
Oregon State University

# Survival Analysis of Treatment Efficacy in Primary Biliary Cirrhosis

**Authors**

Cyril Nana Boakye Benson
Daniel Fowotade Oluwaseun

2024-12-19

## 1 Introduction

Primary biliary cirrhosis (PBC) is a severe and chronic liver disease characterized by progressive bile duct destruction, which can lead to liver failure and life-threatening complications, including the need for liver transplantation or death. The PBC3 dataset originates from a multi-centre randomized clinical trial conducted between January 1983 and January 1987 across six European hospitals. This trial aimed to evaluate the efficacy of Cyclosporin A (`CyA`) in improving outcomes for patients diagnosed with PBC. A total of 349 patients were randomized to either `CyA` (176 patients) or placebo (173 patients). Initially, the primary outcome was defined as survival time; however, due to the increased use of liver transplantation during the study, it was redefined as âfailure of medical treatment,â encompassing death or liver transplantation. Patients were followed until January 1, 1989, or until one of three outcomes occurred: treatment failure, dropout, or the end of the study. By the end of the trial, 61 patients had died (`CyA`: 30, placebo: 31), 29 had undergone liver transplantation (`CyA`: 14, placebo: 15), and 4 were lost to follow-up. This analysis aims to evaluate the effect of `CyA` treatment on time to failure of medical treatment, while accounting for critical patient characteristics, such as `age`, `sex`, and clinical measures. The dataset includes both continuous variables, such as `crea`, `alb`, and `bili`, and categorical variables, such as `tment`, `sex`, and `stage`. Using survival analysis techniques, we will conduct descriptive analyses, univariate analyses using the Kaplan-Meier estimator, and regression modeling with Accelerated Failure Time (`AFT`) models. A critical aspect of this study is assessing potential limitations, including the multi-centre study design, which may introduce correlated observations within the same hospital. The assumptions of regression models will be carefully evaluated to ensure valid and reliable conclusions. Ultimately, this analysis aims to provide a comprehensive understanding of the effect of treatment and patient characteristics on time to failure, informing clinical decision-making and optimizing treatment strategies for PBC.

## 2 Summary Statistics

The dataset summary in (Table 1) shows a diverse patient population. `Ages` ranged from 26 to 75 years, averaging 54.1 years. Key clinical markers varied widely: `creatinine levels` ranged from 35 to 199 (mean 78), `albumin` from 20.63 to 58 (mean 38.4), and bilirubin from 2.33 to 453.1 (mean 45.47). `Alkaline phosphatase` spanned 66.33 to 5108, with a mean of 996.5, and `aspartate transaminase` ranged from 10.5 to 316.3 (mean 95.4). `Patients` weighed between 38 and 98 kg (mean 60.3 kg), and `time to failure` ranged from 1 to 2146 days, with an average of 942.7 days. These variations reflect the differing severity of the condition and treatment outcomes in the study. The frequency analysis provides insights into the distribution of key categorical variables in the dataset. For the treatment variable (`tment`), 50.43% of patients were assigned to Cyclosporin A (CyA), while 49.57% received the placebo. The majority of patients were female (`sex`=0, 85.39%), with males making up 14.61% of the population. Regarding the disease stage (`stage`), 25.79% of patients were at stage 4, followed by 24.07% at stage 2, and 19.48% at stage 3, with 14.04% in stage 1. Notably, 16.62% of entries for this variable are missing. A history of gastrointestinal bleeding (`gibleed`) was reported for 14.33% of patients, with the remaining 85.67% having no such history. In terms of outcomes (`status`), 74.21% of patients were censored (no failure of medical treatment), 8.31% underwent liver transplantation, and 17.48% died during the study.

## 3 Data Visualization

The visualizations reveal several important observations. The histogram (figure 1) of time to failure indicates a wide range of survival times, with a mean of approximately 942 days, though the distribution is not perfectly normal. The box plot (figure 4) comparing treatment types suggests that the median survival times are similar between the two groups, with no major differences in spread or outliers. Scatter plots (figure 2 and 3) of time to failure against age and weight show no clear linear relationships, indicating that these variables may not have strong individual effects on survival time. The scatter plots illustrate the relationship between time to failure (`days`) and key biochemical markers: creatinine (`crea`), albumin (`alb`), and bilirubin (`bili`). The plot of `days` against `crea`(figure 5) shows a scattered distribution without a clear trend, suggesting limited direct association between creatinine levels and survival time. In contrast, `alb` appears to have a positive association with `days` in (figure 6), as higher albumin levels generally correspond to longer survival times. For `bili` (figure 7), the plot indicates a potential negative relationship, with higher bilirubin levels often associated with shorter survival times. These observations highlight the differing impacts of biochemical markers on patient outcomes, warranting further investigation through multivariate analysis.

## 4 Summary of Kaplan-Meier Analysis

The Kaplan-Meier survival analysis revealed several key findings. For treatment groups (Figure 8), the survival curves for Cyclosporin A (`CyA`) and placebo showed overlapping confidence intervals, with a log-rank test p-value of 0.7813, indicating no statistically significant difference in survival probabilities between the two arms. Stratification by disease stage (Figure 9) demonstrated clear separation of survival curves, with advanced stages (Stage 4) associated with markedly lower survival probabilities compared to earlier stages. This was supported by a highly significant log-rank p-value ($<$**0.0001**), underscoring the critical impact of disease stage on survival outcomes.

For sex (Figure 10), a significant difference was observed, with the log-rank test yielding a p-value of 0.0060, suggesting an association between sex and survival time. Patients without a history of gastrointestinal bleeding (`gibleed = 0`) exhibited higher survival probabilities compared to those with bleeding history (Figure 11), with a log-rank test p-value of $<$**0.05**. Finally, survival probabilities varied significantly by patient exit status (Figure 12), where censoring, transplantation, and death showed distinct patterns, highlighting the need to accurately interpret censored observations in survival models. These results indicate that while treatment with `CyA` does not significantly improve survival outcomes, disease stage and gastrointestinal bleeding history are important prognostic factors. Additionally, the association between sex and survival time warrants further exploration to understand potential biological or demographic contributions to this disparity. Early identification and management of critical factors, such as advanced disease stages, remain essential in optimizing patient outcomes.

## 5 Implementing the Cox Model

The Cox Proportional Hazards Model in (Table 2) was applied to evaluate the effect of treatment and other covariates on survival outcomes. The global test for the model was highly significant ($p < 0.0001$), indicating that the covariates collectively explained variations in survival. Among individual predictors, `sex`, `bili`, `alb`, and `stage` were significant contributors. Males exhibited a significantly higher hazard compared to females ($HR = 2.59, p = 0.0015$), while elevated `bili` ($HR = 1.009, p < 0.0001$) and lower `alb` levels ($HR = 0.936, p = 0.0071$) were associated with poorer survival. Advanced disease stages, particularly `stage 4`, were linked to higher risk of failure compared to `stage 2` ($HR = 2.11, p = 0.0409$). Although `tment` (Cyclosporin A treatment) suggested a potential reduction in risk ($HR = 0.65$), this effect was not statistically significant ($p = 0.0680$). `age` and `crea` levels showed no significant association with survival, while `gibleed` and variability across `unit` (hospitals) also had no meaningful effect. These findings highlight key predictors of survival, with `sex`, `bili`, `alb`, and `stage` emerging as significant factors, while the effect of `tment` warrants further investigation.

## 6 Proportional Hazards Assumption Check Analysis

The proportional hazards assumption(Table 3) was evaluated for all variables in the Cox model using Schoenfeld residuals. Statistical tests and graphical diagnostics provided insights into the behavior of each covariate over time. For treatment (`tment`), the test statistic ($p = 0.3520$) and the residual plot showed no discernible trend, indicating that the proportional hazards assumption is satisfied. Similarly, for sex (`sex`), the test yielded $p = 0.7790$, with residuals randomly distributed around zero, further supporting the assumption. `age`, `bili`, `alb`, and `crea` also met the assumption, with $p > 0.05$ in all cases and no systematic patterns observed in the residual plots. For the disease stage (`stage`), a significant violation was detected for `stage 1` ($p < 0.0001$), as evidenced by a clear departure from randomness in the residual plot. Other stages showed no significant issues ($p > 0.05$). Residuals for the hospital (`unit`) and gastrointestinal bleeding (`gibleed`) variables appeared random and within acceptable bounds, with $p > 0.05$ in all cases, confirming that the proportional hazards assumption holds for these variables. Overall, the assumption of proportional hazards was satisfied for most covariates, except for `stage 1`, where the violation suggests potential time-dependent effects. This could be addressed by stratifying the Cox model by `stage` or incorporating time-varying covariates to account for the non-proportionality. The residual plots and test results collectively provide confidence in the model's robustness, with minor adjustments required to handle specific violations.

### 6.1 Interpretation of the Weibull Accelerated Failure Time Model (AFT) Model

The Accelerated Failure Time model for all covariates is given by:

$$\log(\texttt{days}) = \beta_0 + \beta_1 \cdot \texttt{age} + \beta_2 \cdot \texttt{sex} + \beta_3 \cdot \texttt{tment} + \beta_4 \cdot \texttt{stage} + \beta_5 \cdot \texttt{crea} +$$
$$\beta_6 \cdot \texttt{alb} + \beta_7 \cdot \texttt{bili} + \beta_8 \cdot \texttt{alkph} + \beta_9 \cdot \texttt{asptr} + \beta_{10} \cdot \texttt{weight} + \epsilon$$

In this model, $\log(\texttt{days})$ represents the log-transformed survival time, and $\beta_0$ is the intercept term. The coefficients $\beta_i$ quantify the effect of covariates on survival, including `age` (patient age), `sex` (patient sex: 1 for male, 0 for female), `tment` (treatment group: 1 for CyA, 0 for placebo), `stage` (disease stage, reference = Stage 2), `crea` (creatinine levels), `alb` (albumin levels), `bili` (bilirubin levels), `alkph` (alkaline phosphatase levels), `asptr` (aspartate transaminase levels), and `weight` (patient weight). The term $\epsilon$ represents the error term, which follows a Weibull distribution.

The Accelerated Failure Time (AFT) model using the Weibull distribution revealed significant insights into the factors influencing survival time. The model converged successfully, with fit statistics (AIC = 375.539) indicating a reasonable fit to the data. The Weibull shape parameter (shape = 1.6369) suggests an increasing hazard over time, consistent with the progressive nature of the disease. Among the covariates, treatment (`tment`) was significant ($\beta = -0.3083, p = 0.0294$), indicating that CyA treatment accelerates survival time compared to placebo. Sex (`sex`) also showed a significant effect ($\beta = 0.7787, p = 0.0001$), with males experiencing longer survival times. Bilirubin (`bili`, $\beta = -0.0051, p < 0.0001$) and albumin (`alb`, $\beta = 0.0390, p = 0.0086$) were significant predictors, where higher bilirubin reduced survival time and higher

albumin extended it. Elevated aspartate transaminase (`asptr`, $\beta = -0.0029, p = 0.0341$) also decreased survival. Other covariates, such as `age`, `crea`, `alkph`, `stage`, and `weight`, were not significant predictors ($p > 0.05$). Diagnostic checks, including negative log-log survival plots and Cox-Snell residuals, support the adequacy of the Weibull model. The linear pattern in the negative log-log plot validates the Weibull assumption, and the residuals showed no major violations, confirming the robustness of the model for this analysis.

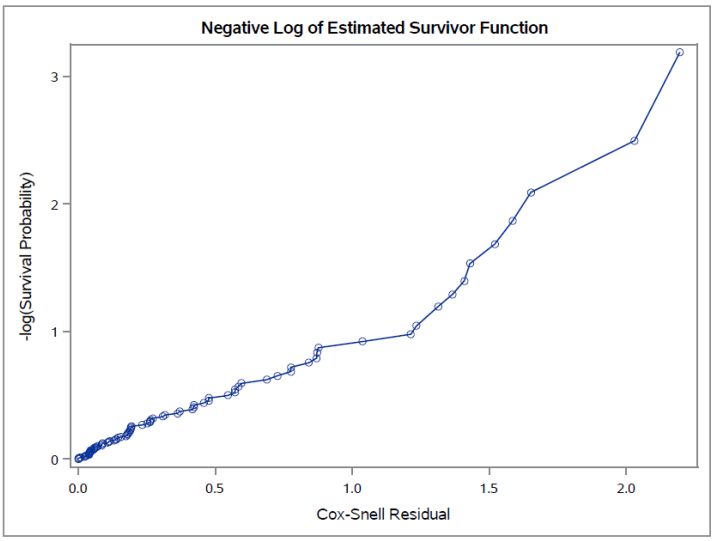
### Significant Weibull AFT Model

The Accelerated Failure Time model with significant covariates is given by:

$$\log(\texttt{days}) = \beta_0 + \beta_1 \cdot \texttt{sex} + \beta_2 \cdot \texttt{tment} + \beta_3 \cdot \texttt{alb} + \\ \beta_4 \cdot \texttt{bili} + \beta_5 \cdot \texttt{asptr} + \epsilon$$

In this model, $\log(\texttt{days})$ represents the log-transformed survival time, and $\beta_0$ is the intercept term. The coefficients $\beta_i$ quantify the effect of the significant covariates, including `sex` (patient sex: 1 for male, 0 for female), `tment` (treatment group: 1 for CyA, 0 for placebo), `alb` (albumin levels), `bili` (bilirubin levels), and `asptr` (aspartate transaminase levels). The term $\epsilon$ represents the error term, which follows a Weibull distribution.

## 6.2 Assumption Check for Weibull Regression




The diagnostic evaluation of the Weibull AFT model using negative log-log (NLL) survival plots above suggests that the Weibull assumption is largely appropriate for the data. The NLL plot exhibits a reasonably linear pattern, indicating that the hazard function aligns well with the Weibull distribution's characteristics. While some deviations are observed in specific segments, these appear minor and do not significantly compromise the modelâs fit. The overall consistency of the linear trend across most of the range provides confidence in the adequacy of the Weibull model. Additionally, the residual diagnostics further support the model's validity. The Cox-Snell residuals align well with the 45-degree reference line, demonstrating that the observed data conform to the expected survival distribution. Standardized residuals display a random scatter around zero without extreme outliers, suggesting that the model captures the underlying data patterns effectively and that no major violations of model assumptions are present. These diagnostics collectively affirm the robustness and suitability of the Weibull model for analyzing the given survival data.

## 7 Discussion/Conclusion

The analysis of the PBC3 dataset provided key insights into the survival outcomes of patients with primary biliary cirrhosis (PBC) based on treatment with Cyclosporin A (`CyA`) and other covariates. Kaplan-Meier analysis revealed no statistically significant improvement in survival for patients treated with `CyA` compared to placebo ($p = 0.7813$). However, significant survival differences were observed based on disease stage, sex, and gastrointestinal bleeding history. The Cox Proportional Hazards model confirmed these findings, identifying `sex`, `bili`, `alb`, and `stage` as significant predictors of survival, while `tment` had a non-significant effect. Advanced disease stages and higher bilirubin levels were associated with poorer survival outcomes, while higher albumin levels were protective.

The Accelerated Failure Time (AFT) model using the Weibull distribution provided additional insights, confirming the significance of `sex`, `tment`, `alb`, `bili`, and `asptr`. Diagnostic checks, including Cox-Snell residuals and standardized residuals, validated the adequacy of the Weibull model, with no severe violations observed. Negative log-log survival plots further supported the appropriateness of the Weibull assumption, showing a largely linear pattern with only minor deviations.

### 7.1 Pros and Cons of the Methods

The Kaplan-Meier estimator provided a simple and intuitive method for comparing survival probabilities across groups, but it could not account for multiple covariates simultaneously. The Cox model allowed for the inclusion of multiple covariates, providing hazard ratios that quantify the effect of predictors on survival. However, the violation of the proportional hazards assumption for `stage 1` indicated that the Cox model may not fully capture time-dependent effects. The AFT model, on the other hand, directly modeled survival time and performed well with the Weibull distribution, offering a robust alternative when the proportional hazards assumption was in question.

### 7.2 Model Assumptions

While most assumptions were satisfied, the violation of proportional hazards for `stage 1` in the Cox model suggests the need for refinements, such as stratification or the inclusion of time-dependent covariates. These limitations highlight the importance of verifying assumptions to ensure model validity. The AFT model assumptions were well-supported by residual diagnostics, affirming the suitability of the Weibull distribution.

### 7.3 Future Directions

Further analyses could explore alternative parametric distributions for the AFT model or stratified Cox models to address the violation of assumptions for `stage 1`. Interaction terms between covariates, such as between `tment` and `stage`, could provide additional insights into the effect of treatment across different stages of the disease. Additionally, incorporating time-dependent covariates in the Cox model could address non-proportional hazards and improve model accuracy.

This comprehensive analysis underscores the complexity of survival outcomes in PBC and highlights the importance of leveraging multiple analytical methods to derive robust conclusions.

## References


[1] Liu, X. (2012). *Survival Analysis: Models and Applications*. ProQuest Ebook Central. Available at: `https://ebookcentral.proquest.com/lib/osu/detail.action?docID=939628`. [Accessed 11 Dec. 2024].

[2] Lee, E. and Wang, J.W. (2013). *Statistical Methods for Survival Data Analysis*. Wiley Online Library. Available at: `https://onlinelibrary-wiley-com.oregonstate.idm.oclc.org/doi/book/10.1002/0471458546`. [Accessed 11 Dec. 2024].

[3] Therneau, T.M. and Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag. Available at: `https://link.springer.com/book/10.1007/978-1-4757-3294-8`. [Accessed 11 Dec. 2024].

[4] Collett, D. (2015). *Modelling Survival Data in Medical Research*. 3rd ed. CRC Press. Available at: `https://www.routledge.com/Modelling-Survival-Data-in-Medical-Research/Collett/p/book/9781439856789`. [Accessed 11 Dec. 2024].

[5] OpenAI (2024). Generated by ChatGPT, OpenAI. [Accessed 11 Dec. 2024].

Table 1: Summary Statistics of Key Variables

| Variable | N | Mean | Median | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|---|
| age | 349 | 54.0888 | 55.0000 | 9.9438 | 26.0000 | 75.0000 |
| crea | 342 | 77.9893 | 78.0000 | 18.2781 | 35.0000 | 199.0000 |
| alb | 343 | 38.3663 | 38.7100 | 5.6627 | 20.6300 | 58.0000 |
| bili | 349 | 45.4730 | 19.2700 | 67.6521 | 2.3330 | 453.1000 |
| alkph | 349 | 996.5439 | 812.3000 | 751.5998 | 66.3300 | 5108.0000 |
| asptr | 346 | 95.3791 | 87.6700 | 52.9139 | 10.5000 | 316.3000 |
| weight | 339 | 60.3422 | 59.0000 | 10.1966 | 38.0000 | 98.0000 |
| days | 349 | 942.6533 | 929.0000 | 514.0021 | 1.0000 | 2146.0000 |

Table 4: Combined Analysis of Maximum Likelihood Estimates and Cox Proportional Hazards Model for Time to Failure

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| tment | 1 | -0.43085 | 0.23060 | 3.3312 | 0.0680 | 0.650 |
| age | 1 | 0.01738 | 0.01304 | 1.7759 | 0.1827 | 1.018 |
| sex | 1 | 0.95288 | 0.29552 | 10.1209 | 0.0015 | 2.593 |
| bili | 1 | 0.00923 | 0.00142 | 42.4623 | <0.0001 | 1.009 |
| alb | 1 | -0.06593 | 0.02451 | 7.2353 | 0.0071 | 0.936 |
| crea | 1 | -0.01464 | 0.00813 | 2.0795 | 0.1490 | 0.985 |
| stage 1 | 1 | -14.67542 | 606.57545 | 0.0060 | 0.9807 | - |
| stage 3 | 1 | 0.44003 | 0.39652 | 1.2315 | 0.2671 | 1.553 |
| stage 4 | 1 | 0.74855 | 0.36620 | 4.1796 | 0.0409 | 2.114 |
| **Cox Proportional Hazards Model for Time to Failure (All Variables)** | | | | | | |
| stage NA | NA | 0.82032 | 0.42712 | 3.6886 | 0.0548 | 2.271 |
| unit 1 | 1 | -0.50188 | 0.56699 | 0.7835 | 0.3761 | 0.605 |
| unit 2 | 1 | -0.48446 | 0.42193 | 1.3218 | 0.2503 | 0.616 |
| unit 4 | 1 | -0.33037 | 0.36046 | 0.8292 | 0.3615 | 0.718 |
| unit 5 | 1 | 0.48038 | 0.64531 | 0.5556 | 0.4558 | 1.616 |
| gibleed | 1 | 0.00257 | 0.00637 | 0.0041 | 0.9487 | 1.002 |

Table 3: Combined Analysis of Maximum Likelihood Estimates and Proportional Hazards Assumption Check

| Parameter | DF | Parameter Estimate | Standard Error | Chi-Square | Pr > ChiSq | Hazard Ratio |
|---|---|---|---|---|---|---|
| **Analysis of Maximum Likelihood Estimates** | | | | | | |
| tment | 1 | -0.43085 | 0.23060 | 3.3312 | 0.0680 | 0.650 |
| sex | 1 | 0.95288 | 0.29552 | 10.1209 | 0.0015 | 2.593 |
| age | 1 | 0.01738 | 0.01304 | 1.7759 | 0.1827 | 1.018 |
| bili | 1 | 0.00923 | 0.00140 | 42.4623 | <0.0001 | 1.009 |
| alb | 1 | -0.06593 | 0.02451 | 7.2353 | 0.0071 | 0.936 |
| crea | 1 | -0.01464 | 0.00813 | 2.3390 | 0.1263 | 0.985 |
| stage 1 | 1 | -14.67542 | 66.57545 | 0.0485 | 0.8260 | - |
| stage 3 | 1 | 0.44003 | 0.39652 | 1.2315 | 0.2671 | 1.553 |
| stage 4 | 1 | 0.74855 | 0.36620 | 4.1796 | 0.0409 | 2.114 |
| **Proportional Hazards Assumption Check for All Variables** | | | | | | |
| stage NA | NA | 0.82032 | 0.42712 | 3.6886 | 0.0548 | 2.271 |
| unit 1 | 1 | -0.50188 | 0.56699 | 0.7835 | 0.3761 | 0.605 |
| unit 2 | 1 | -0.48446 | 0.42193 | 1.3218 | 0.2503 | 0.616 |
| unit 4 | 1 | -0.33037 | 0.36046 | 0.8292 | 0.3615 | 0.718 |
| unit 5 | 1 | 0.48038 | 0.64531 | 0.5556 | 0.4558 | 1.616 |
| gibleed | 1 | 0.00257 | 0.00637 | 0.0041 | 0.9487 | 1.002 |

**The UNIVARIATE Procedure**

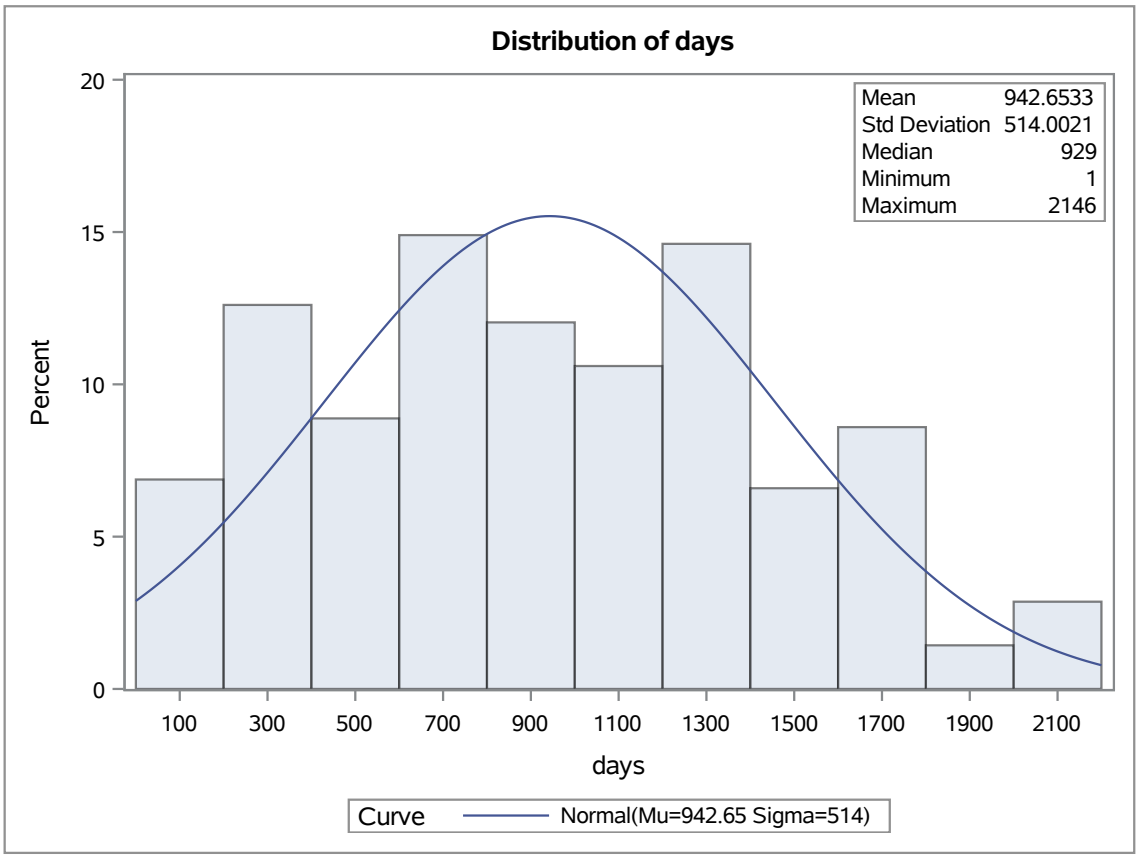



Figure 1: Histogram of Time to Failure (days)

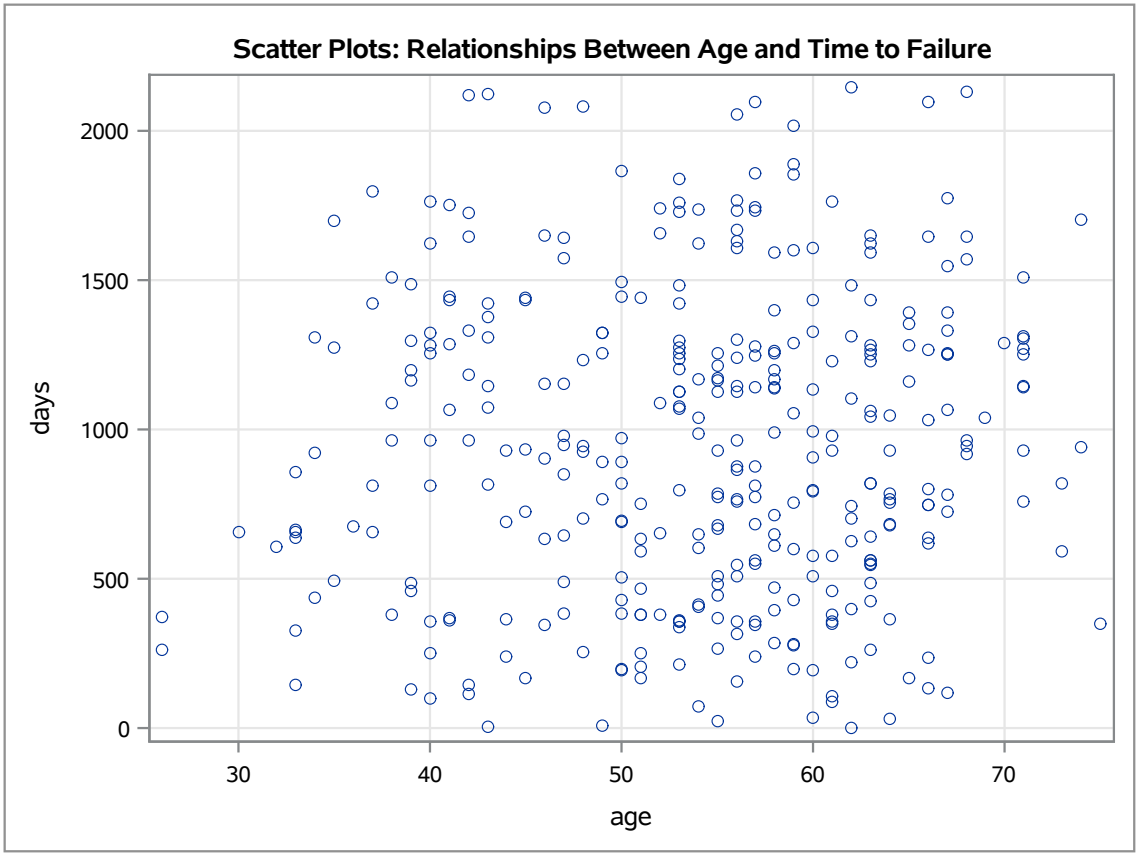



Figure 2: Relationships Between Age and Time to Failure


Figure 3: Relationships Between Age and Time to Failure

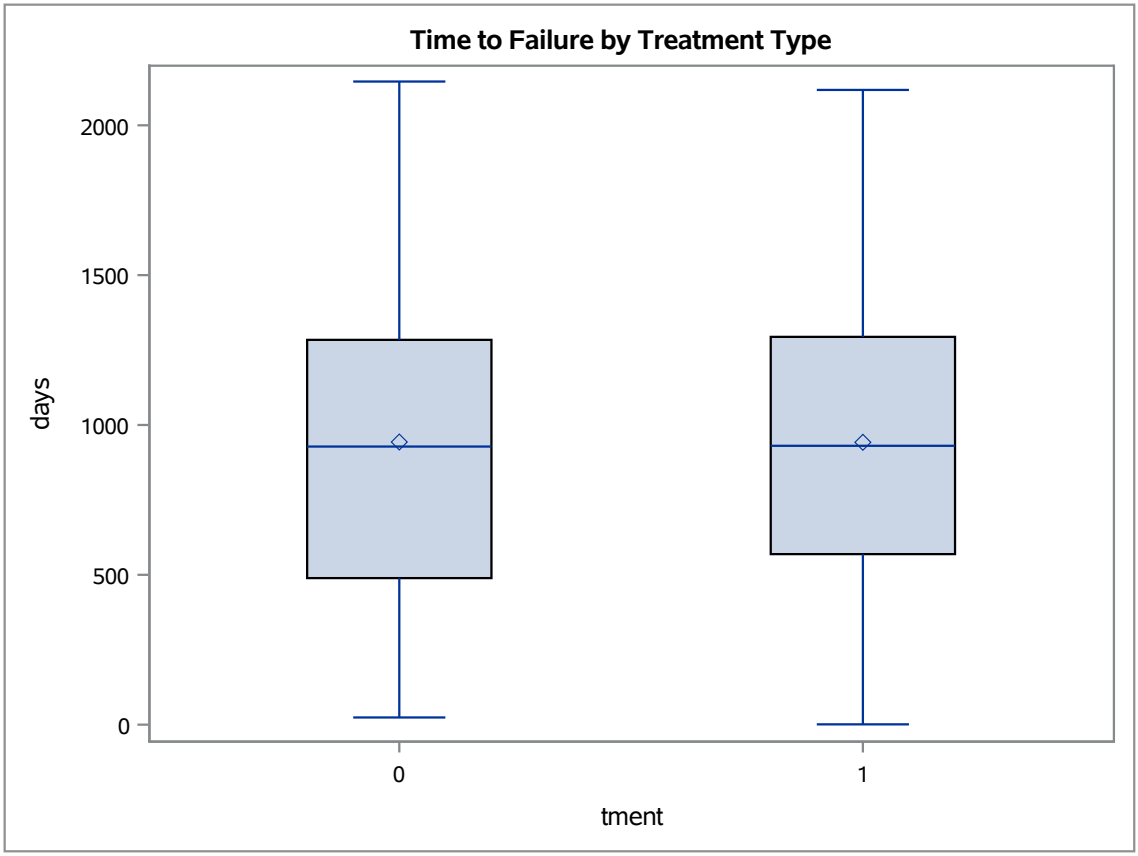



Figure 4: Boxplot of treatments

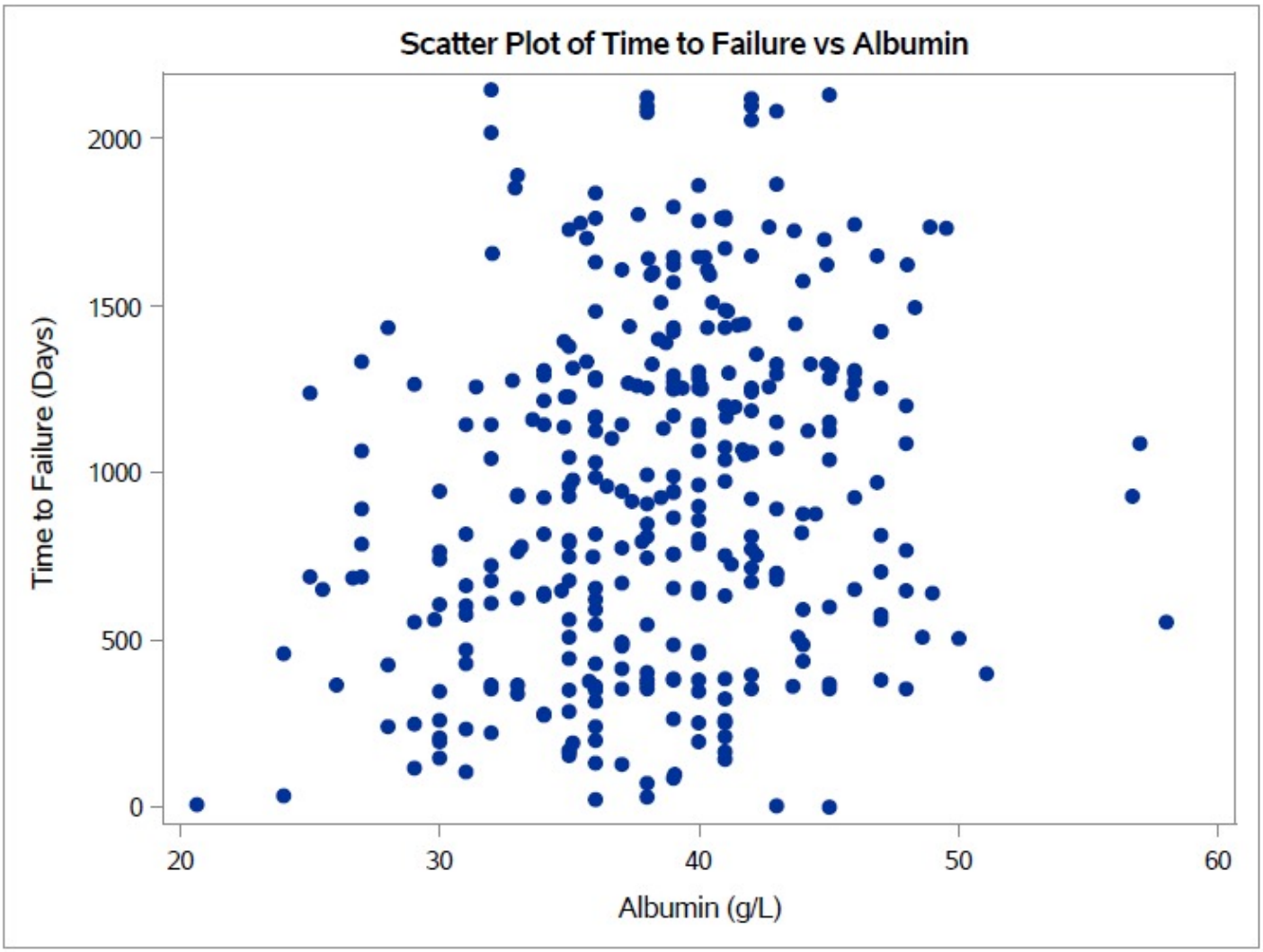



Figure 5: Scatter Plot of Time to Failure vs. Albumin

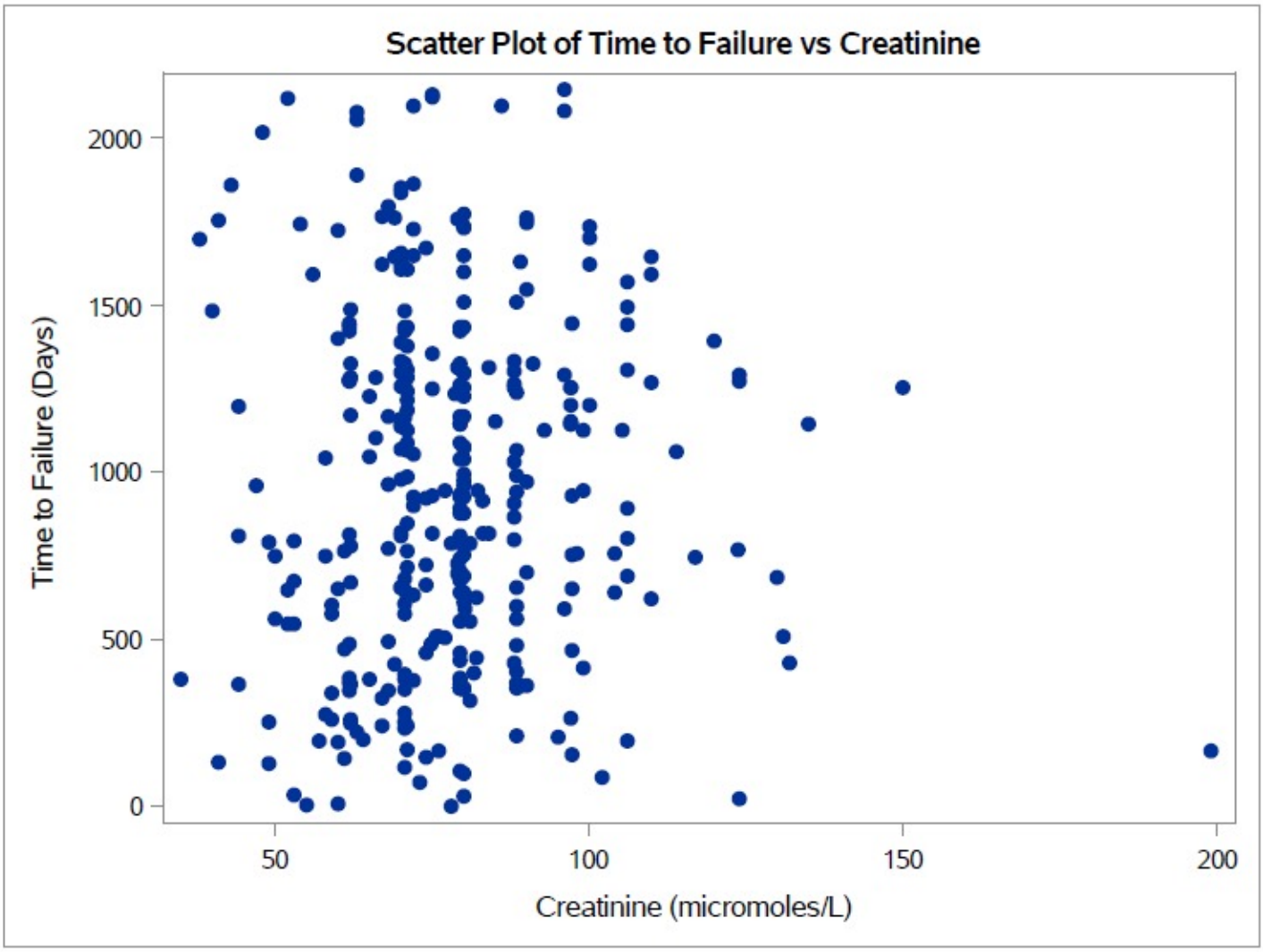



Figure 6: Scatter Plot of Time to Failure vs. Creatinine

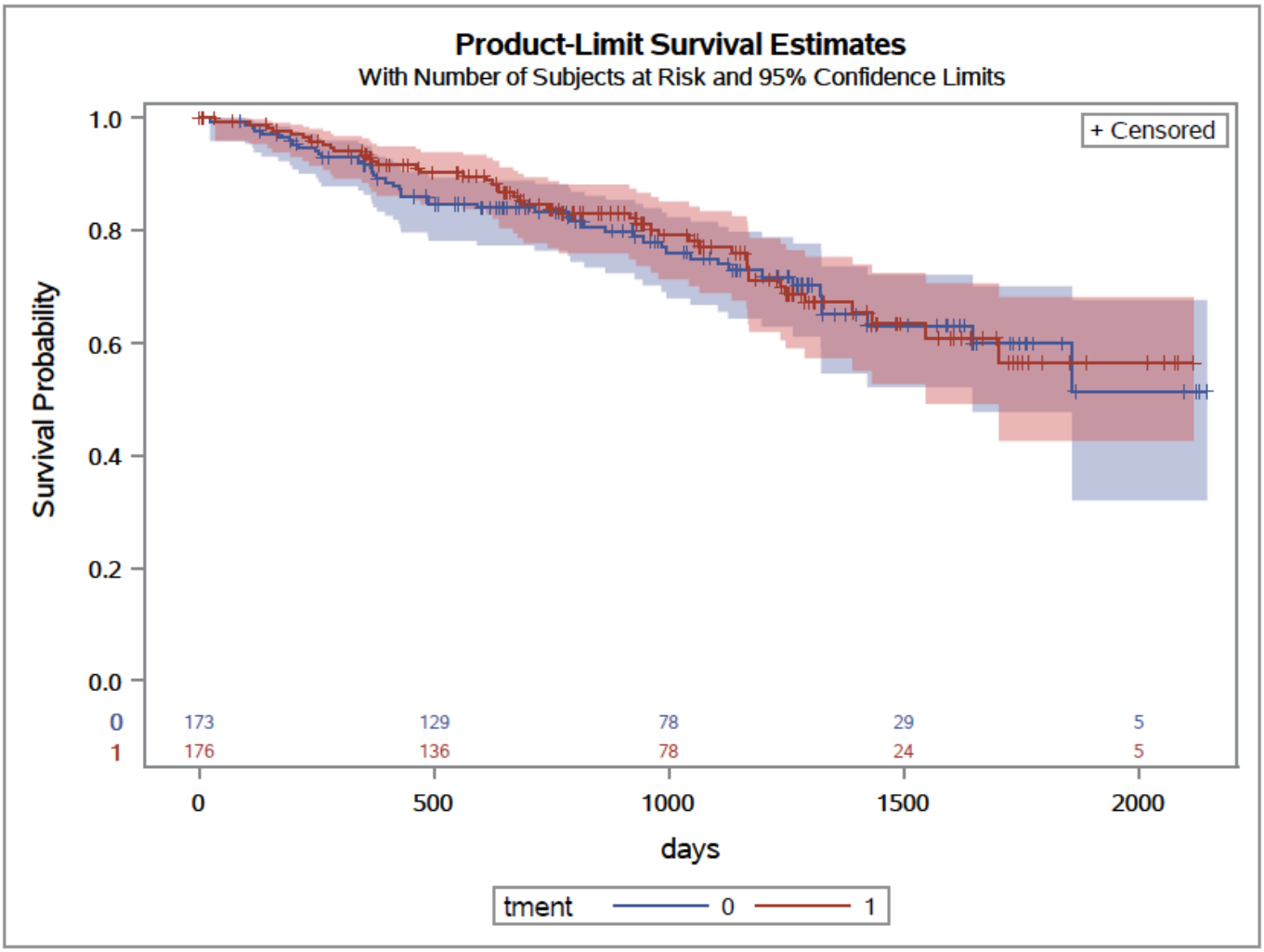



Figure 7: 95% Confidence Limit Kaplan-Meier survival for Treatment

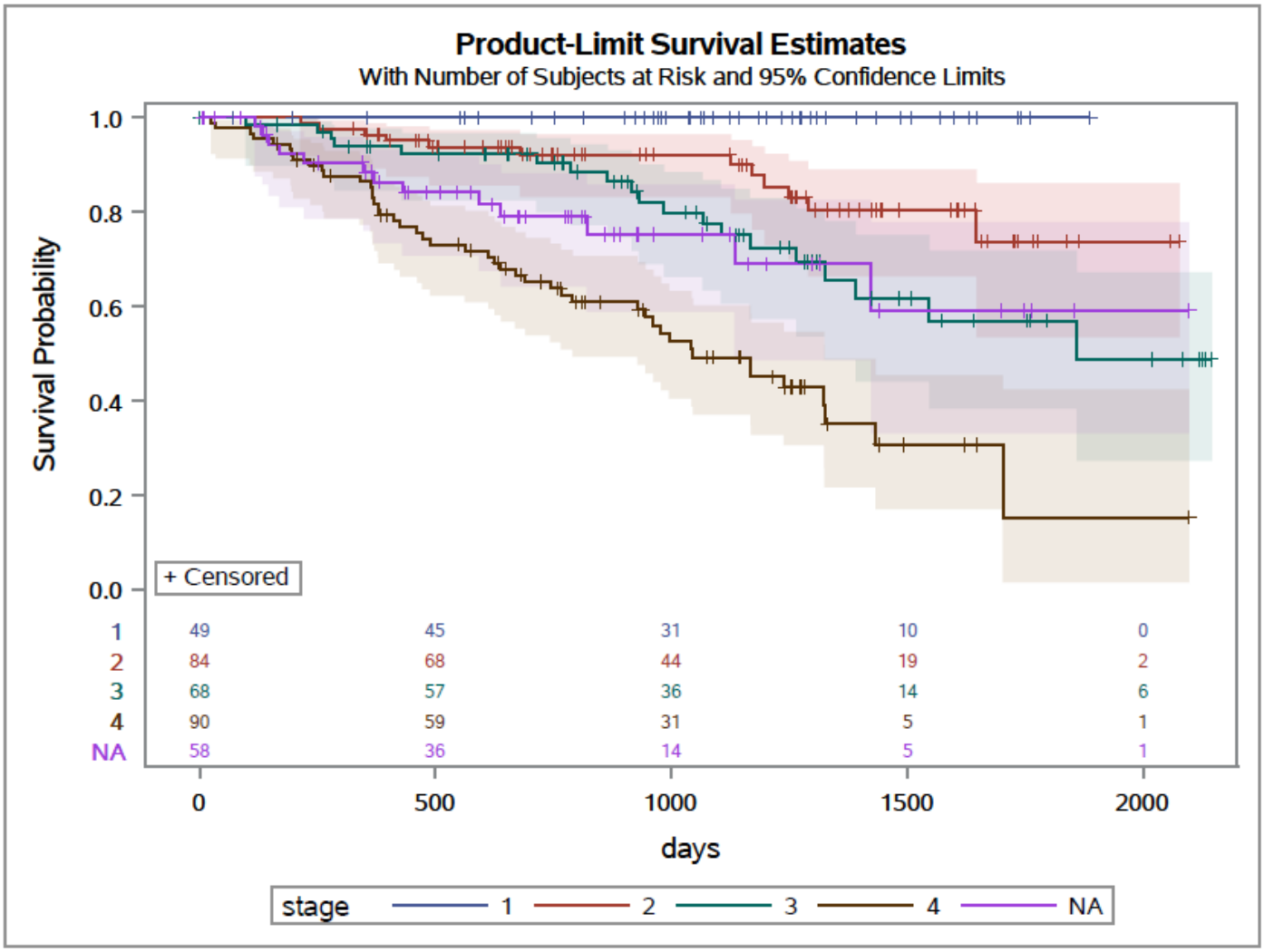



Figure 8: 95% Confidence Limit Kaplan-Meier survival for Stages

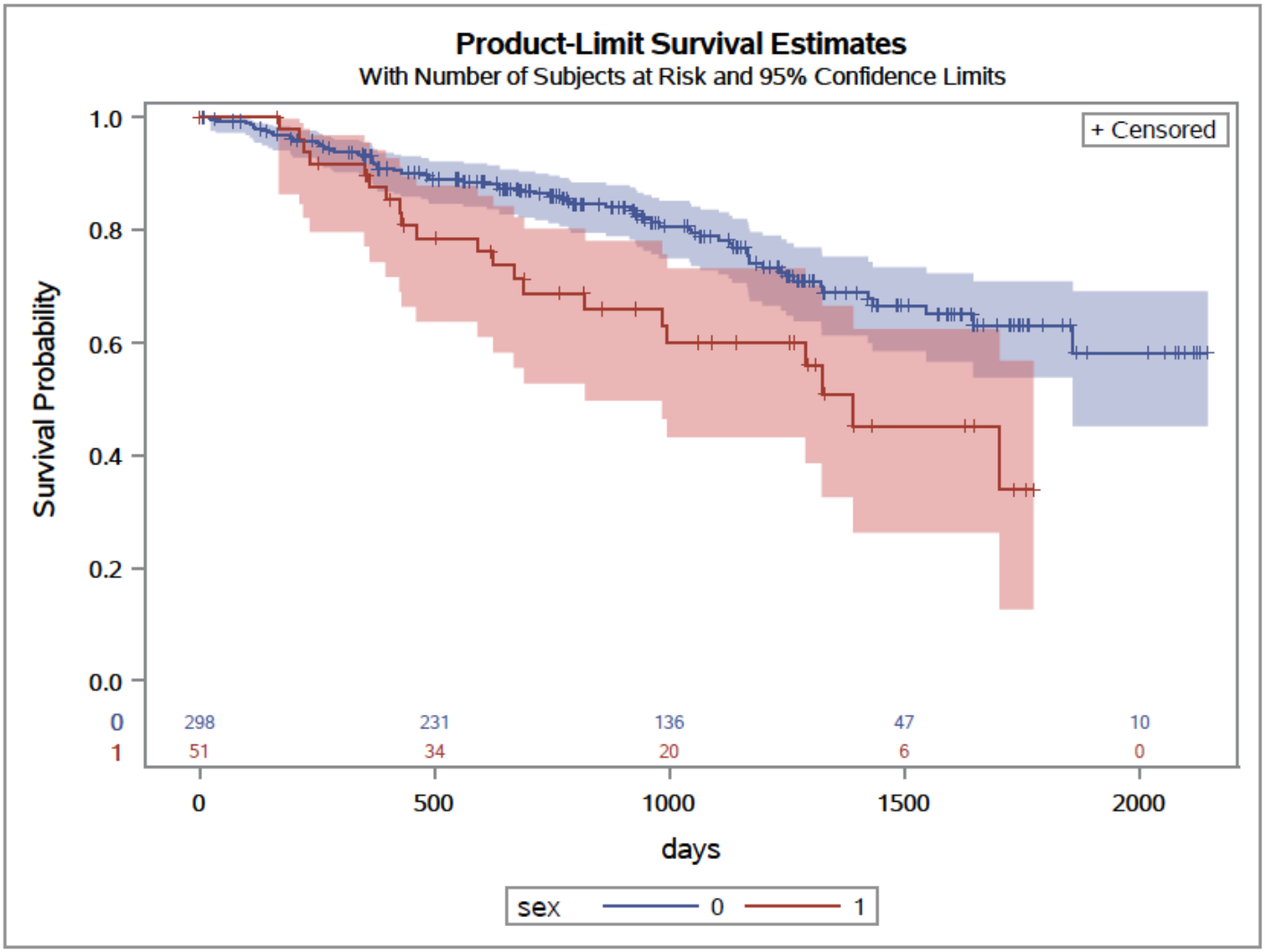



Figure 9: 95% Confidence Limit Kaplan-Meier survival for Sex

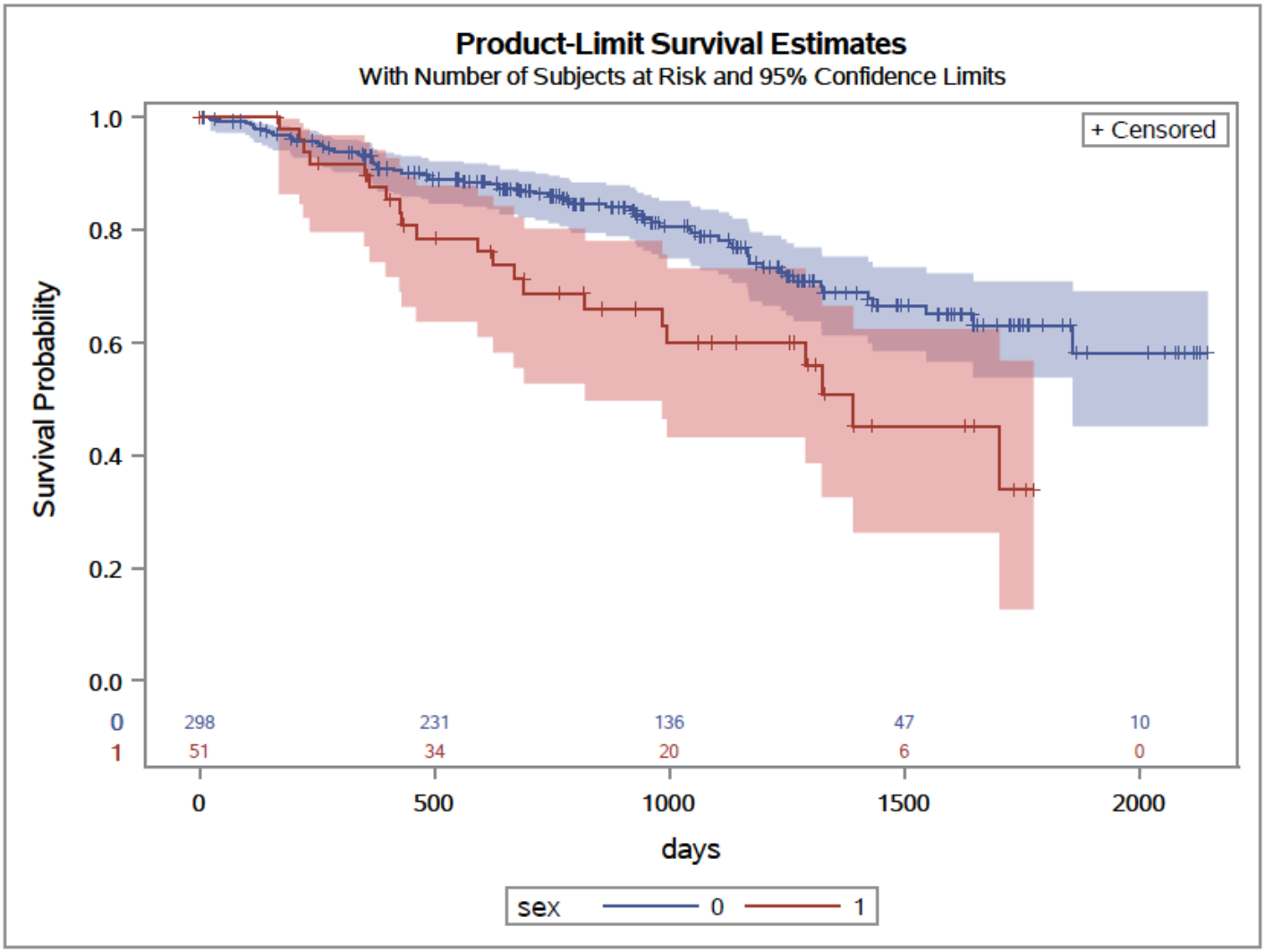



Figure 10: 95% Confidence Limit Kaplan-Meier survival for Sex

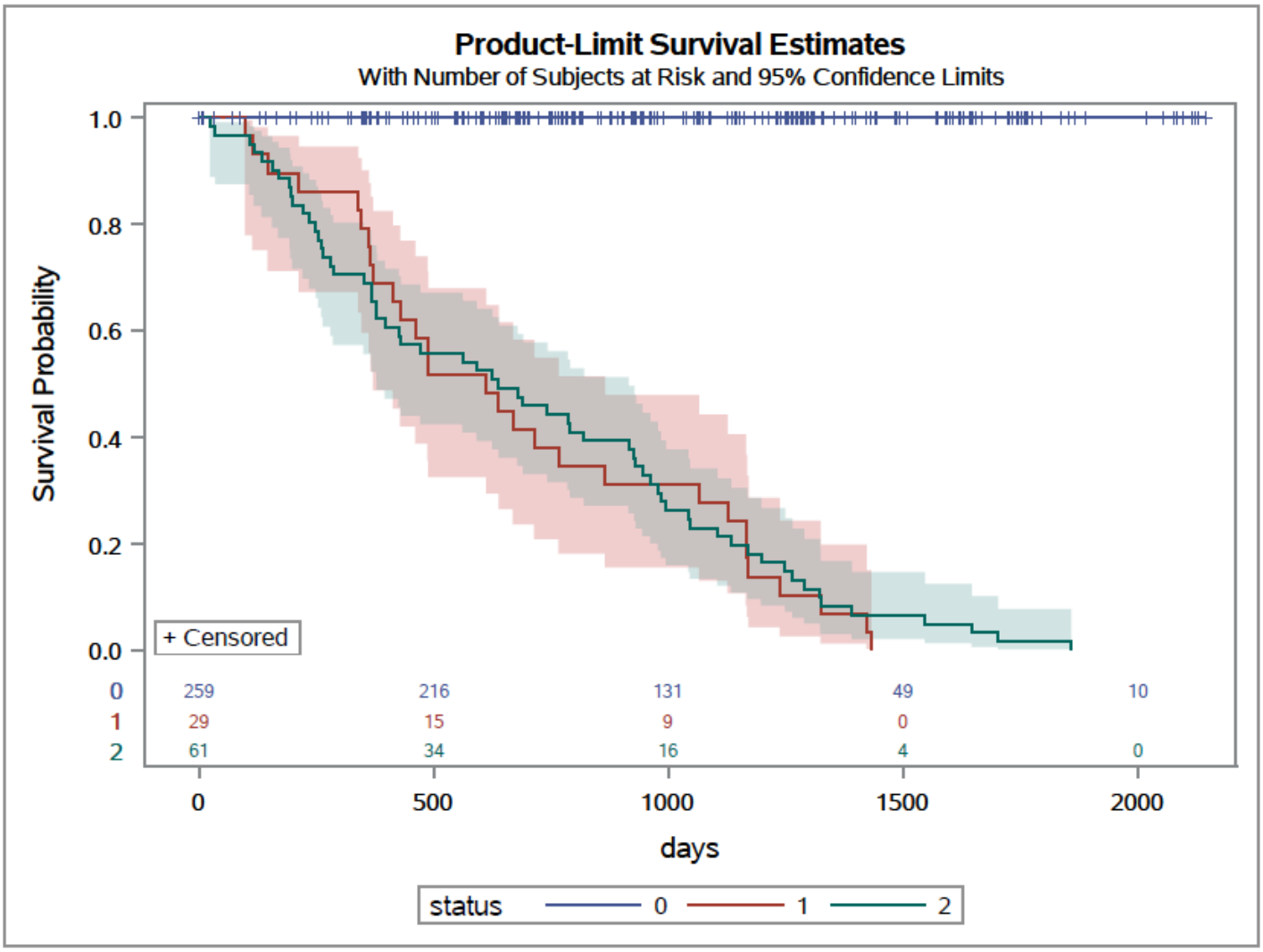



Figure 11: 95% Confidence Limit Kaplan-Meier survival for Sex