1.

**Methods:** All analyses were performed in R. Descriptive statistics are reported stratified by levels of baseline serum bilirubin levels. For the continuous variable of baseline age, the mean, standard deviation, and minimum and maximum values are reported. For categorical variables (number of deaths, sex, and treatment status) count and percentages are reported. Missing observations are noted with a superscript and the corresponding number with data within each subset. Kaplan-Meier estimates of survival and restricted mean survival were made stratified by bilirubin levels using the “survival” package. Arbitrary distinctions in bilirubin were made by dividing the continuous variable into subsets of 5 mg/dl (0-5, 5-10, 10-15, >15) for the purposes of plotting survival stratified by baseline bilirubin levels.

**Results:** There was no consistent trend in either mean age at baseline or in the proportion female in gender across bilirubin levels. However, there appeared to be a trend toward decreasing receipt of the D-Penicillamine treatment with increasing bilirubin levels (52% > 51% > 45% > 31%). As well, there was a general trend toward decreasing survival with increasing bilirubin levels (restricted mean survival: 2526 days > 1502 days > 848 days > 846 days and 30% < 62% < 88% < 94% mortality), though this trend was more inconsistent when considering median survival time (3762 days > 1427 days > 679 days < 853 days).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Bilirubin 0<5 mg/dl****(N=340)** | **Bilirubin ≥5<10 mg/dl****(N=45)** | **Bilirubin ≥10<15 mg/dl****(N=16)** | **Bilirubin ≥15 mg/dl****(N=17)** |
| N Deaths, # (%) | 103 (30%) | 28 (62%) | 14 (88%) | 16 (94%) |
| Median Survival, days | 3762 | 1427 | 679 | 853 |
| R-Mean Survival, days | 2526 | 1502 | 848 | 846 |
| Age, years\* | 50.72 (10.55; 26.28-76.71) | 50.83 (10.92; 30.86-78.44) | 50.57 (10.68; 33.18-70.56) | 51.16 (7.26; 41.95-66.0) |
| Sex, # (%) | 224 (89%)a | 28 (80%)c | 11 (100%)d | 13 (100%)e |
| Treatment, # (%) | 130 (52%)b |  18 (51%)c | 5 (45%)d | 4 (31%)e |

R-Mean = restricted mean; a N = 253; b N=251; c N=35; d N=11; e N=13



2. The largest difference between prior datasets from the Cardiovascular Health Study (CHS) and the current PBC dataset is that follow-up was complete for roughly 4 years in prior data, while in the current dataset, events (death) and censoring occur with regularity from the onset of the study (minimum time to censoring/death = 41 days, out of ~5000 total days of observation). As a result, special censoring methods are required to provide descriptive statistics comparing the two groups’ survival over time. As well, in this particular analysis, there are numerous baseline characteristics that represent liver disease at different stages (e.g., ascites, which is highly associated with liver failure and portal hypertension) that would be expected to predict the outcome (i.e., they affect the time to outcome). Thus, in this case and in analyzing this data, survival methods are preferred to logistic regression.

3a.

**Methods:** Proportional hazards regression modeling baseline serum bilirubin as an untransformed, continuous random variable was used to assess the relationship between time-to-death distributions from any cause across groups defined by baseline bilirubin levels. The hazard ratio (HR) from proportional hazards regression output was used to quantify the association between all-cause mortality and serum bilirubin. Wald statistics and the Huber-White sandwich estimator were used to calculate the 95% confidence intervals and two-sided p values, respectively. Participants missing bilirubin or follow-up data were excluded from analyses.

**Results:** Data was available on 418 total participants, who had a mean serum bilirubin of 3.22 mg/dl and a standard deviation of 4.41 mg/dl (range: 0.3 – 28 mg/dl). Over a mean observation time of 1917.8 days (5.25 years), 161 deaths were observed. From proportional hazards regression estimates, we estimate that there is a 15.26% increase in the instantaneous risk of death (HR = 1.1526) for each 1 mg/dl increase in serum bilirubin at baseline. This observed hazard ratio would not be unusual if the true population instantaneous risk of death was between 12% and 18% higher for a group with a baseline bilirubin 1 mg/dl higher than another group (95% CI for the HR: 1.12-1.18). Based on the two-sided p-value of p < 0.00005, we can reject the null hypothesis that there is no association between baseline serum bilirubin and risk of death from any cause in favor of a hypothesis that higher baseline bilirubin is associated with increased mortality.

3b. Please see graph in question 6.

4a. The distribution of serum bilirubin has a strong right skew/tail, thus, using the natural log of serum bilirubin would be preferred in that it would downweight the outliers with very high bilirubin levels (this would also probably lead to increased precision through decreased variance at the outliers). As well, since the hazard ratio is modeled multiplicatively, it makes more intuitive sense to also model the predictor on such a scale to aid in interpretation (e.g., “a two-fold increase in bilirubin resulted in an x-fold instantaneous increase in risk of death”).

b.

**Methods:** Proportional hazards regression modeling baseline serum bilirubin as a natural log transformed, continuous random variable was used to assess the relationship between time-to-death distributions from any cause across groups defined by baseline bilirubin levels. The hazard ratio (HR) from proportional hazards regression output was used to quantify the association between all-cause mortality and serum bilirubin. Wald statistics and the Huber-White sandwich estimator were used to calculate the 95% confidence intervals and two-sided p values, respectively. Participants missing ln(bilirubin) or follow-up data were excluded from analyses. As no observations had bilirubin = 0, no special transformations were performed prior to natural log transformation of bilirubin.

**Results:** Data was available on 418 total participants, who had a mean ln(serum bilirubin) of 0.571 and a standard deviation of 1.024 (range: -1.204 – 3.332). Over a mean observation time of 1917.8 days (5.25 years), 161 deaths were observed. From proportional hazards regression estimates, we estimate that there is a 98.48% increase in the instantaneous risk of death (HR = 1.9848) for each 2-fold increase in serum bilirubin at baseline. This observed hazard ratio would not be unusual if the true population instantaneous risk of death was between 87.80% and 109.76% higher for a group with a baseline bilirubin 2-fold greater than another group (95% CI for the HR: 1.8780-2.0976). Based on the two-sided p-value of p < 0.00005, we can reject the null hypothesis that there is no association between baseline ln(serum bilirubin) and risk of death from any cause in favor of a hypothesis that higher baseline natural log transformed bilirubin is associated with increased mortality.

4c. Please see graph in question 6.

5a.

**Methods:** Proportional hazards regression modeling both baseline serum bilirubin as a continuous variable and baseline serum bilirubin as a natural log transformed continuous random variable was used to assess the relationship between time-to-death distributions from any cause across groups defined by baseline bilirubin levels. The hazard ratio (HR) from proportional hazards regression output was used to quantify the association between all-cause mortality and the two models forms of serum bilirubin (continuous and natural log transformed). Wald statistics and the Huber-White sandwich estimator were used to calculate the 95% confidence intervals and two-sided p values, respectively. Participants missing ln(bilirubin) or follow-up data were excluded from analyses. As no observations had bilirubin = 0, no special transformations were performed prior to natural log transformation of bilirubin.

**Results:** Data was available on 418 total participants, who had a mean serum bilirubin of 3.22 mg/dl and a standard deviation of 4.41 mg/dl (range: 0.3 – 28 mg/dl) and a mean ln(serum bilirubin) of 0.571 and a standard deviation of 1.024 (range: -1.204 – 3.332). Over a mean observation time of 1917.8 days (5.25 years), 161 deaths were observed. Baseline serum bilirubin modeled as a continuous random variable was not significantly associated with instantaneous risk of death (HR = 0.9614, p=0.191) in a multivariate proportional hazards regression model. However, from the same multivariate proportional hazards regression, we estimate that there is a 127.10% increase in the instantaneous risk of death (HR = 2.2710) for each 2-fold increase in serum bilirubin at baseline, as modeled by natural log transformed bilirubin. This observed hazard ratio would not be unusual if the true population instantaneous risk of death was between 102.23% and 155.04% higher for a group with a baseline bilirubin 2-fold greater than another group (95% CI for the HR: 2.0223-2.5504). Based on the two-sided p-value of p < 0.00005, we can reject the null hypothesis that there is no association between baseline ln(serum bilirubin) and risk of death from any cause in favor of a hypothesis that higher baseline natural log transformed bilirubin is associated with increased mortality. We note that the association of ln(serum bilirubin) was markedly increased in effect size when untransformed bilirubin was added to the proportional hazards regression model (HR=2.2710 vs. HR = 1.9848 for a 2-fold increase in baseline serum bilirubin).

5b. Please see graph in question 6.

6. At low ranges of bilirubin (0-3 mg/dl; the mean bilirubin value being 3.22 mg/dl), all 3 estimates were similar. However, at the higher ranges of baseline serum bilirubin, the three estimates diverge drastically. The HRs estimated by continuous bilirubin are exponential, with great increases at the high ranges (>15 mg/dl bilirubin). In contrast, the HRs estimated by the natural log transformed bilirubin from #4 and #5 are linear across all ranges of bilirubin; however, the slope of the estimates from #5 are greater when compared to the HR estimates from #4 due to the greater effect size (HR).



7a/b. Both age and sex are significantly associated with all-cause mortality in this data in univariate proportional hazards regression. Female gender was significantly associated with decreased risk of death (HR = 0.6164 with a 95% CI of 0.385-0.987; p = 0.0407; 125 events among 312 participants with data). Age was also significantly associated with risk of death, with an estimated 4% increase in risk of death for a group 1 year older than a comparative group (HR = 1.04 with a 95% CI of 1.02-1.06; p < 0.00005; 161 events among 418 participants with data).

However, there is no reason to believe that either age or sex are associated with serum bilirubin either through *a priori* evidence or from the observed data (see Table 1). Female gender is associated with a slight increase in baseline serum bilirubin (0.4339) as compared to males, but this increase is not significant at the 0.05 level (p=0.589); thus we accept the null hypothesis that there is no association between female gender and serum bilirubin. Similarly, an increase of 1 year is associated with a slight increase in serum bilirubin (0.0009965), but this increase is not significant at the 0.05 level (p=0.961); thus we accept the null hypothesis that there is no association between age and serum bilirubin.

Given that age and sex are significantly associated with death but not with serum bilirubin, it is my interpretation that both are precision variables (but not confounders) in this analysis.

7c.

**Methods:** Proportional hazards regression modeling baseline serum bilirubin as a natural log transformed, continuous random variable was used to assess the relationship between time-to-death distributions from any cause across groups defined by baseline bilirubin levels, with adjustment for the precision variables sex and age at baseline. The hazard ratio (HR) from proportional hazards regression output was used to quantify the association between age/sex-adjusted all-cause mortality and serum bilirubin. Wald statistics and the Huber-White sandwich estimator were used to calculate the 95% confidence intervals and two-sided p values, respectively. Participants missing ln(bilirubin), age, sex, or follow-up data were excluded from analyses. As no observations had bilirubin = 0, no special transformations were performed prior to natural log transformation of bilirubin.

**Results:** Data was available on 312 total participants, who had a mean ln(serum bilirubin) of 0.575 and a standard deviation of 1.0321 (range: -1.204 – 3.332). Over a mean observation time of 2006.36 days (5.49 years), 125 deaths were observed. From proportional hazards regression estimates, age (p < 0.00001) was significantly associated with a 3.8% increase in risk of death with a 95% CI of 1.0215 to 1.056. Female gender was not significantly associated with instantaneous risk of death (p=0.781); female gender was associated with a 6.55% decrease (95% CI of 0.5799 – 1.506). For natural log transformed bilirubin, we estimate that there is a 110.88% increase in the instantaneous risk of death (HR = 2.1088) for each 2-fold increase in serum bilirubin at baseline. This observed hazard ratio would not be unusual if the true population instantaneous risk of death was between 97.85% and 124.78% higher for a group with a baseline bilirubin 2-fold greater than another group (95% CI for the HR: 1.9784-2.2478). Based on the two-sided p-value of p < 0.00005, we can reject the null hypothesis that there is no association between baseline ln(serum bilirubin) and age/sex-adjusted risk of death from any cause in favor of a hypothesis that higher baseline natural log transformed bilirubin is associated with increased age/sex-adjusted mortality.

8. The intervention of D-Penicallamine could potentially be a confounder, effect modifier, or precision variable. However, as treatment is not associated with long-term mortality from proportional hazards regression (p=0.751), it is unlikely that D-Penicallamine acts as either a confounder or precision variable in these analyses. Cursory proportional hazards regression with an interaction term for serum bilirubin-by-treatment status on the outcome of long-term mortality demonstrates a strong interaction effect (p=0.000572). Thus, it is likely that treatment with D-Penicillamine acts as an effect modifier in this analysis of serum bilirubin and long-term mortality.