
# plotting Cox-Snell for tdc model
# ========================================================================== #
par(mfrow=c(1,1), mar = c(2, 3, 3, 1))
#fit martingale for full model
burn.martTDC <- residuals(burn.tdc.coxRef,type="martingale")
#find cox-snell residuals: martingales subtracted from event indicator
burn.csTDC <- burn.tdc$status - burn.martTDC
#cumulative hazard of cs residuals
surv.csr <- survfit(Surv(burn.csTDC, burn.tdc$status)~1,type="fleming-harrington")
# plotting cumulative hazard of cs residuals
plot(surv.csr,fun="cumhaz")
abline(0,1)
title("Cumulative Hazard of Cox-Snell Residuals")
# ========================================================================== #
# residual metrics for tdc model
# ========================================================================== #
#fit residuals: martingale, deviance, and df beta
burntdc.mart <- residuals(burn.tdc.coxRef,type="martingale")
burntdc.dev <- residuals( burn.tdc.coxRef,type="deviance")
burntdc.dfb <- residuals( burn.tdc.coxRef,type="dfbeta")
#find linear predictor
burntdc.preds <- predict(burn.tdc.coxRef)
n <- length(sort(abs(burntdc.preds)))
# ========================================================================== #
# observing greatest outliers
# ========================================================================== #
```

```
print(sort(abs(burntdc.mart))[(n-7):n])
print(sort(abs(burntdc.dev))[(n-7):n])
print(sort(abs(burntdc.dfb[,1]))[(n-7):n])
print(sort(abs(burntdc.dfb[,2]))[(n-7):n])
print(sort(abs(burntdc.dfb[,3]))[(n-7):n])
unusuals.burntdc <- c(14, 146, 149, 165)
burn.tdc[unusuals.burntdc,]
# ========================================================================== #
# plotting residuals and deviance residuals
# ========================================================================== #
par(mfrow=c(2,1), mar = c(2, 3, 3, 1))
plot(burntdc.preds,burntdc.mart,xlab="Linear Predictor",
ylab="Martingale Residual", ylim = c(-2,2), pch = 19, cex = 0.5, col=12)
text(burntdc.preds,burntdc.mart+0.2, labels = rownames(burn.tdc))
title("Martingale Residuals vs. Linear Predictor")
plot(burntdc.preds,burntdc.dev,xlab="Linear Predictor",ylab="Deviance Residual",
ylim = c(-3,4), pch = 19, cex = 0.5)
text(burntdc.preds,burntdc.dev+0.23, labels = rownames(burn.tdc))
title("Deviance Residuals vs. Linear Predictor")
# ========================================================================== #
# plotting tdc model dfbeta values
# ========================================================================== #
par(mfrow=c(3,1), mar = c(2, 3, 3, 1))
plot(burntdc.dfb[,1],xlab="Observation Number",ylab="dfbeta for Treatment",
ylim=c(-.06,.08), pch = 19, cex = 0.5)
text(burntdc.dfb[,1]+.01, labels = rownames(burn.tdc))
title("dfbeta Values by Observation Number for Bathing/Cleansing")
plot(burntdc.dfb[,2],xlab="Observation Number",ylab="dfbeta for Race",
ylim=c(-.05,.1), pch = 19, cex = 0.5)
text(burntdc.dfb[,2]+.01, labels = rownames(burn.tdc))
title("dfbeta Values by Observation Number for Nonwhite/White")
plot(burn.dfb[,3],xlab="Observation Number",ylab="dfbeta for Time to Surgical Excision",
ylim=c(-.4,.35), pch = 19, cex = 0.5)
text(burn.dfb[,3]+.01, labels = rownames(burn.recode))
title("dfbeta Values by Observation Number for Time to Surgical Excision")
burn.recode[unusuals.burn3,]
burn.tdc[unusuals.burntdc,]
# ========================================================================== #
# checking if we can include continuous covariates
# ========================================================================== #
mres2 <- residuals(burn3.cox, type = "martingale")
# martingale residual for categorical
plot(burn.recode$PercentBurnedGroup,mres2,xlab="Methadone Dosage (mg)", ylab="Martingale Residuals")
lines(lowess(burn.recode$PercentBurnedGroup,mres2))
title("Martingale Residuals vs. Methadone Dosage")
# martingale residual for categorical
plot(burn.recode$PercentBurned,mres2,xlab="Methadone Dosage (mg)", ylab="Martingale Residuals")
lines(lowess(burn.recode$PercentBurned,mres2))
title("Martingale Residuals vs. Methadone Dosage")
```

6.3 Suggested Prompts

1. Plot the Kaplan-Meier curves for the treated and untreated patients and use `survdiff` to test for whether the curves are different

---

2. Plot the cumulative hazards vs. time and the complimentary log-log survival vs. log time.

---

3. Construct Cox model using only the time-independent predictors, Maybe start with one using only Treatment. Decide if the burn site variables will be separately included after analysis or included or excluded as a group. Note that this is not a factor, because a patient may have burns at many sites. The respiratory tract burn site variable is different from the others since it does not focus on skin.

---

4. Run the usual suite of model checking methods and report any interesting findings. Possibly alter the model as a result.

---

5. Construct the data set with the the time-dependent covariates for surgical excision and prophylactic antibiotic treatment and find a good model which includes useful time-dependent covariates as well as useful time-independent ones.

---

6. Run the usual suite of model checking methods and report any interesting findings. Possibly alter the model as a result

---

7. Interpret the results and comment on the implications for clinical management. Note that, from other studies, for the endpoint of survival (an outcome not included in this data set), burn percentage and burn degree (first, second, third) are quite important, but the first one may or may not be important for infection control, and information on the second factor is not included in the data set.