

# BIOS 522: Project 1

Falcons Group

2020-09-14

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Methods</b>	<b>2</b>
<b>3</b>	<b>Results</b>	<b>2</b>
3.1	Linear Regression . . . . .	2
3.2	Logistic Regression . . . . .	4
<b>4</b>	<b>Discussion</b>	<b>5</b>
4.1	Critique . . . . .	5
4.2	Parametric Survival Analysis . . . . .	6
<b>5</b>	<b>Conclusions</b>	<b>7</b>
<b>6</b>	<b>Appendix</b>	<b>8</b>
6.1	Estimates . . . . .	8
6.2	Standard errors . . . . .	8
6.3	Right Censored Data . . . . .	9
<b>7</b>	<b>Supplementary Tables and Figures</b>	<b>10</b>
7.1	Linear Regression Models . . . . .	10
7.2	Logistic Regression Models . . . . .	13
7.3	Parametric Survival Analysis . . . . .	15

# 1 Introduction

From 1974 to 1984, a randomized placebo-controlled Mayo Clinic trial was conducted to investigate the effect of the drug D-penicillamine on the mortality of 424 patients diagnosed with primary biliary cirrhosis (PBC) of the liver. Right censorship due to surviving or receiving a liver transplant is present in the data. More details can be found in the original text by Fleming and Harrington that includes this dataset (Counting Processes and Survival Analysis, Wiley, 1991). In this report we develop a regression model to describe the time to death as a function of treatment, accounting for age and serum bilirubin levels. We first implement a suggestion from a fictional collaborator, Dr. Blum, to use linear and logistic regression models to analyze the survival data. At the end, we conduct a log-linear Weibull survival model to address the shortcomings of the linear and logistic approaches.

## 2 Methods

Our goal is to develop a model to quantify the relationship between time to death and certain predictors in our data. The effect of the drug D-penicillamine on time to PBC death is of particular interest. This model can then be used for statistical inference of the model parameters, as well as prediction of time to death for future observations. We take four different approaches to analyze the survival data.

1. A linear model that excludes all censored observations from the analysis
2. A linear model but treats all censored times as death times
3. A logistic regression, where the binary outcome is either death, or censorship
4. A log-linear Weibull regression that addresses the censoring nature of the survival data in the likelihood

Each regression was run multiple times on the same sets of predictors. The sets consist of three univariate analyses on the predictors administered drug type (D-penicillamine vs. Placebo), age in years, and level of serum bilirubin in mg/dL as a categorical variable, and one multivariate analysis including all three predictors. The model with the lowest Akaike information criterion (AIC), a measure of in-sample fit that adjusts for model complexity, is selected as the best performing model for describing the dataset and future predictive accuracy.

## 3 Results

### 3.1 Linear Regression

Here we will analyze the results of the initial linear regression on the survival data under two scenarios:

- i) complete exclusion of censored data
- ii) treating the time of censorship as time of death.

Our initial analysis of the data after input and cleaning shows that we have 187 patients alive at the end of the study and 125 patients who died at the end of the study ( $\delta(0 = \text{alive}, 1 = \text{dead})$ ). Our initial analysis also shows the distribution of serum bilirubin levels impacting survival among the patients with 116 patients having less than 1.1 mg/dL serum bilirubin, 113 patients having between 1.1 to 3.3 mg/dL of serum bilirubin, and 83 patients having greater than 3.3 mg/dL.

Our linear models use the following general format,

$$Y = \beta_0 + \beta^\top \mathbf{X} + \epsilon, \quad \text{where } \epsilon \sim N(0, \sigma^2)$$

The coefficients in linear regression represent how the response variable changes with a unit increase for the given parameter, given all other parameters remain constant. Specific values and interpretations of regression coefficients will be presented in Section 7: Supplementary Tables and Figures.

### 3.1.1 Excluding censored observations

**3.1.1.1 Univariate regressions on treatment (drug), age (in years), and serum bilirubin (mg/dL)** Univariate linear regression models were separately fit to model survival time using treatment (drug), age (in years), and serum bilirubin (mg/dL) as predictors. For these models, censored observations were excluded from the data set and we were left with 125 observations. Supplementary Tables 1-3 organize the results of these regressions.

We determine that the linear regression coefficients  $\beta_{\text{drug}}$  and  $\beta_{\text{age}}$  for treatment and age, respectively, both do not represent the difference in mean survival time between the placebo and treatment groups with statistical significance, when all other factors are held constant. We do not have enough evidence to conclude that either of the covariates are statistically significant predictors of survival due to their p-values being larger than the prespecified confidence level of  $\alpha = 0.05$  (treatment p-value: 0.6474; age p-value: 0.05945). This conclusion is further supported by the adjusted  $R^2$  values for each regression being close to 0 (treatment adj.  $R^2 = -0.00641$ ; age adj.  $R^2 = 0.02069$ ).

We determine that the linear regression coefficient  $\beta_{\text{SB}}$  for serum bilirubin levels does represent the difference in mean survival time between the placebo and treatment groups with statistical significance, when all other factors are held constant. In other words, we do have enough evidence to conclude that serum bilirubin level is a statistically significant predictor of survival due to our p-value being smaller than our alpha value ( $3.712 \times 10^{-6} < 0.05$ ) when we control for treatment effect, and age. Due to this, serum bilirubin levels (especially when the concentration is over 3.3 mg/dL) is the best predictor of survival compared to treatment and age. Our  $R^2$  value is low (Adjusted  $R^2 = 0.172$ ) which suggests that the model has very small amount of explanation towards our response variable. Therefore, it is strongly recommended to use a different modeling technique to predict survival due to survival analysis data not being truly linear.

**3.1.1.2 Regression using all three covariates** From Supplementary Table 4, we determine that all three linear regression coefficients ( $\beta_{\text{drug}} + \beta_{\text{age}} + \beta_{\text{SB}}$ ) taken together does represent the difference in mean survival time between the placebo group and treatment group with statistical significance. In other words, we do have enough evidence to conclude that the all three covariates can be used together with statistical significance to predict survival due to our p-value being smaller than our alpha value ( $1.533 \times 10^{-6} < 0.05$ ). Our  $R^2$  value is low (Adjusted  $R^2 = 0.2099$ ) which suggests that the model has very small amount of explanation towards our response variable. Therefore, it is strongly recommended to use a different modeling technique to predict survival due to survival analysis data not being truly linear.

If the aim is to reduce the number of covariates in the linear regression model to strive for the most parsimonious model, then it is recommended to remove the treatment (drug) and potentially the age covariates. Treatment has not shown to be significant in any of the models that have been fit so we can remove it from the final model equation. Age by itself was not statistically significant but taken with the other covariates it is significant so this would require more investigation. The final model would be written as  $Y = \beta_0 + \beta_{\text{SB}}X_{\text{SB}} + \epsilon$ , where  $\epsilon \sim N(0, \sigma^2)$  and the intercept is time.

### 3.1.2 Treating censored times as death times

**3.1.2.1 Univariate regressions on treatment (drug), age (in years), and serum bilirubin (mg/dL)** Univariate linear regression models were separately fit to model survival time using treatment (drug), age (in years), and serum bilirubin (mg/dL) as predictors. For these models, censored observations were treated like death times and we had 312 observations. Supplementary Tables 5-7 organize the results of these regressions.

We determine that the linear regression coefficient  $\beta_{\text{drug}}$  for the treatment does not represent the difference in mean survival time between the placebo group and treatment group with statistical significance, when we treat censored times as death times and we hold all other factors constant. Even after we treat censored times as death times, we do not have enough evidence to conclude that the covariate of treatment (drug) effect is a statistically significant predictor of survival due to our p-value being larger than our alpha value ( $0.883 > 0.05$ ) when we control for age, and serum bilirubin levels. Furthermore, our  $R^2$  value is close to 0 (Adjusted  $R^2 = 0.0147$ ) which suggests that the model has very small amount of explanation towards our response variable. Therefore, it is strongly recommended to use a different modeling technique to predict survival.

We determine that the linear regression coefficients  $\beta_{\text{age}}$  and  $\beta_{\text{SB}}$  for age and serum bilirubin levels, respectively, do represent the difference in mean survival time between the placebo group and treatment group with statistical significance, when including censored observations and holding all other factors constant. There is enough evidence to conclude that both covariates of age and serum bilirubin levels are statistically significant predictors of survival due to their p-value being smaller than our prespecified confidence level of  $\alpha = 0.05$  (age p-value: 0.0182; serum bilirubin p-value:  $2.2 \times 10^{-16}$ ) when we control for treatment, and age. However, both adjusted  $R^2$  values are low (age adj.  $R^2 = 0.0147$ ; serum bilirubin adj  $R^2 = 0.2176$ ) which suggests that both models have little explanatory power towards our response variable. Therefore, it is strongly recommended to use a different modeling technique to predict survival.

**3.1.2.2 Regression on all three covariates** From Supplementary Table 8, we determine that all three linear regression coefficients ( $\beta_{\text{drug}} + \beta_{\text{age}} + \beta_{\text{serum bilirubin}}$ ) taken together does represent the difference in mean survival time between the placebo group and treatment group with statistical significance when we treat censored times as death times. In other words, after we treat censored times as death times, we do have enough evidence to conclude that the all three covariates can be used together with statistical significance to predict survival due to our p-value being smaller than our alpha value ( $2.2 \times 10^{-16} < 0.05$ ). Our  $R^2$  value is low (Adjusted  $R^2 = 0.2283$ ) which suggests that the model has very small amount of explanation towards our response variable. Due to this being the largest  $R^2$  value from all of the previous models, this is the best performing linear regression model in terms of capturing as much explanation of the response with the covariates.

If the aim is to reduce the number of covariates in the linear regression model to strive for the most parsimonious model, then it is recommended remove the treatment (drug) covariate. Treatment has not shown to be significant in any of the models that treats censored times as death times so we can remove it from the final model equation. The final model would be written as  $Y = \beta_0 + \beta_{\text{age}}X_{\text{age}} + \beta_{\text{serum bilirubin}}X_{\text{serum bilirubin}} + \epsilon$ , where  $\epsilon \sim N(0, \sigma^2)$  and the intercept is time.

## 3.2 Logistic Regression

As in Section 4.1, we will analyze the results of the logistic regression models on the survival data. Our logistic regression models use the general format,

$$Y_i \sim \text{Bernoulli}(p_i) \quad \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \sum_{k=1}^p \beta_k x_{ki}.$$

In logistic regression, a coefficient value represents the change in the log odds when the covariate is increased by 1. Specific values of coefficients and their interpretations for each model will be discussed in Section 7: Supplementary Tables and Figures.

### 3.2.1 Univariate regressions on treatment (drug), age (in years), and serum bilirubin (mg/dL)

Univariate logistic regression models were separately fit to model survival time using treatment (drug), age (in years), and serum bilirubin (mg/dL) as predictors. For these models, censored observations were treated

like death times and we had 312 observations. Supplementary Tables 9-11 organize the results of these regressions.

When all other factors are held constant, we determine that the logistic regression coefficient for treatment does not represent the change in the log odds with statistical significance. In other words, we do not have enough evidence to conclude that treatment is a statistically significant predictor of survival time when modeled with logistic regression due to our p-value being larger than our prespecified confidence level of  $\alpha = 0.05$  (treatment p-value: 0.695) when we control for age and serum bilirubin levels. The model produces the largest Akaike's information criteria ( $AIC = 423.97$ ) out of the models that we fit for logistic regression.

We determine that the logistic regression coefficients for age and serum bilirubin levels do represent a statistically significant change in log odds when all other factors are held constant. The effects of age and serum bilirubin levels, when modeled with logistic regression, produce p-values smaller than the prespecified confidence level (age p-value:  $1.81 \times 10^{-5}$ ; serum bilirubin p-value:  $6.84 \times 10^{-6}$ ) when we control for other covariates. Both models produce large AIC values when fit (age AIC: 404.3; serum bilirubin AIC: 354.73), thus we suggest alternative methods to model this interaction.

### 3.2.2 Regression on all three covariates

From Supplementary Table 12, we determine that the logistic regression coefficient for all of the predictors taken together does represent the change in the log odds with statistical significance. In other words, we do have enough evidence to conclude that all three covariates (when modeled with logistic regression) can be used together with statistical significance to predict survival due to our p-values being smaller than our alpha value of 0.05 (except the covariate for treatment). Our  $R^2$  value is low ( $\text{Adjusted } R^2 = 0.2283$ ) which suggests that the model has very small amount of explanation towards our response variable. This model produces the lowest Akaike's information criterion ( $AIC = 335.33$ ) when fit for logistic regression so we will choose this model as the final model out of all the models previous stated to capture as much explanation of the response with the covariates.

Striving for a more parsimonious model, it is recommended to remove treatment as a covariate. Treatment has not shown to be significant in any of the models (linear regression or logistic regression) so it can be removed from the final model equation.

## 4 Discussion

### 4.1 Critique

Dr. Blum's suggestion to conduct both linear and logistic regressions are good places to start but remain inappropriate methods to model the data from the trial.

The main reason is the models do not appropriately address the right censorship present in the data. A patient is right censored if the event is not directly observed during the study and all that is known is the time to event is bounded below by the censorship time. In this dataset, patients are censored due to still being alive by the end of the ten-year study period, or due to receiving a liver transplant. The standard linear and logistic regression model objects in R do not address this issue (though it is still possible to incorporate censoring into estimation); therefore, estimation of the model parameters is based on a data likelihood that is not reflective of the actual information present in the sample. In the linear regression, we made some modifications to the response variable to get around this issue. However, excluding censored observations is not an efficient use of the entire dataset, and treating the censored time as the death times is fundamentally incorrect because deaths were never observed for those patients. An ideal model would encode the true state of information present in the dataset into the likelihood.

The fit statistics, particularly the  $R^2$  in the linear models which intuitively can be interpreted as the percentage of total variation in the data that is explained by the model, is very low at 0.23 even for the multivariate

linear model. The linear model does not do a good job of describing changes in time to event between patients that are present in our data.

There are other technical shortcomings of the linear and logistic approaches. In the linear model, each patient's time to death is assumed to follow a normal distribution, so the range of values is the entire real line. However, time to death should be nonnegative and the linear model does not restrict the response to the proper space. The logistic regression is also unsatisfactory because we are interested in a different response variable than what the logistic regression models. Our dataset provides us with rich information on time to death which we can use to estimate a survival curve, but a logistic regression models binary outcome which can be considered a summary of the time to death, and a survival function would not be possible. Therefore, the basic model formulation of a logistic regression does not fully address the question that we are looking to answer: How does the treatment affect the time to death among PBC patients?

## 4.2 Parametric Survival Analysis

We conduct a parametric Weibull survival analysis from the `flexsurvreg` R package on the data to address the shortcomings from the earlier analysis using the same combination of predictors as before. This model assumes that each patient's time to death follows a nonnegative Weibull distribution. Specifically, the log of the scale parameter is assumed to be a linear function of our covariates, and the shape parameter is assumed to be constant. Right censoring is now also incorporated when conducting maximum likelihood estimation (MLE), contrary to the earlier models. More details on estimation of Weibull survival models can be found in the technical appendix, including steps to derive the mean and standard error of each of the estimated parameters.

The best performing model based on lowest AIC is the multivariate model with all three covariates. We focus on this model for interpretation of the coefficient estimates. The first two parameters in the regression output are the shape and scale. If a Weibull survival function is defined as

$$S(x) = \exp\left(-\frac{x^\alpha}{\lambda}\right), \quad x \geq 0$$

Then the shape corresponds to  $\alpha$  and the scale corresponds to  $\lambda$ . This is exactly the notation used for the `dweibull` function in R, which is different from that used in class. The estimate corresponding to shape from the regression output is the estimate of the constant shape parameter  $\alpha$  of our time to event Weibull distribution. However, the scale is assumed to vary by patient and is reparametrized so that its log is a linear function of the covariates.

$$\log(\lambda_i) = \mu + \beta^\top Z_i$$

With covariate values  $Z_i$  for patient  $i$  and model parameters  $\mu$  and  $\beta$ . Therefore, the scale estimate in the regression output can be thought of as an intercept when all the covariate values  $Z$  are set to 0, corresponding to a scale parameter value of  $\lambda_0 = \exp(\mu)$ . In other words the patients in the treatment group that are age 0 and have serum bilirubin levels below 1.1 mg/dL have an estimated Weibull survival curve with a scale parameter of  $\lambda = \exp(\mu)$ .

The estimates for the rest of the parameters are the linear effect each of the three covariates have on  $\log(\lambda)$ . The estimate for the drug predictor is  $-0.167$ , which means the mean survival time of the placebo group is  $\exp(-0.167) = 0.846$  of the treatment group, when all other covariate values are fixed. However, the confidence interval of the estimate overlaps 0, which is not compelling evidence that the drug has a significant effect on survival time. The estimate for the age in years is  $-0.029$ , meaning an increase in age of one year is associated with a decrease in mean survival time by a multiplicative factor of  $\exp(-0.029) = 0.97$ , a 3% decrease, so not a huge change between ages. Our data does span a large age range from 26 to 78 years, so our model may not accurately capture differences in the effect between younger ages and older ages. The serum 1.1-3.3 estimate tells us that the mean survival time of the serum 1.1-3.3 group is lower than the serum  $<1.1$  group by a multiplicative factor of  $\exp(-0.96) = 0.383$  or a 61.7% decrease. For the serum  $>3.3$

estimate, the difference between the serum  $<1.1$  group and the serum  $>3.3$  group is a multiplicative factor of  $\exp(-1.90) = 0.15$ , or a 85% decrease.

The direction of each of the coefficients in this survival model is reasonable. The estimates tell us that mean survival time decreases for those in the placebo group, older patients, and patients with high serum bilirubin levels. The age and serum effects match with our preconceived notions of PBC mortality, and suggests there is weak evidence that the drug improves the prognosis within each age and serum level group.

One notable difference between the univariate analyses and the multivariate analysis is the direction of the effect of the treatment. The effects of age and serum bilirubin remain relatively unchanged between the univariate and multivariate models. However, In the univariate drug survival model, the placebo group is estimated to have an increased multiplicative effect compared to the treatment group on mean survival time of  $\exp(0.04) = 1.043$ , or a 4.3% increase. While this estimate is close to 1 and therefore does not suggest a strong effect, it is contrary to the multivariate model's estimate of 0.846 / 15.4% decrease in mean survival time, where the placebo group has a shorter mean time to event. This result suggests that accounting for age and serum bilirubin levels helps alleviate potential confounding of the drug and time to event relationship.

## 5 Conclusions

In this analysis, we experimented with several regression approaches to model the time to PBC death as a function of treatment, age, and serum bilirubin levels. The linear and logistic models were unsatisfactory because the assumed distributions are not correct descriptions of time to event data, which is a continuous and nonnegative quantity. In addition, the linear and logistic models as used did not account for right censorship in the maximum likelihood estimation.

The Weibull survival model we run in the last section addresses these critical issues, and the model including all three covariates did the best job of capturing the variation in the data. This is supported by the multivariate model having the lowest AIC out of all survival models tested. We recommend that Dr. Blum use this for further research of the data in this trial.

## 6 Appendix

Parameters are estimated by maximum likelihood using the algorithms available in the standard R `optim` function.

A Weibull regression model assumes

$$\log X = \mu + \gamma^T Z + \sigma W, W \sim S_W(x) = \exp\{-\exp(x)\} \quad (1)$$

$X|Z$  follows Weibull distribution, with probability density function

$$f(x) = \alpha \lambda_Z x^{\alpha-1} \exp(-\lambda_Z x^\alpha) \quad (2)$$

and survival function

$$S_X(x|Z) = \exp(-\lambda_Z x^\alpha) \quad (3)$$

where

$$\lambda_Z = \exp\left[-\frac{\mu + \gamma^T Z}{\sigma}\right], \alpha = \frac{1}{\sigma}$$

For right censored sample with  $n$  observations, the likelihood function should be changed to

$$L(x_1, \dots, x_n; k, \lambda) = \prod_{i:c_i=1} \alpha \lambda_Z x^{\alpha-1} \exp(-\lambda_Z x^\alpha) \prod_{i:c_i=0} \exp(-\lambda_Z X^\alpha) \quad (4)$$

The log-likelihood function is

$$l = \ln L = \sum_{i:c_i=1} \{\log[\alpha \lambda_Z X^{\alpha-1}] - \lambda_Z X^\alpha\} - \sum_{i:c_i=0} \lambda_Z X^\alpha \quad (5)$$

### 6.1 Estimates

Take derivatives of the log-likelihood with respect to  $\mu, \sigma$  and  $\gamma$ . Set equations to zero, and solve them through iterative procedures. The BFGS optimization algorithm is the default in `flexsurvreg`. Method “BFGS” is a quasi-Newton method (also known as a variable metric algorithm), specifically that published simultaneously in 1970 by Broyden, Fletcher, Goldfarb and Shanno. This uses function values and gradients to build up a picture of the surface to be optimized. It uses the analytic derivatives of the likelihood with respect to the model parameters, if these are available, to improve the speed of convergence to the maximum.

In `flexsurvreg`, the reported *shape* is  $\alpha = \frac{1}{\sigma}$ , *scale* is  $\lambda_Z^{-\frac{1}{\alpha}} = \exp(\mu + \gamma^T Z)$  and covariate effects  $\gamma$  have the same “accelerated failure time” interpretation, as linear effects on  $\log X$ .

### 6.2 Standard errors

Standard errors are calculated via the observed Fisher information. Let  $\theta = [\mu, \sigma, \gamma]$ . The Hessian matrix of (5) is

$$\mathbf{H} = \nabla^2 l(x_1, \dots, x_n; \theta) = \nabla^2 l(\theta) \quad (6)$$

where,

$$\mathbf{H}_{i,j} = \frac{\partial^2 l(\theta)}{\partial \theta_i \partial \theta_j}$$



The observed Fisher information is

$$\mathbf{I}(\hat{\theta}) = -\mathbf{H} \quad (7)$$

A standard asymptotic approximation to the distribution of the MLE for large  $N$  is a normal distribution with mean  $\theta$  and variance  $[\mathbf{I}(\hat{\theta})]^{-1}$ .

Therefore, the standard errors of  $\hat{\theta}$  are

$$\hat{se}(\hat{\theta}) = [\mathbf{I}(\hat{\theta})]^{-\frac{1}{2}} \quad (8)$$

The corresponding approximate 95% confidence intervals for  $\theta$  is

$$\hat{\theta} \pm 1.96 [\mathbf{I}(\hat{\theta})]^{-\frac{1}{2}} \quad (9)$$

### 6.2.1 Covariate effects ( $\beta$ )

**6.2.1.1 Estimate**  $X|Z$  follows Weibull distribution with parameter derived from above formulas. With the method of generalized weighted least-square, we can compute the MLE of covariate  $\beta$ .

Let  $Y = \log X$ . Given a symmetric weight matrix  $\mathbf{V}$ , which equals to  $\sigma^{-1}W^{-1}$  in (1), we get

$$Y = Z\beta + \mathbf{V}$$

The solution of  $\underset{\beta}{\operatorname{argmin}}(Y - Z\beta)^T \mathbf{V}(Y - Z\beta)$  is

$$\hat{\beta} = (Z^T \mathbf{V} Z)^{-1} Z^T \mathbf{V} Y \quad (17)$$

**6.2.1.2 Standard errors** The corresponding variance is

$$\operatorname{Cov}(\hat{\beta}) = [(Z^T \mathbf{V} Z)^{-1} Z^T \mathbf{V}] \sigma W [\mathbf{V}^T Z (Z^T \mathbf{V} Z)^{-1}] \quad (18)$$

Therefore the standard error of  $\hat{\beta}$  is  $\operatorname{Cov}(\hat{\beta})^{\frac{1}{2}}$

## 6.3 Right Censored Data

For right censored sample with  $n$  observations, the likelihood function should be changed to

$$L(x_1, \dots, x_n; k, \lambda) = \prod_{i:c_i=1} \frac{k}{\lambda} \left(\frac{x_i}{\lambda}\right)^{k-1} \exp\left[-\left(\frac{x_i}{\lambda}\right)^k\right] \prod_{i:c_i=0} \exp\left(-\left(\frac{x_i}{\lambda}\right)^k\right) \quad (19)$$

The log-likelihood function is

$$\ell = \ln L = \sum_{i:c_i=1} \left\{ \log\left[\frac{k}{\lambda} \left(\frac{x_i}{\lambda}\right)^{k-1}\right] - \left(\frac{x_i}{\lambda}\right)^k \right\} - \sum_{i:c_i=0} \left(\frac{x_i}{\lambda}\right)^k \quad (20)$$

The method to derive the maximum likelihood estimations for parameters  $k, \lambda, \beta$  is the same as that for no censoring data, so as the method for calculating standard errors and confidence intervals for these parameters.

## 7 Supplementary Tables and Figures

### 7.1 Linear Regression Models

This part of the supplementary section goes over the results of the linear regression analysis conducted in section 3.1. The survival data was examined under two scenarios:

- i) complete exclusion of censored data
- ii) treating the time of censorship as time of death.

#### 7.1.1 Excluding censored observations

In this section, the data used for the linear models excludes any censored data. As a result we have 125 observations.

##### 7.1.1.1 Regression on treatment (drug)

Table 1: Linear regression results using treatment as the predictor

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	1608	305.3	5.268	5.975e-07
<b>drug</b>	-89.63	195.5	-0.4585	0.6474

Table 1 organizes the results of the linear regression of time  $\sim$  treatment. This model had a multiple  $R^2 = 0.001707$  and adjusted  $R^2 = -0.00641$ . The F-statistic was  $F_{1,123} = 0.2103$  with a p-value of  $p = 0.6474$ .

In this model,  $\beta_{\text{drug}} = -89.63$ . This means being in the placebo group will decrease the survival time by 89.63 days on average compared to being given D-penicillamine. However, because the p-value for this coefficient ( $p = 0.647$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant.

##### 7.1.1.2 Regression on age

Table 2: Linear regression results using age as the predictor

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	2445	518.3	4.716	6.403e-06
<b>ageinyear</b>	-18.2	9.566	-1.902	0.05945

Table 2 organizes the results of the linear regression of time  $\sim$  age. This model had a multiple  $R^2 = 0.02858$  and adjusted  $R^2 = 0.2069$ . The F-statistic was  $F_{1,123} = 3.619$  with a p-value of  $p = 0.05945$ .

In this model,  $\beta_{\text{age}} = -18.199$ . This means a year increase in age will decrease the survival time by 18.199 days. However, because the p-value for this coefficient ( $p = 0.0594$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant.

### 7.1.1.3 Regression on serum bilirubin

Table 3: Linear regression results using serum bilirubin as the predictor

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	2109	240.2	8.779	1.222e-14
<b>bili_cat1.1-3.3</b>	-259.5	279.5	-0.9284	0.355
<b>bili_cat&gt;3.3</b>	-1111	272.1	-4.083	7.97e-05

Table 3 organizes the results of the linear regression of  $\text{time} \sim \text{serum bilirubin}$ . This model had a multiple  $R^2 = 0.1853$  and adjusted  $R^2 = 0.172$ . The F-statistic was  $F_{2,122} = 13.88$  with a p-value of  $p = 3.712 \times 10^{-6}$ .

In this model,  $\beta_{\text{SB} \in [1.1, 3.3]} = -259.5$ . This means if a person's serum bilirubin level is within 1.1 to 3.3 mg/dL it will decrease the survival time by 259.5 days on average compared to people whose serum bilirubin levels are below 1.1 mg/dL. However, because the p-value for this coefficient ( $p = 0.355$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant. Furthermore,  $\beta_{\text{SB} > 3.3} = -1111.1$ . This means if a person's serum bilirubin level is greater than 3.3 mg/dL it will decrease the survival time by 1111.1 days on average compared to people whose serum bilirubin levels are below 1.1 mg/dL.

### 7.1.1.4 Regression on all three covariates

Table 4: Linear regression results using treatment, age, and serum bilirubin as predictors

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	3675	627.6	5.856	4.237e-08
<b>drug</b>	-144.3	174.4	-0.8271	0.4098
<b>ageinyear</b>	-23.98	8.71	-2.753	0.006818
<b>bili_cat1.1-3.3</b>	-336.1	274.5	-1.224	0.2232
<b>bili_cat&gt;3.3</b>	-1208	268	-4.508	1.533e-05

Table 4 organizes the results of the linear regression of  $\text{time} \sim \text{treatment} + \text{age} + \text{serum bilirubin}$ . This model had a multiple  $R^2 = 0.2354$  and adjusted  $R^2 = 0.2099$ . The F-statistic was  $F_{4,120} = 13.88$  with a p-value of  $p = 1.533 \times 10^{-6}$ .

In this model,  $\beta_{\text{drug}} = -144.27$ . This means being in the placebo group will decrease the survival time by 144.27 days on average compared to being given D-penicillamine. However, because the p-value for this coefficient ( $p = 0.40982$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant. For age,  $\beta_{\text{age}} = -23.98$ . This means a year increase in age will decrease the survival time by 23.98 days. For serum bilirubin,  $\beta_{\text{SB} \in [1.1, 3.3]} = -336.07$  and  $\beta_{\text{SB} > 3.3} = -1208.34$ . This means if a person's serum bilirubin level is between 1.1 to 3.3 mg/dL then their survival time decreases by 336.07 days on average compared to people whose serum bilirubin level is below 1.1 mg/dL. If a person's serum bilirubin level is greater than 3.3 mg/dL then their survival time decreases by 1208.34 days on average compared to people whose serum bilirubin level is below 1.1 mg/dL. However, the p-value for  $\beta_{\text{SB} \in [1.1, 3.3]}$  is greater than the prespecified confidence level  $\alpha = 0.05$ , thus this term is not statistically significant. Interpretations for each coefficient are only relevant given that the other coefficients stay constant.

### 7.1.2 Treating censored times as death times

In this section, the data used for the linear models treats censored data as equal to the death time. As a result we have 312 observations.

#### 7.1.2.1 Regression on treatment (drug)

Table 5: Linear regression results using treatment as the predictor

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	2034	200.7	10.14	4.749e-21
<b>drug</b>	-18.76	127.4	-0.1472	0.883

Table 5 organizes the results of the linear regression of  $\text{time} \sim \text{treatment}$ . This model had a multiple  $R^2 = 6.992 \times 10^{-5}$  and adjusted  $R^2 = -0.003156$ . The F-statistic was  $F_{1,310} = 0.02168$  with a p-value of  $p = 0.883$ .

In this model,  $\beta_{\text{drug}} = -18.76$ . This means being in the placebo group will decrease the survival time by 18.76 days on average compared to being given D-penicillamine. However, because the p-value for this coefficient ( $p = 0.883$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant.

#### 7.1.2.2 Regression on age

Table 6: Linear regression results using age as the predictor

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	2716	305.5	8.892	5.059e-17
<b>ageinyear</b>	-14.19	5.975	-2.375	0.01816

Table 6 organizes the results of the linear regression of  $\text{time} \sim \text{age}$ . This model had a multiple  $R^2 = 0.01787$  and adjusted  $R^2 = 0.0147$ . The F-statistic was  $F_{1,310} = 5.641$  with a p-value of  $p = 0.01816$ .

In this model,  $\beta_{\text{age}} = -14.191$ . This means a year increase in age will decrease the survival time by 14.191 days.

#### 7.1.2.3 Regression on serum bilirubin

Table 7: Linear regression results using serum bilirubin as the predictor

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	2504	92.25	27.14	8.397e-84
<b>bili_cat1.1-3.3</b>	-397.7	131.3	-3.028	0.002665
<b>bili_cat&gt;3.3</b>	-1330	142.8	-9.31	2.463e-18

Table 7 organizes the results of the linear regression of  $\text{time} \sim \text{serum bilirubin}$ . This model had a multiple  $R^2 = 0.2226$  and adjusted  $R^2 = 0.2176$ . The F-statistic was  $F_{2,309} = 44.24$  with a p-value of  $p < \epsilon_{\text{Machine}} = 2.22 \times 10^{-16}$ .

In this model,  $\beta_{\text{SB} \in [1.1, 3.3]} = -397.73$ . This means if a person's serum bilirubin level is within 1.1 to 3.3

mg/dL it will decrease the survival time by 397.73 days on average compared to people whose serum bilirubin levels are below 1.1 mg/dL. Furthermore,  $\beta_{\text{SB}>3.3} = -1329.90$ . This means if a person's serum bilirubin level is greater than 3.3 mg/dL it will decrease the survival time by 1329.90 days on average compared to people whose serum bilirubin levels are below 1.1 mg/dL.

#### 7.1.2.4 Regression on all three covariates

Table 8: Linear regression results using treatment, age, and serum bilirubin as predictors

	Estimate	Std. Error	t value	Pr(> t )
<b>(Intercept)</b>	3233	348.6	9.274	3.288e-18
<b>drug</b>	-40.12	113.2	-0.3543	0.7234
<b>ageinyear</b>	-13.39	5.339	-2.509	0.01262
<b>bili_cat1.1-3.3</b>	-400.3	130.8	-3.06	0.00241
<b>bili_cat&gt;3.3</b>	-1322	141.9	-9.319	2.365e-18

Table 8 organizes the results of the linear regression of  $\text{time} \sim \text{treatment} + \text{age} + \text{serum bilirubin}$ . This model had a multiple  $R^2 = 0.2382$  and adjusted  $R^2 = 0.2283$ . The F-statistic was  $F_{4,307} = 24$  with a p-value of  $p < \epsilon_{\text{Machine}} = 2.22 \times 10^{-16}$ .

In this model,  $\beta_{\text{drug}} = -40.120$ . This means being in the placebo group will decrease the survival time by 40.120 days on average compared to being given D-penicillamine. However, because the p-value for this coefficient ( $p = 0.72337$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant. For age,  $\beta_{\text{age}} = -13.394$ . This means a year increase in age will decrease the survival time by 13.394 days. For serum bilirubin,  $\beta_{\text{SB} \in [1.1, 3.3]} = -400.348$  and  $\beta_{\text{SB} > 3.3} = -1322.395$ . This means if a person's serum bilirubin level is between 1.1 to 3.3 mg/dL then their survival time decreases by 400.348 days on average compared to people whose serum bilirubin level is below 1.1 mg/dL. If a person's serum bilirubin level is greater than 3.3 mg/dL then their survival time decreases by 1322.395 days on average compared to people whose serum bilirubin level is below 1.1 mg/dL. Interpretations for each coefficient are only relevant given that the other coefficients stay constant.

## 7.2 Logistic Regression Models

### 7.2.1 Regression on treatment (drug)

Table 9: Logistic regression results using treatment as the predictor

	Estimate	Std. Error	z value	Pr(> z )
<b>(Intercept)</b>	-0.2675	0.3631	-0.7366	0.4614
<b>drug</b>	-0.09074	0.2312	-0.3925	0.6947

Table 9 organizes the results of the logistic regression of  $\text{dead} \sim \text{treatment}$ . This model had an AIC of 423.97.

In this model,  $\beta_{\text{drug}} = -0.09074$ . This means being in the placebo group will decrease the log odds of death by 0.09074 on average compared to being given D-penicillamine. However, because the p-value for this coefficient ( $p = 0.695$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant.

### 7.2.2 Regression on age

Table 10: Logistic regression results using age as the predictor

	Estimate	Std. Error	z value	Pr(> z )
<b>(Intercept)</b>	-2.951	0.6114	-4.828	1.383e-06
<b>ageinyear</b>	0.05045	0.01177	4.287	1.808e-05

Table 10 organizes the results of the logistic regression of  $\text{dead} \sim \text{age}$ . This model had an AIC of 404.3.

In this model,  $\beta_{\text{age}} = 0.01177$ . This means a year increase in age will increase the log odds of death by 0.01177 on average.

### 7.2.3 Regression on serum bilirubin

Table 11: Logistic regression results using serum bilirubin as the predictor

	Estimate	Std. Error	z value	Pr(> z )
<b>(Intercept)</b>	-1.762	0.2625	-6.711	1.931e-11
<b>bili_cat1.1-3.3</b>	1.459	0.3243	4.499	6.838e-06
<b>bili_cat&gt;3.3</b>	2.721	0.3593	7.573	3.642e-14

Table 11 organizes the results of the logistic regression of  $\text{dead} \sim \text{serum bilirubin}$ . This model had an AIC of 354.73.

In this model,  $\beta_{\text{drug}} = -0.09074$ . This means being in the placebo group will decrease the log odds of death by 0.09074 on average compared to being given D-penicillamine. However, because the p-value for this coefficient ( $p = 0.695$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant.

In this model,  $\beta_{\text{SB} \in [1.1, 3.3]} = 1.4587$ . This means if a person's serum bilirubin level is within 1.1 to 3.3 mg/dL it will increase the log odds of death by 1.4587 on average compared to people whose serum bilirubin levels are below 1.1 mg/dL. Furthermore,  $\beta_{\text{SB} > 3.3} = 2.7208$ . This means if a person's serum bilirubin level is greater than 3.3 mg/dL it will increase the log odds of death by 2.7208 on average compared to people whose serum bilirubin levels are below 1.1 mg/dL.

### 7.2.4 Regression on all three covariates

Table 12: Logistic regression results using treatment, age, and serum bilirubin as predictors

	Estimate	Std. Error	z value	Pr(> z )
<b>(Intercept)</b>	-5.15	0.9439	-5.456	4.865e-08
<b>drug</b>	0.07332	0.2761	0.2656	0.7906
<b>ageinyear</b>	0.06316	0.01389	4.547	5.44e-06
<b>bili_cat1.1-3.3</b>	1.562	0.3389	4.607	4.077e-06
<b>bili_cat&gt;3.3</b>	2.914	0.3808	7.653	1.966e-14

Table 12 organizes the results of the logistic regression of  $\text{dead} \sim \text{treatment} + \text{age} + \text{serum bilirubin}$ . This model had an AIC of 335.33.

In this model,  $\beta_{\text{drug}} = 0.07332$ . This means being in the placebo group will increase the log odds of death by 0.07332 on average compared to being given D-penicillamine. However, because the p-value for this coefficient ( $p = 0.791$ ) is greater than the prespecified confidence level  $\alpha = 0.05$ , this term is not statistically significant. For age,  $\beta_{\text{age}} = 0.06316$ . This means a year increase in age will increase log odds of death by 0.06316. For serum bilirubin,  $\beta_{\text{SB} \in [1.1, 3.3]} = 1.56164$  and  $\beta_{\text{SB} > 3.3} = 2.91444$ . This means if a person's serum bilirubin level is between 1.1 to 3.3 mg/dL then their log odds of death increase by 1.56164 days on average compared to people whose serum bilirubin level is below 1.1 mg/dL. If a person's serum bilirubin level is greater than 3.3 mg/dL then their log odds of death increases by 2.91444 days on average compared to people whose serum bilirubin level is below 1.1 mg/dL. Interpretations for each coefficient are only relevant given that the other coefficients stay constant.

### 7.3 Parametric Survival Analysis

#### 7.3.1 Parametric survival analysis on drug

Table 13: Weibull survival analysis results using treatment as the predictor

	Estimate	Std. Error	L95%	U95%
<b>shape</b>	1.127	0.08867	0.9659	1.315
<b>scale</b>	4335	1095	2642	7111
<b>drug</b>	0.04176	0.1589	-0.2696	0.3531

Table 13 organizes the results of the Weibull regression of  $\text{time} \sim \text{treatment}$ . The log-likelihood of -1178.718 with 3 degrees of freedom. This model had an AIC of 2383.436.

In this model,  $\beta_{\text{drug}} = 0.0418$ . This means being in the placebo group decreases the hazard of death by 0.9591 times compared to being given D-penicillamine. However because the 95% confidence interval contains 1 ( $\lambda_{\text{drug}} \in [0.7025, 1.309]$ ), we cannot say that this coefficient is statistically significant at the predetermined confidence level of  $\alpha = 0.05$ .

#### 7.3.2 Parametric survival analysis on age

Table 14: Weibull survival analysis results using age as the predictor

	Estimate	Std. Error	L95%	U95%
<b>shape</b>	1.144	0.08947	0.9815	1.334
<b>scale</b>	26540	11824	11083	63553
<b>ageinyear</b>	-0.03436	0.007959	-0.04996	-0.01877

Table 14 organizes the results of the Weibull regression of  $\text{time} \sim \text{age}$ . The log-likelihood of -1178.773 with 3 degrees of freedom. This model had an AIC of 2363.467.

In this model,  $\beta_{\text{age}} = -0.0344$ . This means a year increase in age increases the hazard of death by 1.035 times compared to someone just born ( $\text{age}_i = 0$ ).

### 7.3.3 Parametric survival analysis on serum bilirubin

Table 15: Weibull survival analysis results using serum bilirubin as the predictor

	Estimate	Std. Error	L95%	U95%
<b>shape</b>	1.337	0.1007	1.153	1.549
<b>scale</b>	10785	2242	7176	16209
<b>bili_cat1.1-3.3</b>	-0.9427	0.2215	-1.377	-0.5085
<b>bili_cat&gt;3.3</b>	-1.902	0.23	-2.353	-1.452

Table 15 organizes the results of the Weibull regression of time  $\sim$  serum bilirubin. The log-likelihood of -1136.177 with 4 degrees of freedom. This model had an AIC of 2280.354.

In this model,  $\beta_{\text{SB} \in [1.1, 3.3]} = -0.943$ . This means a person's serum bilirubin level being between 1.1 and 3.3 mg/dL increases the hazard of death by 2.5677 times compared to a person whose serum bilirubin level is less than 1.1 mg/dL. Furthermore,  $\beta_{\text{SB} > 3.3} = -1.9$ . This means a person's serum bilirubin level being greater than 3.3 mg/dL increases the hazard of death by 6.6859 times compared to a person whose serum bilirubin level is less than 1.1 mg/dL.

### 7.3.4 Parametric survival analysis on all three covariates

Table 16: Weibull survival analysis results using treatment, age, and serum bilirubin as predictors

	Estimate	Std. Error	L95%	U95%
<b>shape</b>	1.365	0.1016	1.18	1.58
<b>scale</b>	61014	30192	23133	160928
<b>drug</b>	-0.1669	0.1349	-0.4313	0.09752
<b>ageinyear</b>	-0.02925	0.006207	-0.04141	-0.01708
<b>bili_cat1.1-3.3</b>	-0.9603	0.2174	-1.386	-0.5342
<b>bili_cat&gt;3.3</b>	-1.9	0.2251	-2.341	-1.459

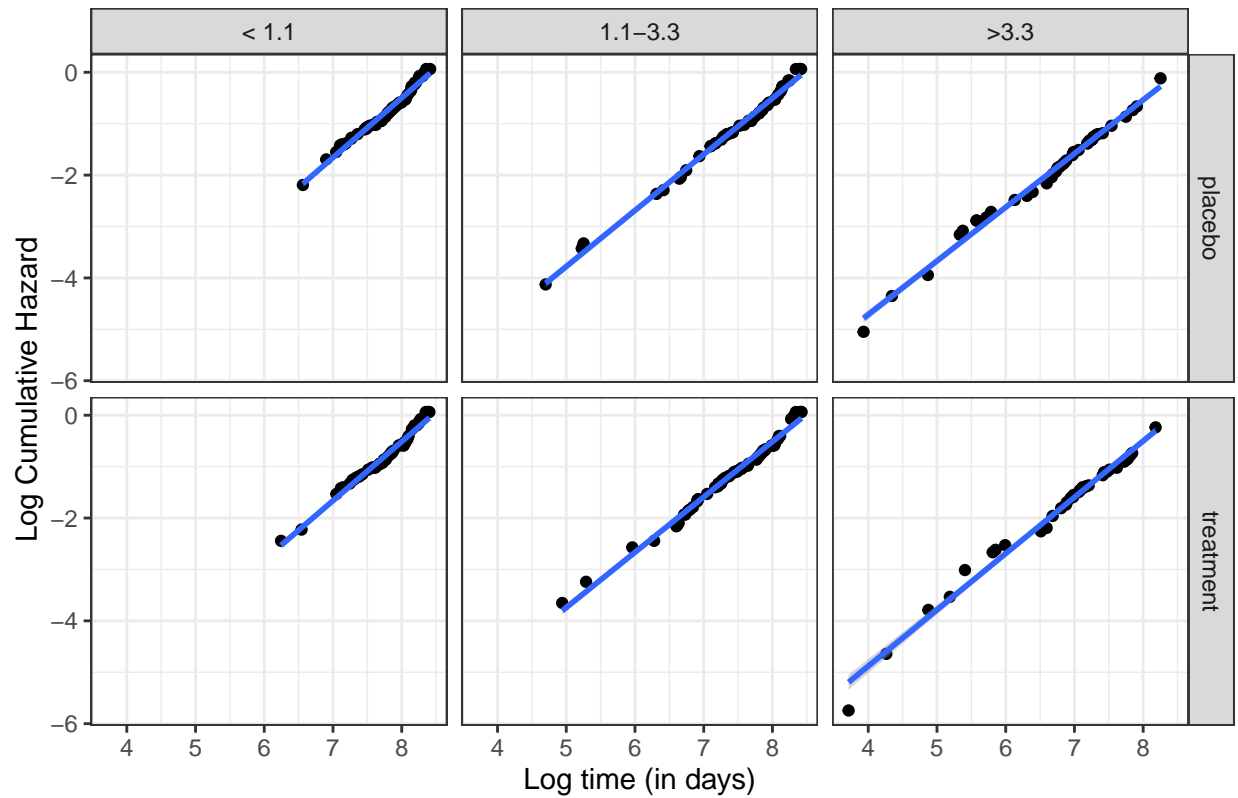
Table 16 organizes the results of the Weibull regression of time  $\sim$  treatment + age + serum bilirubin. The log-likelihood of -1124.358 with 6 degrees of freedom. This model had an AIC of 2260.716.

In this model,  $\beta_{\text{drug}} = -0.167$ . This means being in the placebo group increases the hazard of death by 1.1818 times compared to being given D-penicillamine. However because the 95% confidence interval contains 1 ( $\lambda_{\text{drug}} \in [0.907, 1.539]$ ), we cannot say that this coefficient is statistically significant at the predetermined confidence level of  $\alpha = 0.05$ . For age,  $\beta_{\text{age}} = -0.0292$ . This means a year increase in age increases the hazard of death by 1.0296 times compared to someone just born ( $\text{age}_i = 0$ ). For serum bilirubin,  $\beta_{\text{SB} \in [1.1, 3.3]} = -0.96$  and  $\beta_{\text{SB} > 3.3} = -1.9$ . This means a person's serum bilirubin level being between 1.1 and 3.3 mg/dL increases the hazard of death by 2.6117 times compared to a person whose serum bilirubin level is less than 1.1 mg/dL. A person's serum bilirubin level being greater than 3.3 mg/dL increases the hazard of death by 6.6859 times compared to a person whose serum bilirubin level is less than 1.1 mg/dL. Interpretations for each coefficient are only relevant given that the other coefficients stay constant.



### 7.3.5 Checking Weibull distribution is a good fit with log-log plot

Check fit data to Weibull Distribution



The plot of log time against the log cumulative hazard is mostly linear in each of the treatment / serum bilirubin groups and deviates mostly at the earlier time points. Due to the linearity within each of the groups, we conclude the Weibull distribution is a good fit for this dataset. It is harder to visualize along with age (our third covariate), but based on this diagnostic it seems Weibull is a good fit regardless.