Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Chenyue Yang Date

Effect Of 5-FU Bolus on Survival in Patients with Metastatic Colorectal Cancer Treated with Combination Chemotherapy

By

Chenyue Yang

Master of Science in Public Health

Biostatistics and Bioinformatics

Zhengjia (Nelson) Chen, PhD

Committee Chair

Tianwei Yu, PhD

Committee Member

Effect Of 5-FU Bolus on Survival in Patients with Metastatic Colorectal Cancer Treated with Combination Chemotherapy

By

Chenyue Yang

B.A.

Emory University

2017

Thesis Committee Chair: Zhengjia (Nelson) Chen, PhD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Biostatistics and Bioinformatics

2019

Abstract

Effect Of 5-FU Bolus on Survival in Patients with Metastatic Colorectal Cancer Treated with Combination Chemotherapy

By Chenyue Yang

Background: The standard treatment for metastatic colorectal cancer (CRC) includes the use of FOLFOX chemotherapy in combination with 5-fluorouacil (5-FU) bolus. However, the use of 5-FU bolus has been associated with toxicity effects in other cancers. The primary objective of this study is to compare the progression-free survival (PFS) of CRC patients receiving 5-FU bolus to PFS of CRC patients omitting 5-FU bolus. Secondary objectives include comparing overall survival (OS), toxicity events, dose reductions due to toxicity, and genetic mutations between two patient groups.

Methods: The dataset included 110 metastatic CRC patients, with 74 in the with-bolus group, and 36 in the no-bolus group. Kaplan-Meier plots were used to examine the PFS and OS of the patients, and log-rank tests were used to compare survival pattern between two treatment groups. Univariate analysis was done on PFS and OS, and variables that were significant in univariate analysis were included in multivariate Cox proportional hazard models. Forward selection was used to choose the best model, and residual analysis was performed.

Results: Overall, 66.36% of patients died, and 74.55% had disease progression during the study period. 72.22% of patients died, and 69.44% had disease progression in no-bolus group. 63.51% of patients died, and 77.03% had disease progression in with-bolus group. Among 110 patients, percentage of patients censored were 25.45% for PFS and 33.64% for OS. The log-rank test p-value between two treatment groups was 0.1988 (>0.05) for PFS, and 0.2856 (>0.05) for OS. After adjusting for diabetes status, the hazard of disease progression for no-bolus group is 1.581 (0.958, 2.612) times the hazard of disease progression for with-bolus group ($p = 0.073$).

Conclusion: The unadjusted hazard of disease progression and death are not significantly different between no-bolus group and with-bolus group. However, after adjusting for diabetes status, the effect of using bolus is beneficial under $\alpha = 0.1$, but not under $\alpha = 0.05$. Given the relatively small sample size of this study, further consideration is needed before omitting the bolus from metastatic CRC treatments.

Effect Of 5-FU Bolus on Survival in Patients with Metastatic Colorectal Cancer Treated with Combination Chemotherapy

By

Chenyue Yang

B.A.

Emory University

2017

Thesis Committee Chair: Zhengjia (Nelson) Chen, PhD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Science in Public Health

in Biostatistics and Bioinformatics

2019

TABLE OF CONTENTS

I. INTRODUCTION

Disease Background

Colorectal cancer (CRC) is cancer that arises in colon, or rectum found in lower parts of the digestive system [1]. Even though CRC dropped from the third most common cancer to fourth from 2017 to 2018, the disease remains to be the second most common cause of cancer death in the United States [2, 3]. The American Cancer Society determined that about 75,610 new male cases and 64,640 new female cases of CRC are estimated to arise in the United States in 2018 [4]. Moreover, CRC accounts for about 8% of death among all cancer cases [4]. Even though CRC is highly treatable when the disease is still localized, only 39% of patients do not develop metastasis within 5 years of diagnosis [5]. The 5-year survival is significantly higher at 90% when the cancer is diagnosed before metastasis compared to overall 5-year survival at 65% [6]. Demographic factors that may influence the risk and treatment outcomes of CRC include age, gender, and race; the risks of disease increase with age, males have 30% higher incidence rates compare to females, and Non-Hispanic Blacks are most at risk for CRC among all races [1]. Studies have also shown that sidedness is clinically important when examining the Progression-free Survival of patients with metastatic CRC [7].

Over 95% of CRC arise from mutations in the inner lining of colorectal regions[1]. Genetic alterations leading to the disease include both activation of proto-oncogenes and loss of function of tumor suppressor genes [8]. KRAS mutation is one of the major gene alterations involved in CRC that causes instabilities in the chromosome, eventually contributing to cancer progression [8]. In addition, many patients also develop BRAF mutations that

constitutively activate the mitogen-activated protein kinase (MAPK) pathways, aiding in cancer cell division and growth [9].

Current Treatments

Different methods exist for treating CRC; traditional treatments include surgery and radiation therapy. However, those methods do not completely eliminate metastatic CRC, and are often associated with side effects including diarrhea, intestine bleeding, fatigue, skin reactions, and vomiting [10]. The use of chemotherapy and targeted therapy provides more treatment options and can be delivered to patients through non-invasive ways such as oral ingestion and intravenous administration. Common chemotherapeutic drugs to treat CRC include 5-fluorouacil (5-FU), which interferes with DNA synthesis through folate metabolic pathways; adjustments to 5-FU also exist, including FOLFOX (5-FU, leucovorin, and oxaliplatin), and FOLFIRI (5-FU, leucovorin, and irinotecan) [10, 11]. 5-FU medications may also be combined with targeted therapies such as Cetuximab and Bevacizumab that targets endothelial growth factor [12]. However, medications containing those cancer drugs may only be effective on a subgroup of CRC patients, and the outcome depends on patient's genetic mutation status; for instance, studies have shown that KRAS mutation status influences patient's response to Cetuximab therapy, and we therefore recorded mutation status of varies genes for each patient and considered them as covariates in this study [13].

Current Situation

Overall survival (OS) is the survival period between study time origin and event (death) time of the patient; it is the traditional standard for assessing effectiveness of cancer treatment. However, OS sometimes require long follow-up periods, and does not take account into secondary events such as disease progression. Progression-free Survival (PFS) examines the survival period between study time origin to first progression of the disease and is sometimes used as an alternative or in addition to OS in clinical trials. Studies have shown that PFS is a decent substitute to OS for survival analysis involving colorectal cancers [14].

The methods for delivery may influence how cancer drugs are transported and metabolized, leading to different patient responses. A randomized trial compared monthly low-dose fluorouracil bolus with bimonthly high-dose fluorouracil bolus and continuous administration for colorectal cancer, and found reduction in toxicity response under bimonthly treatment, but no difference in patient survival between the two groups [15]. A similar study examined the toxicity of delivering 5-FU drug through bolus compared to continuous infusion for local rectal cancer [16]. The study found significantly less diarrhea, weight loss, and White Blood Cell toxicity in the infusion group, who also have nearly twice the drug tolerance compared to the bolus group [16]. Emory Winship Cancer Institute did comparable studies on pancreatic cancer treated with FOLFIRINOX and concluded the elimination of 5-FU bolus did not affect drug efficacy, but reduced drug toxicity [17].

Purpose Statement

This retrospective study is motivated by positive results in studies substituting the traditional bolus administration in varies cancers. In this study, colorectal cancer (CRC) patients received FOLFOX treatment either with 5-FU bolus, or without 5-FU bolus. The primary goal of this study is to compare the progression-free survival (PFS) of CRC patients receiving 5-FU bolus to PFS of CRC patients omitting the 5-FU bolus. Secondary objectives include comparing the overall survival (OS) between the two patient groups. Moreover, the toxicity effects of two treatments are measured through grade 3 or 4 neurotoxicity events. In addition, dose reduction due to toxicity in two groups are examined. Finally, genetic mutations including KRAS and BRAF and sidedness of cancer are analyzed in the two groups. Survival analysis including Kaplan-Meier, log-rank test, and cox proportional hazard models will be used to achieve these objectives. Statistical tests will also be used to analyze and compare continuous and categorical covariates between the two treatment groups.

II. METHODS

Study Design and Patient Enrollment

The dataset was constructed based on a retrospective study for CRC patients treated at the Emory Winship Cancer Institute. Patient data was entered at different time points ranging between January 01, 2010 and study end date on June 27, 2018. All subjects were treated with FOLFOX chemotherapy. Among the selected patients, 100 received the drug together with 5-FU bolus, and 100 received the drug without 5-FU bolus. After eliminating incomplete observations and disqualified subjects, the final dataset contains 74

observations for the treatment group with 5-FU bolus, and 36 observations for the treatment group without 5-FU bolus.

The inclusion criteria for patients in this study include clinically diagnosed metastatic CRC, proper administration of at least two doses of FOLFOX treatment, and no history of dihydropyrimidine dehydrogenase (DPD) or other enzyme deficiencies that could potentially affect the response to 5-FU bolus. Despite having the above inclusion criteria, patients were excluded from this study if they are pregnant individuals, prisoners, adolescents younger than 18 years old, patients on maintenance therapy, clinical trials, or capecitabine, patients who develop severe side effects from 5-FU bolus, and patients who experience poorly differentiated side effects.

Each patient present in the final dataset was assigned a unique id, and confidential information such as name and medical record number are permanently deleted from the database before any analysis.

Variables

The original dataset has 34 variables containing information about the subjects' demographics, disease progression, survival, treatment cycles, side effects, gene mutation status, and various disease status.

Demographical variables include patient Age (in years), Gender ('male', 'female'), and Race ('White', 'Black', 'Hispanic', 'Asian', 'Others'). For gender, value '23' was recoded

as missing. For race, ‘Hispanic’ and ‘Asian’ are collapsed into ‘Others’ due to the small number of occurrences in each category.

Disease progression and death information for each subject were recorded in variables First Cycle Date, Last Cycle Date, Death (0, 1), Death date, and Date of Radiological Progression. The study time origin is set to be the First Cycle date for each patient, and the study end date is June 27, 2018. Based on this information, Progression-free Survival (PFS) and Overall Survival (OS) were created as two new variables. Radiological progression and death were treated as events for calculating PFS and OS. If radiological progression was observed, PFS was calculated as the time between Date of Radiological Progression and First Cycle Date. If death was observed but radiological progression is censored, PFS was calculated as the time between Death Date and First Cycle Date. If both radiological progression and death were censored, PFS was calculated as the time between Study End Date and First Cycle Date. If death was observed, OS was calculated as the time between Death Date and First Cycle Date. If death was censored, OS was calculated as the time between Study End Date and First Cycle Date. Death indicator and disease progression indicator were coded as 0 for censored subjects, and were coded as 1 for observed subjects.

Information about treatment cycles were recorded in variables Number of Complete Cycles, Length of Treatment Delays (days), and Number of Treatment Delays. Side effects experienced by patients were recorded in variables Grade 3 Absolute Neutrophil Count (ANC) for $500\ mm^3 < ANC < 1000\ mm^3$, and Grade 4 ANC for $ANC < 500\ mm^2$ events. Moreover, variables Treatment Delayed (‘yes’, ‘no’), Reason for Delay

('Thrombocytopenia', 'Neutropenia', 'Other'), Growth Factor Added ('yes, 'no') were also recorded to indicate the effects of toxicity on patients.

The mutation status for various genes were coded as 0 for no mutation and 1 for mutation. These genes include CHF, KRAS, BRAF, CKD, HTN, MSS, MAb, Bev, Cetux, Pani, PNI. The dataset also contains disease-related characteristics including Sidedness of Cancer, Eastern Cooperative Oncology Group (ECOG) Scores (0, 1, 2), and Diabetes ('yes', 'no').

Statistical Analysis

All statistical analysis for this study were performed via SAS Version 9.4 (SAS Institute, Inc., Cary, North Carolina). The significant level was set at $\alpha = 0.05$ for all statistical tests in this study.

Descriptive statistics for all variables mentioned above were summarized for patients overall, and for two treatment groups separately. For categorical variables, the count of patients was recorded with the column percentage in each category within a group. For continuous variables, the mean was recorded for each variable. P-values were used to determine if any variables are significantly different between with-bolus group and no-bolus group. Chi-Square Test was used for categorical variables with 2 levels, ANOVA Test was used for categorical variables with >2 levels, and two-sample t-Test was used for continuous variables to calculate p-values. Fisher's Exact Test was used as a substitute to Chi-Square Test when any cell count is less than 5.

Kaplan-Meier estimators were calculated to examine the survival functions for the overall patient cohort, and for with-bolus group vs. no-bolus group separately. The Kaplan-Meier estimator was constructed based on the risk set and failure incidences at time *t*. The risk set *R(t)* represents the set of subjects at risk at time *t*, and was obtained as follows for each time point *t*:

$$\mathrm{R(t)} = \{\mathrm{i}: \mathrm{T_i} \geq \mathrm{t}\},$$

Here, T_i is the event or censor time of subject *i*. Let n_t be the size of risk set, and d_t be the number of events (disease progressions) at time *t*. The survival function under Kaplan-Meier estimator was then calculated as:

$$\hat{\mathrm{S}}_{\mathrm{KM}} = \prod_{i:t(i)\leq t} (1 - \frac{d_i}{n_i})$$

The above function was plotted as Kaplan-Meier curves against time *t* to visualize any difference of PFS between two treatment groups.

Univariate analysis was then carried out to identify the relevant variables associated with PFS. Log-rank test was used to examine if the survival function for PFS is significantly different between each level of the categorical variables. For two-sample log-rank test used in this study, the null hypothesis was H_0: $S_1(t) = S_2(t)$, tested against the alternative H_A: $S_1(t) \neq S_2(t)$. Where $S_1(t)$ and $S_2(t)$ are the survival functions of PFS for different levels in the tested variable. The test statistics was calculated and compared to standard normal distribution for significance:

$$\mathrm{Z} = \frac{\mathrm{O} - \mathrm{E}}{\sqrt{V}} \sim N(0,1)$$

O is the total number of observed failures for group 2, and *E* is the expected number of failures for group 2 under H_0, calculated by multiplying the total number of patients from group 2 with hazard of the study cohort. *V* represents the conditional variance of group 2 under H_0.

As further steps of univariate analysis, single-variable Cox Proportional-Hazard Models was fitted for each predictor variable on PFS:

$$h(t|X) = h_0(t) \cdot e^{\beta_1 x_1}$$

The above Cox Model has no intercept, and $h_0(t)$ represents the hazard function for baseline level of the variable. e^{β_1} can be interpreted as the hazard ratio of the variable relative to the baseline level at given *t*. Global Wald test was used to calculate the p-value for null hypothesis H_0: $\beta_1 = 0$ against the alternative H_A: $\beta_1 \neq 0$ for all single-variable Cox Models.

Multivariate analysis was then performed based on results from the univariate analysis. Variables with significant p-values in the univariate analysis were included in multivariate Cox Model:

$$h(t|X) = h_0(t) \cdot e^{\sum_{i=1}^{p} \beta_i x_i}$$

e^{β_i} can be interpreted as the hazard ratio of variable *i* relative to the baseline level at given *t*, controlling for other variables in the model. The variable 'Group' representing treatment groups was forced into the model for PFS even when it was not significant univariately, as treatment effect is the primary focus of this study. Forward selection was then implemented on the model including all variables that were significant univariately, and the best

multivariate model was chosen based on AIC. The steps for forward selection are as follows:

1. Fit a Cox Model on PFS with only the primary variable Group (treatment group) as predictor. Run a global test on the model without adjusting for other covariates.
2. Fit all 2-variable models by adding another covariate to the model and calculating the AIC. Run a local test for each variable and compare the p-value with specified significance level (0.05) for entry in the model. Repeat for all chosen covariates and select the one that resulted in the smallest AIC to add to the model.
3. Repeat the step above until the AIC is not further improved and no variable can be entered in the model. The resulting model with smallest AIC is the final model.

Only main effects models were considered without interaction term due to limitations set by the sample size. Global test for H_0: $\boldsymbol{\beta} = 0$ vs. H_A: $\boldsymbol{\beta} \neq 0$ was used to determine if all coefficients are significantly different from zero. Local Wald tests for H_0: $\beta_i = 0$ vs. H_A: $\beta_i \neq 0$ were also used to test for each coefficient separately.

The Cox PH Models used are based on several assumptions. First, non-informative censoring should be satisfied, where the censoring status of patients is independent from factors related to survival. Moreover, the models assume proportional hazard, meaning the ratio of hazard function at all levels of *t* remains constant. The proportional hazard assumption will be evaluated for each variable included in the final model via Mallow's Plot, Cox-Snell Residual Plot, and testing for time-dependent covariates. If the assumption is violated for any variable, a time-dependent covariate will be created and included in the model with the following form:

$$h(t|X) = h_0(t) \cdot e^{\sum_{i=1}^{p} \beta_i x_i + \alpha x_j(t)}$$

where x_j is the variable that violated proportional hazard assumption, and $x_j(t) = x_j \cdot log(t)$. Finally, the fit for final model on the data was assessed using Cox-Snell residual plot.

Similar procedures were carried out to compare the OS between with-bolus vs. no-bolus group, and to analyze covariate effects related to OS of patients in this study.

III. RESULTS

Descriptive Statistics

The descriptive statistics for categorical variables in this dataset are shown in *Table 1*. The dataset is not balanced between treatment groups, but relatively balanced for variables including Gender, Reason for Delay, HTN, and sidedness. Overall, 66.36% of patients died, and 74.55% experienced disease progression during the study period. For the no-bolus group, 72.22% of patients died, and 69.44% experienced disease progression. For the with-bolus group, 63.51% of patients died, and 77.03% experienced disease progression. From Chi-Square Test p-values alone, there is no significant difference between either disease progression status or death status between two treatment groups. In addition, from the calculated p-values, we concluded that Gender, Treatment Delayed, Reason for Delay, Growth Factor Added, and Sidedness are significantly different between no-bolus group and with-bolus group at $\alpha = 0.05$.

The missing values were also recorded for each categorical variable. Most variables did not have more than 10 missing values, but others contained relatively large numbers of

missing values. Those variables include Reason for Delay (74 missing), MSS (44 missing), BRAF (78 missing), KRAS (36 missing), and PNI (68 missing). Careful considerations were taken when examining the significance of these variables in the proceeding statistical tests and models.

For continuous variables, the overall mean, and mean for each treatment group are recorded in *Table 2*. Patients in this study have average age of 57.66 years, but the no-bolus group patients generally have higher age with a mean of 64.33 years compared to with-bolus group with a mean of 54.42 years. The mean OS is about 594.23 days, and mean PFS is about 318.36 days for all patients. The mean OS is 549.78 days for no-bolus group, and 615.85 for with-bolus group. The mean PFS is 257.86 days for no-bolus group, and 347.80 days for with-bolus group. However, two-sample t tests showed no significant difference for OS and PFS between treatment groups. Continuous variables that are significantly different between no-bolus and with-bolus groups are Age, Number of Completed Cycles, and Days of Treatment Delay at $\alpha = 0.05$.

Survival Analysis for PFS

The Kaplan-Meier plot for PFS of all patients in the study is shown in *Figure 1*. Among the 110 patients, 25.45% were censored, and only 4 patients were progression-free after 1000 days from their First Cycle Date. The Kaplan-Meier plot for PFS stratified by treatment group is shown in *Figure 2*. Among the 36 no-bolus patients, 30.56% were censored, and among the 74 with-bolus patients, 22.97% were censored. The PFS Kaplan-Meier curve for with-bolus group is visually above no-bolus group, with a cross-over at

around 1000 days. After 1000 days from First Cycle Date, only 2 patients from each group remained progression free. The p-value associated with log-rank test is $p = 0.1988$, indicating the PFS for two treatment groups are not significantly different.

Univariate Cox Models for PFS were then fitted separately for each variable. The results for categorical variables are shown in *Table 3.*, and the results for continuous variables were shown in *Table 4*. For categorical variables, the p-value for both hazard ratio and log-rank test between two treatment groups are recorded; for continuous variables, only p-values for hazard ratio are recorded. The variables that significantly contribute to PFS univariately include Diabetes status, HTN status, Pani status, and Age at $\alpha = 0.05$. The variables that significantly contribute to PFS univariately include Reason for Delay, and Number of Completed Cycles at $\alpha = 0.1$. From the univariate analysis, patients with diabetes, with HTN, with Pani, of higher age, whose treatments were delayed due to Neutropenia, and who had lower number of completed cycles had higher hazard for disease progression compared to the baseline level.

Variables that are related to PFS in the univariate analysis were included in a multivariate Cox Model to assess the adjusted effects. Even though Group was not significant univariately, it was still included in the model as treatment group is the primary interest of this study for PFS. Hazard ratios and associated p-values from multivariate analysis are shown in *Table 5*. Only the hazard ratio for Diabetes is significantly different from 1 at $\alpha = 0.05$, but hazard ratios for all variables (Group, Diabetes, HTN, and Pani) are

significantly different from 1 at $\alpha = 0.1$. Forward model selection was then implemented on the multivariate model, and the final selected model was:

$$\textbf{Model 1}: h(t|X) = h_0(t) \cdot e^{\beta_1 \text{Group} + \beta_2 \text{Diabetes}}$$

The associated hazard ratios for Group and Diabetes in this model are shown in *Table 6.* The p-value of hazard ratio for Diabetes is significant at $\alpha = 0.05$, and the p-value of hazard ratio for Group is significant at $\alpha = 0.1$. The hazard of disease progression is 1.581 [0.958, 2.612] for no-bolus group relative to the with-bolus group at given *t*, controlling for the diabetes status of patients. The hazard of disease progression is 1.955 [1.067, 3.584] for patients with diabetes relative to patients without diabetes at given *t*, controlling for treatment group.

Proportional hazard assumption for the final model was assessed through Mallows' C(p) plot later) and inclusion of time-dependent covariate. In the appendix section, *Figure 5.* shows the Mallows' plot and *Table 12.* shows the p-values associated with the hazard ratio for Group and Diabetes separately based on the test. Both the plot and p-values do not indicate significance of the time-dependent effect. Therefore, the proportional hazard assumption was not violated for any variables in the model. The fit of final model (Model 1) is assessed via Cox-Snell residual plot shown in *Figure 6.* The plot indicates the cumulative hazard of Cox-Snell residuals approximately follows the 45° line, indicating good fit of the model. The deviation of plot from the line near the tail may be resulted from the relatively small number of patients left near the end of the study, thus larger variability involved in estimating β coefficients.

Survival Analysis for OS

The Kaplan-Meier plot for OS of all patients in the study is shown in *Figure 3*. Among the 110 patients, 33.64% were censored, and only 5 patients were progression-free after 1500 days from their First Cycle Date. The Kaplan-Meier plot for OS stratified by treatment group is shown in *Figure 4*. Among the 36 no-bolus patients, 27.78% were censored, and among the 74 with-bolus patients, 36.49% were censored. The OS Kaplan-Meier curve for with-bolus group is visually above no-bolus group before 1000 days. After 100 days, the two curves approximately overlap. The p-value associated with log-rank test was $p = 0.2856$, indicating OS for two treatment groups are not significantly different.

Univariate Cox Models for OS were then fitted separately for each variable. The results for categorical variables are shown in *Table 7.*, and the results for continuous variables were shown in *Table 8*. For categorical variables, the p-value for both hazard ratio and log-rank test between two treatment groups are recorded; for continuous variables, only p-values for hazard ratio are recorded. At $\alpha = 0.05$, the variables that significantly contribute to OS univariately include Grade 3 ANC, HTN status, Pani, Age, Number of Completed Cycles, and Number of Treatment Delays. At $\alpha = 0.1$, the variables that significantly contribute to OS univariately include Reason for Delay, and Number of Completed Cycles. From the univariate analysis, patients with Grade 3 ANC events, with HTN, with Pani, have higher Age, have less Number of Completed Cycles, and have more Number of Treatment Delays had higher hazard for death compared to the baseline level.

Variables that are related to OS in the univariate analysis were included in a multivariate Cox Model to assess the adjusted effects. Hazard ratios and associated p-values from

multivariate analysis are shown in *Table 9.* The hazard ratio of HTN status and Number of Completed Cycles are significant at $\alpha = 0.05$. The hazard ratio of Number of Treatment Delay is significant at $\alpha = 0.1$. Forward model selection was then implemented on the multivariate model, and the selected model was:

$$\textbf{Model 2}: h(t|X) = h_0(t) \cdot e^{\beta_1 \text{NumDelay} + \beta_2 \text{HTN} + \beta_3 \text{NumCycle} + \beta_4 \text{Grade3ANC}}$$

The associated hazard ratios for covariates in the above model are shown in *Table 10.* The hazard ratio for Number of Treatment Delays, HTN, and Number of Completed Cycles are significant at $\alpha = 0.05$. The hazard ratio for Grade 3 ANC is significant at $\alpha = 0.1$.

Proportional hazard assumption for the model was assessed through Mallow's C(p) plot and inclusion of time-dependent covariate. In the appendix section, *Figure 7.* shows the Mallow's plot, and *Table 13.* shows the p-values associated with the hazard ratio for time-dependent variables of the variables included in the model. The p-value associated with time-dependent variable of Number of Completed Cycles is <.0001, which indicated the proportional hazard assumption for this variable is violated in the model. Therefore, this time-dependent covariate was included in the final model:

$$\textbf{Model 3}: h(t|X) = h_0(t) \cdot e^{\beta 1 \text{NumDelay} + \beta 2 \text{HTN} + \beta 3 \text{NumCycle} + \beta 4 \text{Grade3ANC} + \beta 5 \text{NumCycle.log(OS)}}$$

The p-values and hazard ratios for the final model for OS is shown in *Table 11.*

Additional Results

Table 1. contains the p-value associated with Chi-Square Test comparing variables between two treatment groups. Grade 3 and Grade 4 toxicity events were examined as a secondary objective, and the p-values associated with both variables indicate no significant

difference between no-bolus and with-bolus group. Moreover, KRAS mutation status was also compared between two treatment groups, and no significant results were obtained. Finally, the sidedness of cancer was recorded, and a p-value of $0.031 < 0.05$ was obtained, indicating significant difference of the side of cancer for no-bolus group compared to with-bolus group.

IV. CONCLUSION

Based on univariate analysis, the unadjusted hazard of disease progression is not significantly different between no-bolus group and with-bolus group. However, after adjusting for diabetes status, the effect of using a bolus on PFS is significantly different from 1 under $\alpha = 0.1$, but not under $\alpha = 0.05$. The adjusted hazard ratio of disease progression for no-bolus compared to with-bolus was 1.581 (>1.0). This result indicates relative benefits of keeping the bolus treatment. Therefore, more considerations are needed before omitting the bolus treatment, as the bolus is contributing to the delay of disease progression for metastatic CRC patients.

Based on both univariate analysis and multivariate analysis, the hazard of death is not significantly different between no-bolus and with-bolus group. Therefore, the use of bolus does not seem to affect the overall survival time of patients with metastatic CRC.

V. DISCUSSION

This study aimed to examine the PFS and OS of patients with CRC in order to determine if 5-FU bolus can be omitted from the traditional treatment without detrimental effects on

disease progression or survival. Results from this study indicated that omitting a bolus may not be beneficial on disease progression but has no effect on the overall survival for patients with metastatic CRC. In other studies, the use of bolus was shown to be related to toxicity events [15]. In this study, although the number of Grade 3 ANC and Grade 4 ANC records did not differ by treatment group, there was a significantly higher number of patients in with-bolus group who experienced treatment delay due to side effects. Considering the potential side effects of using bolus, clinical trials can be conducted in the future to achieve solid conclusion of whether it is safe to remove the bolus from standard CRC treatments.

The relatively small sample size (110 patients) limited our ability to achieve less biased results. Interaction terms were not considered in any models, since the number of features should not be similar or even larger than the number of data points. Moreover, using a significance level of $\alpha = 0.05$ may not be optimal for such small sample size, so some variables to include in the model are based on $\alpha = 0.1$ instead. Furthermore, for many variables in the dataset, especially the treatment group is not balanced. There were 74 patients who received the traditional bolus treatment, and only 36 received the treatment with bolus omitted. Imbalanced dataset may cause bias in statistical tests and models. In addition, some variables such as BRAF had more than 50% missing values, and due to the small sample size, methods such as median or predictive mean matching missing imputation did not seem feasible for this dataset. The Cox model is treated as a robust model, for this dataset, the use of parametric models did not seem feasible even under assumption violations. If given future larger datasets, the ‘cross-over’ of survival curves

between two treatment groups may be anticipated and weighted analysis or stratified analysis may be considered.

Overall, this study provides insights and results about the benefits and drawbacks of the use of 5-FU bolus in patients with metastatic CRC. However, further studies should be conducted on larger datasets if possible, to fully examine the effect of bolus on metastatic CRC patients.

VI. REFERENCES

1. Society, A.C., *Colorectal Cancer Facts & Figures 2017-2019.* American Cancer Society, 2017.
2. Rebecca L. Siegel, K.D.M., Stacey A. Fedewa, *Colorectal cancer statistics, 2017.* CA, 2017. **67**(3): p. 177-193.
3. Rebecca L. Siegel, K.D.M., Stacey A. Fedewa, *Cacner Statistics, 2018.* CA, 2018. **68**(1): p. 7-30.
4. Andrew Wolf, E.T.H., *Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society.* CA, 2018. **68**(4): p. 250-281.
5. Wilkes, G.M., *Metastatic Colorectal Cancer: Management Challenges and Opportunities.* CancerNetwork, 2011. **25**(7).
6. Kimberly D. Miller, R.L.S., *Cancer treatment and survivorship statistics, 2016.* CA, 2016. **66**(4): p. 271-289.
7. Lee GH, M.G., *Is right-sided colon cancer different to left-sided colorectal cancer?* PubMed, 2015.
8. Tannaz Armaghany, J.D.W., *Genetic Alterations in Colorectal Cancer.* PMC, 2012.
9. Barras, D., *BRAF Mutation in Colorectal Cancer: An Update.* PMC, 2015.
10. Board, C.N.E., *Colorectal Cancer: Types of Treatment.* Cancer.Net, 2018.
11. Melissa Stoppler, B.K.H., *Generic drugs vs. brand name drugs.* MedicineNet, 2009.
12. Herbert Hurwitz, L.F., William Novotny, *Bevacizumab plus Irinotecan, Fluorouracil, and Leucovorin for Metastatics Colorectal Cancer.* The New England Journal of Medicine, 2004.

13. Astrid Lievre, J.-B.B., Delphine Le Corre, *KRAS Mutation Status Is Predictive of Response to Cetuximab Therapy in Colorectal Cancer.* American Association for Cancer Research, 2006. **66**(8).

14. Marc Buyse, T.B., Kevin Carroll, Stefan Michiels, Daniel J. Sargent, *Progression-Free Survival Is a Surrogate for Survival in Advanced Colorectal Cancer.* Journal of Clinical Oncology, 2007. **25**.

15. A de Gramont, J.F.B., *Randomized trial comparing monthly low-dose leucovorin and fluorouracil bolus with bimonthly high-dose leucovorin and fluorouracil bolus plus continuous infusion for advanced colorectal cancer: a French intergroup study.* Journal of Clinical Oncology, 1997. **15**.

16. Melissa M Thrall, P.W., *Investigation of the comparative toxicity of 5-FU bolus versus 5-FU continuous infusion circadian chemotherapy with concurrent radiation therapy in locally advanced rectal cancer.* International Journal of Radiation Oncology, 2000. **46**(873-881).

17. Mahaseth H, B.E., Kauh J, *Modified FOLFIRINOX regimen with improved safety and maintained efficacy in pancreatic adenocarcinoma.* PubMed, 2013. **42**(8).

VII. TABLES AND FIGURES

Table 1: Summary of categorical variables for all subjects, and within each treatment group.

Variable	Levels	Overall (col %) N = 110	No Bolus (col %) N = 36	With Bolus (col %) N = 74	P-value*
Gender†	Male	56 (51.38)	12 (34.29)	44 (59.46)	0.014
	Female	53 (48.62)	23 (65.71)	30 (40.54)	
Race	White	61 (55.45)	15 (41.67)	46 (62.16)	0.114
	Black	41 (37.27)	17 (47.22)	24 (32.43)	
	Other	8 (7.27)	4 (11.11)	4 (5.41)	
Death	Yes	73 (66.36)	26 (72.22)	47 (63.51)	0.364
	No	37 (33.64)	10 (27.78)	27 (36.49)	
Progression	Yes	82 (74.55)	25 (69.44)	57 (77.03)	0.392
	No	28 (25.45)	11 (30.56)	17 (22.97)	
Grade 3 ANC	Yes	10 (9.09)	4 (11.11)	6 (8.11)	0.726
	No	100 (90.91)	32 (88.89)	68 (91.89)	
Grade 4 ANC	Yes	3 (2.73)	1 (2.78)	2 (2.70)	1.000
	No	107 (97.27)	35 (97.22)	72 (97.30)	
Treatment Delayed†	Yes	30 (28.04)	2 (6.06)	28 (37.84)	0.001
	No	77 (71.96)	31 (93.94)	46 (62.16)	
	Missing	3	3	0	
Reason for Delay†	Thrombocytopenia	7 (19.44)	0 (0.00)	7 (25.00)	0.007
	Neutropenia	11 (30.56)	5 (75.00)	5 (17.86)	
	Other	18 (50.00)	2 (25.00)	16 (57.14)	
	Missing	74	28	46	
Diabetes	Yes	18 (16.36)	6 (16.67)	12 (16.22)	0.952
	No	92 (83.64)	30 (83.33)	62 (83.78)	

Variable	Levels	Overall (col %) N = 110	No Bolus (col %) N = 36	With Bolus (col %) N = 74	P-value*
Growth Factor Added†	Yes	29	2 (6.25)	27 (36.49)	0.001
	No	77	30 (93.75)	47 (63.51)	
	Missing	4	4	0	
CHF	Yes	3 (2.73)	1 (2.78)	2 (2.70)	1.000
	No	107 (97.27)	35 (97.22)	72 (97.30)	
CKD	Yes	2 (1.82)	1 (2.78)	1 (1.35)	0.550
	No	108 (98.18)	35 (97.22)	73 (98.65)	
HTN	Yes	48 (43.64)	20 (55.56)	28 (37.84)	0.079
	No	62 (56.36)	16 (44.44)	46 (62.16)	
MSS	Yes	5 (7.58)	2 (9.09)	3 (6.82)	1.000
	No	61 (92.42)	20 (90.91)	41 (93.18)	
	Missing	44	14	30	
BRAF	Yes	7 (21.88)	5 (35.71)	2 (11.11)	0.195
	No	25 (78.13)	9 (64.29)	16 (88.89)	
	Missing	78	22	56	
KRAS	Yes	45 (60.81)	17 (65.38)	28 (58.33)	0.553
	No	29 (39.19)	9 (34.62)	20 (41.67)	
	Missing	36	10	26	
MAb	Yes	78 (70.91)	25 (69.44)	53 (71.62)	0.814
	No	32 (29.09)	11 (30.56)	21 (28.38)	
Bev	Yes	69 (62.73)	24 (66.67)	45 (60.81)	0.551
	No	41 (37.27)	12 (33.33)	29 (39.19)	
Cetux	Yes	3 (2.73)	0 (0.00)	3 (4.05)	0.550
	No	107 (97.27)	36 (100.00)	71 (95.95)	

Variable	Levels	Overall (col %) N = 110	No Bolus (col %) N = 36	With Bolus (col %) N = 74	P-value*
Pani	Yes	4 (3.64)	1 (2.78)	3 (4.05)	1.000
	No	106 (96.36)	35 (97.22)	71 (95.95)	
Sidedness†	Right	45 (48.91)	19 (65.62)	26 (41.27)	0.031
	Left	47 (51.09)	10 (34.48)	37 (58.73)	
	Missing	18	7	11	
ECOG	0	17 (25.37)	3 (10.71)	14 (35.90)	0.062
	1	35 (52.24)	18 (64.29)	17 (43.59)	
	2	15 (22.39)	7 (25.00)	8 (20.51)	
PNI	Yes	14 (33.33)	2 (16.67)	12 (40.00)	0.147
	No	28 (66.67)	10 (83.33)	18 (60.00)	
		68	24	44	

* p-values are calculated by Chi-Square Test, Fisher's Exact Test (for cell counts < 5), or ANOVA (for variables with >2 levels) comparing treatment group counts.

† The highlighted variables are significantly different between two treatment groups at $\alpha = 0.05$.

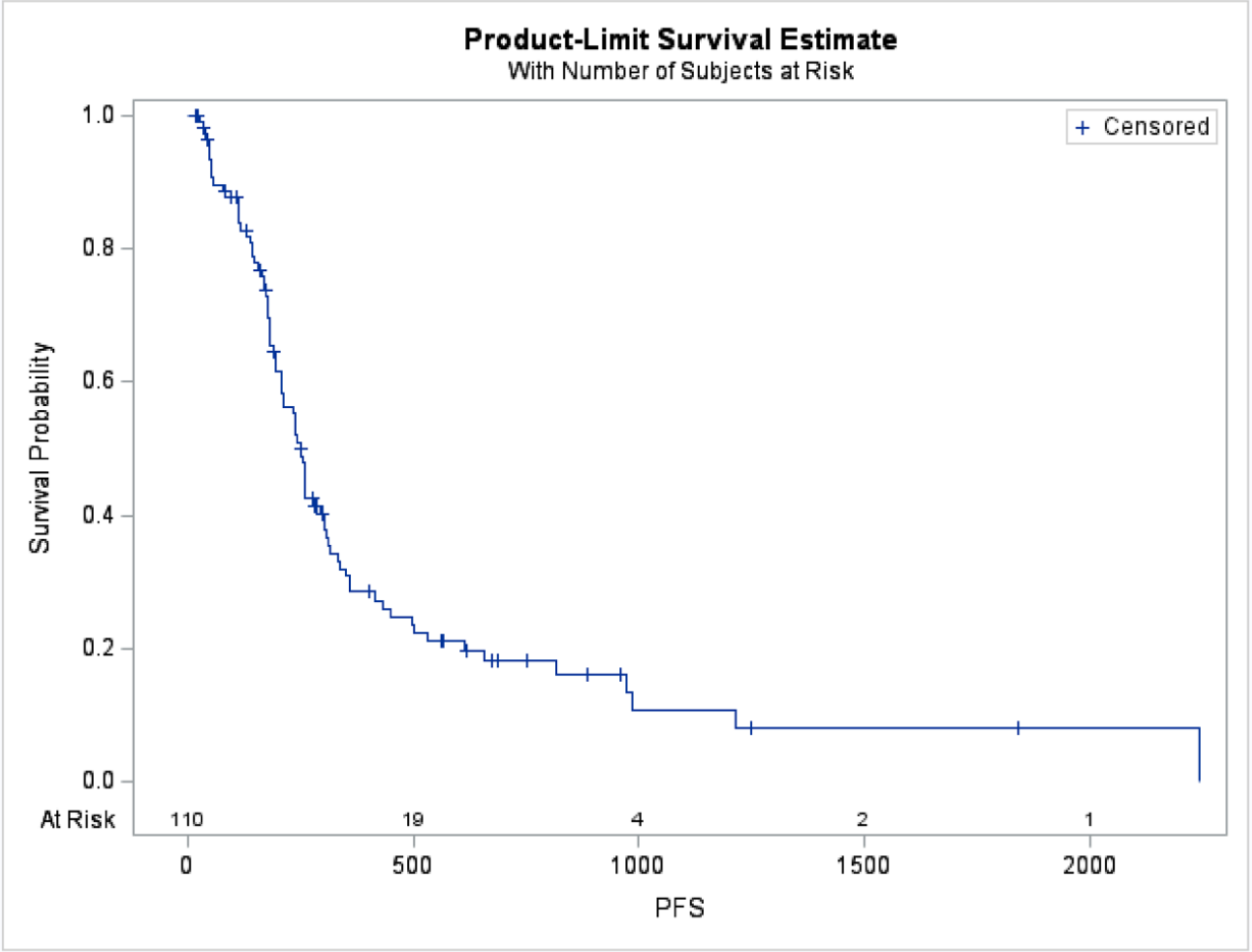
Table 2: Mean of continuous variables for all subjects, and within each treatment group.

Variable	Overall (SD) N = 110	No Bolus (SD) N = 36	With Bolus (SD) N = 74	P-value*
Age†	57.66 (13.88)	64.33 (13.11)	54.42 (13.13)	<.0001
Number of Completed Cycles†	4.30 (2.17)	3.69 (2.04)	4.59 (2.19)	0.037
Number of Treatment Delays	0.97 (3.15)	2.11 (5.31)	0.42 (0.60)	0.065
Days of Treatment Delay†	5.39 (11.62)	0.89 (3.99)	7.58 (13.39)	<.0001
Overall Survival (OS)	594.23 (453.63)	549.78 (574.71)	615.85 (384.01)	0.535
Progression-free Survival (PFS)	318.36 (344.62)	257.86 (396.18)	347.80 (315.27)	0.239

* p-values are calculated by two-sample Welch t tests comparing treatment group means.

† The highlighted variables are significantly different between two treatment groups at $\alpha = 0.05$.

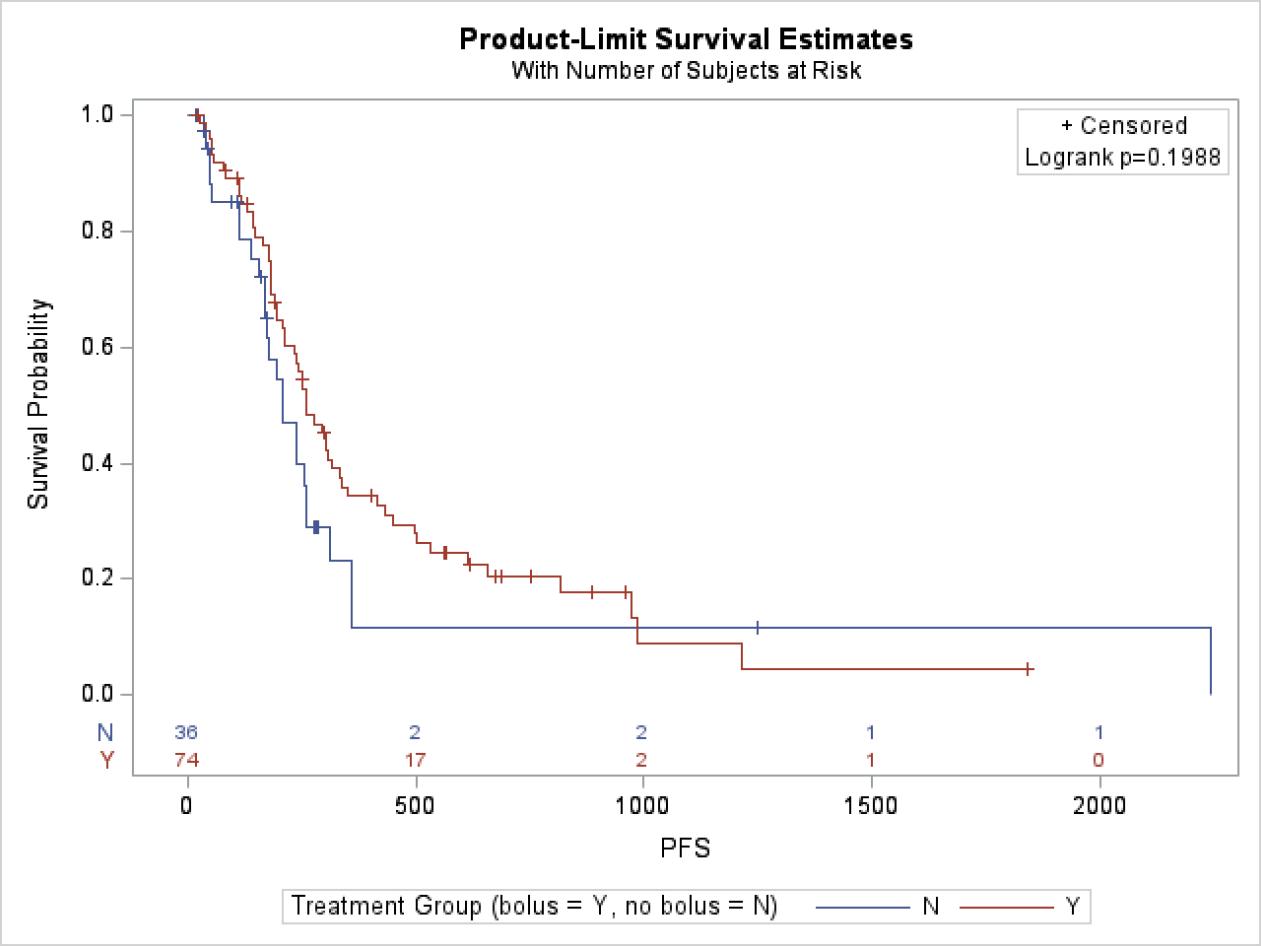
Figure 1: Kaplan-Meier plot of all patients' Progression-free Survival (PFS).



Summary of Progression-free Survival (PFS) pattern for patients overall

Total	Failed	Censored	Percent Censored
110	82	28	25.45

Figure 2: Kaplan-Meier plot of patients' Progression-free Survival (PFS) by treatment group.



Summary of Progression-free Survival (PFS) pattern for patients by treatment group

Group	Total	Failed	Censored	Percent Censored
No Bolus	36	25	11	30.56
With Bolus	74	57	17	22.97

Table 3: Univariate analysis for Progression-free Survival (PFS) on categorical variables.

Variable	Levels	Hazard Ratio (95% CI)	Hazard Ratio p-value*	Log-Rank p-value$^{\Delta}$
Group	No Bolus	1.372 (0.884, 2.229)	0.202	0.199
	With Bolus	Ref		
Gender	Male	1.046 (0.676, 1.620)	0.840	0.839
	Female	Ref		
Race	White	1.125 (0.473, 2.674)	0.790	0.947
	Black	1.160 (0.478, 2.815)	0.743	
	Other	Ref		
Grade 3 ANC	Yes	1.277 (0.583, 2.793)	0.540	0.538
	No	Ref		
Grade 4 ANC	Yes	1.119 (0.352, 3.559)	0.849	0.849
	No	Ref		
Treatment Delayed	Yes	1.025 (0.624, 1.681)	0.924	0.924
	No	Ref		
Reason for Delay^{+}	Thrombocytopenia	0.402 (0.126, 1.276)	0.122	0.053
	Neutropenia	1.531 (0.659, 3.554)	0.322	
	Other	Ref		
Diabetes†	Yes	2.188 (1.300, 3.817)	0.006	0.004
	No	Ref		
Growth Factor Added	Yes	0.760 (0.444, 1.302)	0.318	0.316
	No	Ref		
CHF	Yes	1.171 (0.368, 3.731)	0.789	0.789
	No	Ref		
CKD	Yes	0.819 (0.000, 5.917)	0.843	0.842
	No	Ref		

Variable	Levels	Hazard Ratio (95% CI)	Hazard Ratio p-value*	Log-Rank p-value$^{\Delta}$
HTN†	Yes	1.677 (1.073, 2.618)	0.023	0.022
	No	Ref		
MSS	Yes	1.552 (0.475, 5.076)	0.467	0.461
	No	Ref		
BRAF	Yes	2.229 (0.760, 6.536)	0.144	0.134
	No	Ref		
KRAS	Yes	1.415 (0.829, 2.416)	0.204	0.200
	No	Ref		
MAb	Yes	1.380 (0.816, 2.336)	0.230	0.226
	No	Ref		
Bev	Yes	1.208 (0.757, 1.931)	0.428	0.426
	No	Ref		
Cetux	Yes	1.146 (0.360, 3.650)	0.818	0.817
	No	Ref		
Pani†	Yes	2.935 (1.052, 8.197)	0.040	0.031
	No	Ref		
Sidedness	Right	1.155 (0.718, 1.859)	0.552	0.550
	Left	Ref		
ECOG	0	0.734 (0.302, 1.785)	0.464	0.446
	1	1.122 (0.510, 2.468)	0.082	
	2	Ref		
PNI	Yes	1.318 (0.615, 2.825)	0.478	0.476
	No	Ref		

* p-values associated with the hazard ratio estimate comparing each level with reference level for each variable.

Δ p-values associated with the log-rank test comparing the PFS between each level for categorical variables.

† The survival function for PFS of highlighted variables is significantly different at each level compared to the reference level according to the log-rank test at $\alpha = 0.05$.

+ The survival function for PFS of highlighted variables is significantly different at each level compared to the reference level according to the log-rank test at $\alpha = 0.1$.

Table 4: Univariate analysis for Progression-free Survival (PFS) on continuous variables.

Variable	Hazard Ratio (95% CI)	Hazard Ratio p-value*
Age†	1.017 (1.000, 1.033)	0.045
Number of Completed Cycles+	0.899 (0.799, 1.011)	0.075
Number of Treatment Delays	1.023 (0.954, 1.096)	0.532
Days of Treatment Delay	0.846 (0.983, 1.022)	0.846

* p-values associated with the hazard ratio estimate for one-unit increase in the variable.

† The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.05$.

+ The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.1$.

Table 5: Multivariate analysis for Progression-free Survival (PFS).

Variable	Hazard Ratio (95% CI)	Hazard Ratio p-value*
Group+	1.580 (0.954, 2.615)	0.076
Diabetes†	1.955 (1.067, 3.584)	0.030
HTN+	1.495 (0.935, 2.392)	0.093
Pani+	2.669 (0.935, 7.634)	0.066

The variables that are associated with PFS with p-value < 0.1 from univariate analysis were selected for multivariate analysis.

* p-values associated with the hazard ratio estimate for one-unit increase in the variable.

† The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.05$.

+ The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.1$.

The reference levels are: Group = with-bolus, Diabetes = No, HTN = No, Pani = No.

Table 6: Hazard ratio for variables included in final model for Progression-free Survival (PFS).

Variable	**Hazard Ratio (95% CI)**	**Hazard Ratio p-value***
Group+	1.581 (0.958, 2.612)	0.073
Diabetes†	2.452 (1.381, 4.348)	0.002

* p-values associated with the hazard ratio estimate for one-unit increase in the variable.

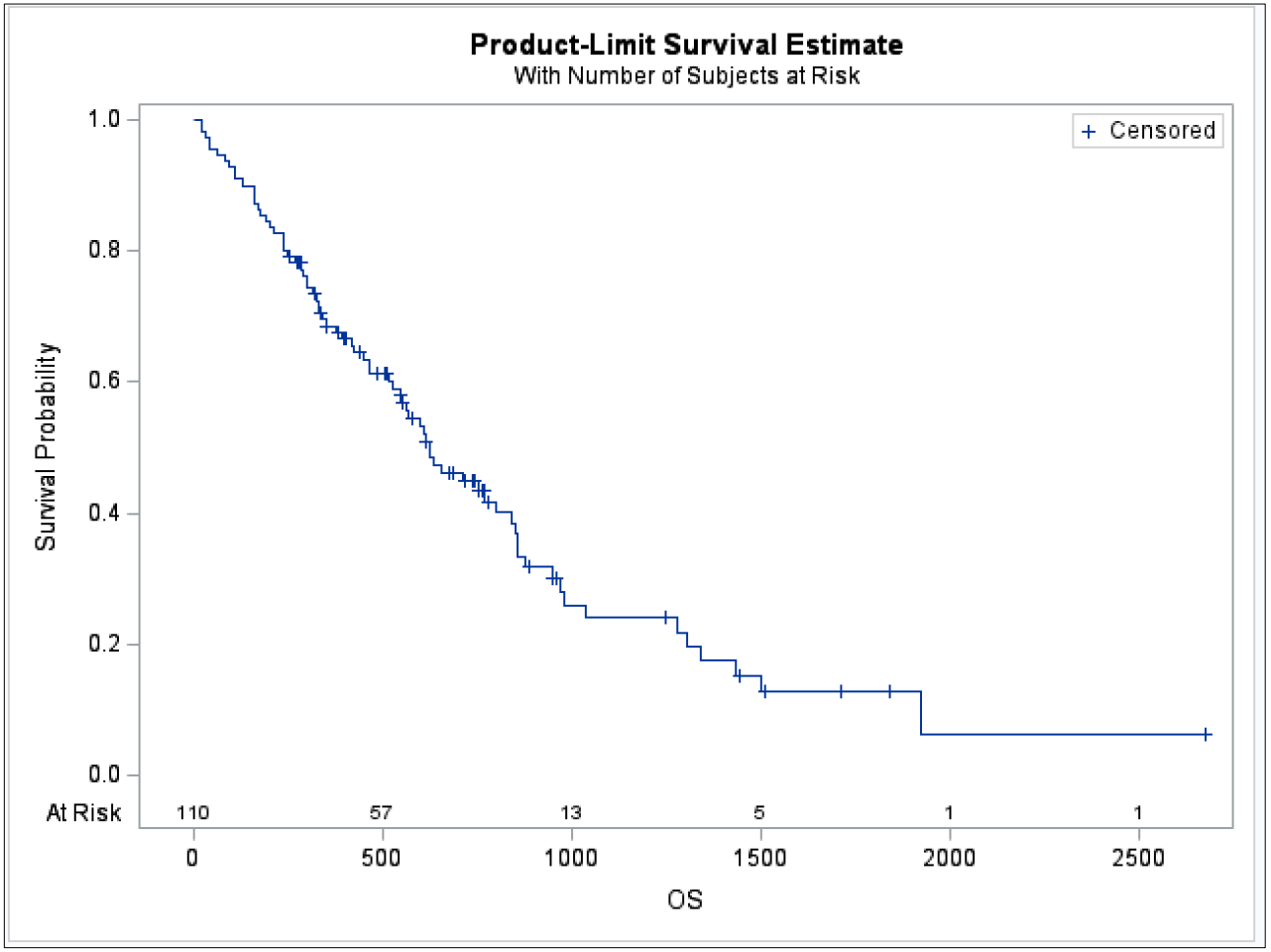
† The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.05$.

+ The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.1$.

The variables included in the final model were selected by forward selection based on smallest AIC.

The reference levels are: Group = with-bolus, Diabetes = No.

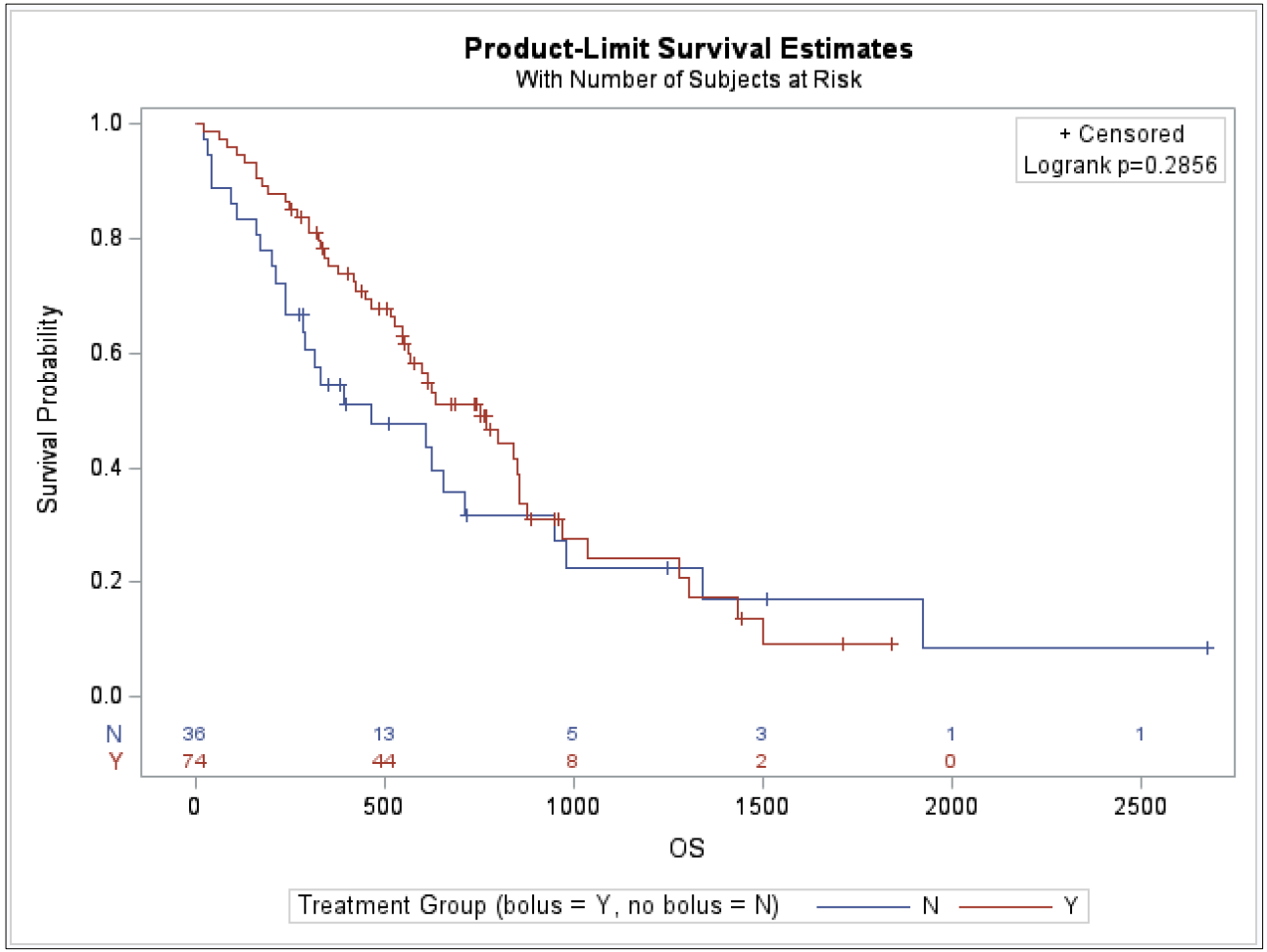
Figure 3: Kaplan-Meier plot of all patients' Overall Survival (OS).



Summary of Overall Survival (OS) pattern for patients overall

Total	Failed	Censored	Percent Censored
110	73	37	33.64

Figure 4: Kaplan-Meier plot of patients' Overall Survival (OS) by treatment group.



Summary of Overall Survival (OS) pattern for patients by treatment group

Group	Total	Failed	Censored	Percent Censored
No Bolus	36	26	10	27.78
With Bolus	74	47	27	36.49

Table 7: Univariate analysis for Overall Survival (OS) on categorical variables:

Variable	Levels	Hazard Ratio (95% CI)	Hazard Ratio p-value*	Log-Rank p-value$^{\Delta}$
Group	No Bolus	1.304 (0.800, 2.127)	0.287	0.286
	With Bolus	Ref		
Gender	Male	0.681 (0.427, 1.084)	0.105	0.103
	Female	Ref		
Race	White	1.038 (0.432, 2.494)	0.934	0.335
	Black	1.469 (0.600, 3.594)	0.400	
	Other	Ref		
Grade 3 ANC†	Yes	2.468 (1.107, 5.495)	0.027	0.022
	No	Ref		
Grade 4 ANC	Yes	0.962 (0.235, 3.937)	0.957	0.956
	No	Ref		
Treatment Delayed	Yes	1.528 (0.917, 2.545)	0.104	0.101
	No	Ref		
Reason for Delay^{+}	Thrombocytopenia	0.263 (0.075, 0.914)	0.036	0.080
	Neutropenia	0.694 (0.292, 1.653)	0.410	
	Other	Ref		
Diabetes	Yes	1.592 (0.857, 2.959)	0.141	0.138
	No	Ref		
Growth Factor Added	Yes	1.311 (0.760, 2.262)	0.330	0.329
	No	Ref		
CHF	Yes	1.686 (0.527, 5.406)	0.379	0.373
	No	Ref		
CKD	Yes	1.087 (0.150, 7.874)	0.934	0.934
	No	Ref		

Variable	Levels	Hazard Ratio (95% CI)	Hazard Ratio p-value*	Log-Rank p-value$^{\Delta}$
HTN†	Yes	1.970 (1.239, 3.135)	0.004	0.004
	No	Ref		
MSS	Yes	1.166 (0.276, 1.166)	0.835	0.834
	No	Ref		
BRAF	Yes	1.400 (0.377, 5.208)	0.615	0.614
	No	Ref		
KRAS	Yes	0.961 (0.537, 1.721)	0.894	0.894
	No	Ref		
MAb	Yes	1.055 (0.633, 1.761)	0.837	0.837
	No	Ref		
Bev	Yes	0.990 (0.613, 1.597)	0.968	0.968
	No	Ref		
Cetux	Yes	- (0.000, -)	0.987	0.184
	No	Ref		
Pani†	Yes	3.005 (1.085, 8.333)	0.034	0.026
	No	Ref		
Sidedness	Right	0.963 (0.618, 1.75)	0.887	0.887
	Left	Ref		
ECOG	0	0.412 (0.173, 0.984)	0.046	0.124
	1	0.720 (0.364, 1.427)	0.347	
	2	Ref		
PNI+	Yes	2.397 (0.864, 6.623)	0.093	0.084
	No	Ref		

* p-values associated with the hazard ratio estimate comparing each level with reference level for each variable.

Δ p-values associated with the log-rank test comparing the OS between each level for categorical variables.

† The survival function for OS of highlighted variables is significantly different at each level compared to the reference level according to the log-rank test at $\alpha = 0.05$.

+ The survival function for OS of highlighted variables is significantly different at each level compared to the reference level according to the log-rank test at $\alpha = 0.1$.

Table 8: Univariate analysis for Overall Survival (OS) on continuous variables:

Variable	Hazard Ratio (95% CI)	Hazard Ratio p-value*
Age†	1.024 (1.006, 1.043)	0.011
Number of Completed Cycles†	0.876 (0.772, 0.993)	0.038
Number of Treatment Delays†	1.087 (1.024, 1.154)	0.006
Days of Treatment Delay+	1.017 (0.999, 1.037)	0.068

* p-values associated with the hazard ratio estimate for one-unit increase in the variable.

† The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.05$.

+ The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.1$.

Table 9: Multivariate analysis for Overall Survival (OS).

Variable	Hazard Ratio (95% CI)	Hazard Ratio p-value*
Group	0.938 (0.536, 1.643)	0.824
Grade 3 ANC	1.961 (0.789, 4.878)	0.147
HTN†	1.975 (1.164, 3.356)	0.012
Pani	1.933 (0.602, 6.211)	0.268
Age	1.007 (0.972, 1.014)	0.526
Number of Completed Cycles†	0.856 (1.028, 1.326)	0.017
Number of Treatment Delay+	1.051 (0.880, 1.010)	0.096

The variables that are associated with OS with p-value < 0.05 from univariate analysis were selected for multivariate analysis.

* p-values associated with the hazard ratio estimate for one-unit increase in the variable.

† The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.05$.

+ The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.1$.

The reference levels are: Group = with-bolus, Grade 3 ANC = No, HTN = No, Pani = No.

Table 10: Hazard ratio for variables included in Model 2 for Overall Survival (OS).

Variable	Hazard Ratio (95% CI)	Hazard Ratio p-value*
Number of Treatment Delays†	1.072 (1.010, 1.139)	0.023
HTN†	2.026 (1.272, 3.226)	0.003
Number of Completed Cycles†	0.845 (1.045, 1.339)	0.008
Grade 3 ANC+	2.230 (0.947, 5.236)	0.066

* p-values associated with the hazard ratio estimate for one-unit increase in the variable.
† The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.05$.
+ The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.1$.
The variables included in the final model were selected by forward selection based on smallest AIC.
The reference levels are: Group = with-bolus, HTN = No, Grade 3 ANC = No.

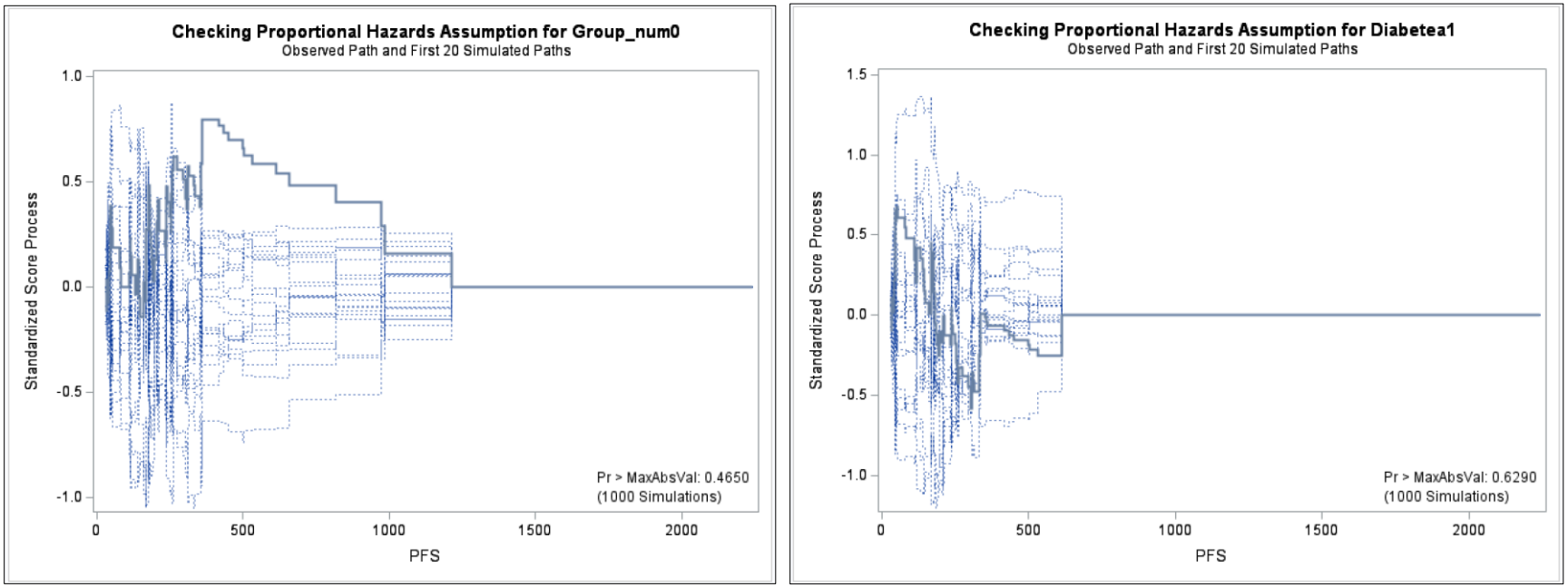
Table 11: Hazard ratio for variables included in final model for Overall Survival (OS) including time-dependent covariate.

Variable	Hazard Ratio (95% CI)	Hazard Ratio p-value*
Number of Treatment Delays†	1.068 (1.006, 1.134)	0.031
HTN†	1.780 (1.106, 2.865)	0.017
Number of Completed Cycles†	0.075 (0.025, 0.230)	<.0001
Grade 3 ANC+	2.869 (1.258, 6.536)	0.012
Number of Completed Cycles · log(OS)†	1.494 (1.249, 1.788)	<.0001

* p-values associated with the hazard ratio estimate for one-unit increase in the variable.
† The hazard ratio associated with one-unit increase in the highlighted variables is significantly different from 1 at $\alpha = 0.05$.
The variables included in the final model were selected by forward selection based on smallest AIC.
The reference levels are: Group = with-bolus, HTN = No, Grade 3 ANC = No.

VIII. APPENDIX

Figure 5: Mallow's C(p) plot to assess proportional hazard assumption for variables in Model 1 (PFS).



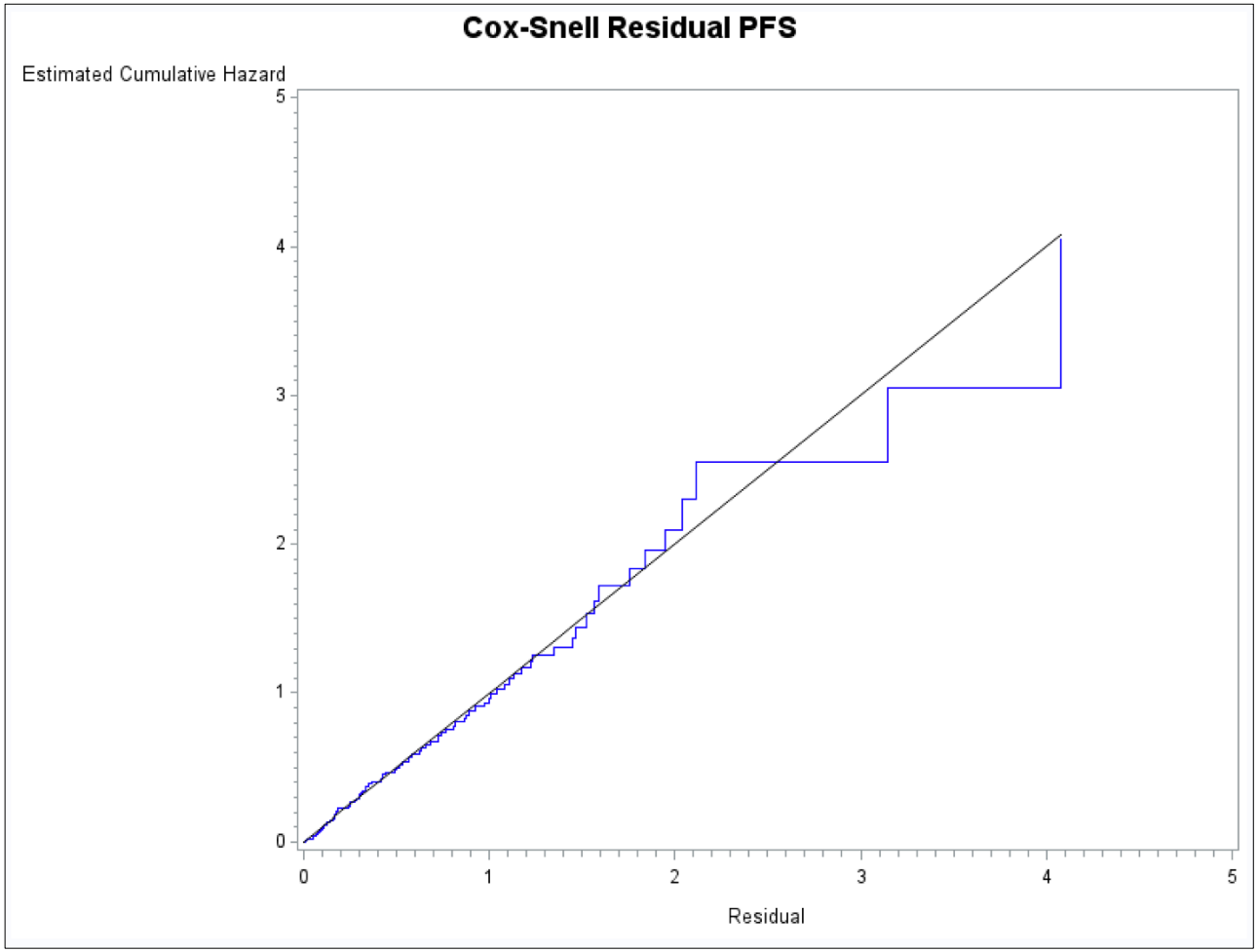
The curves for both Group and Diabetes variables fluctuate around 0 within – 1.0 to 1.0 range. Therefore, proportional hazard assumption holds for both variables.

Table 12: Check for time-dependent covariate in Model 1.

Variable	Hazard Ratio (95% CI)	Hazard Ratio p-value*
Group · log(PFS)	1.458 (0.789, 2.692)	0.229
Diabetes · log(PFS)	0.775 (0.355, 1.691)	0.522

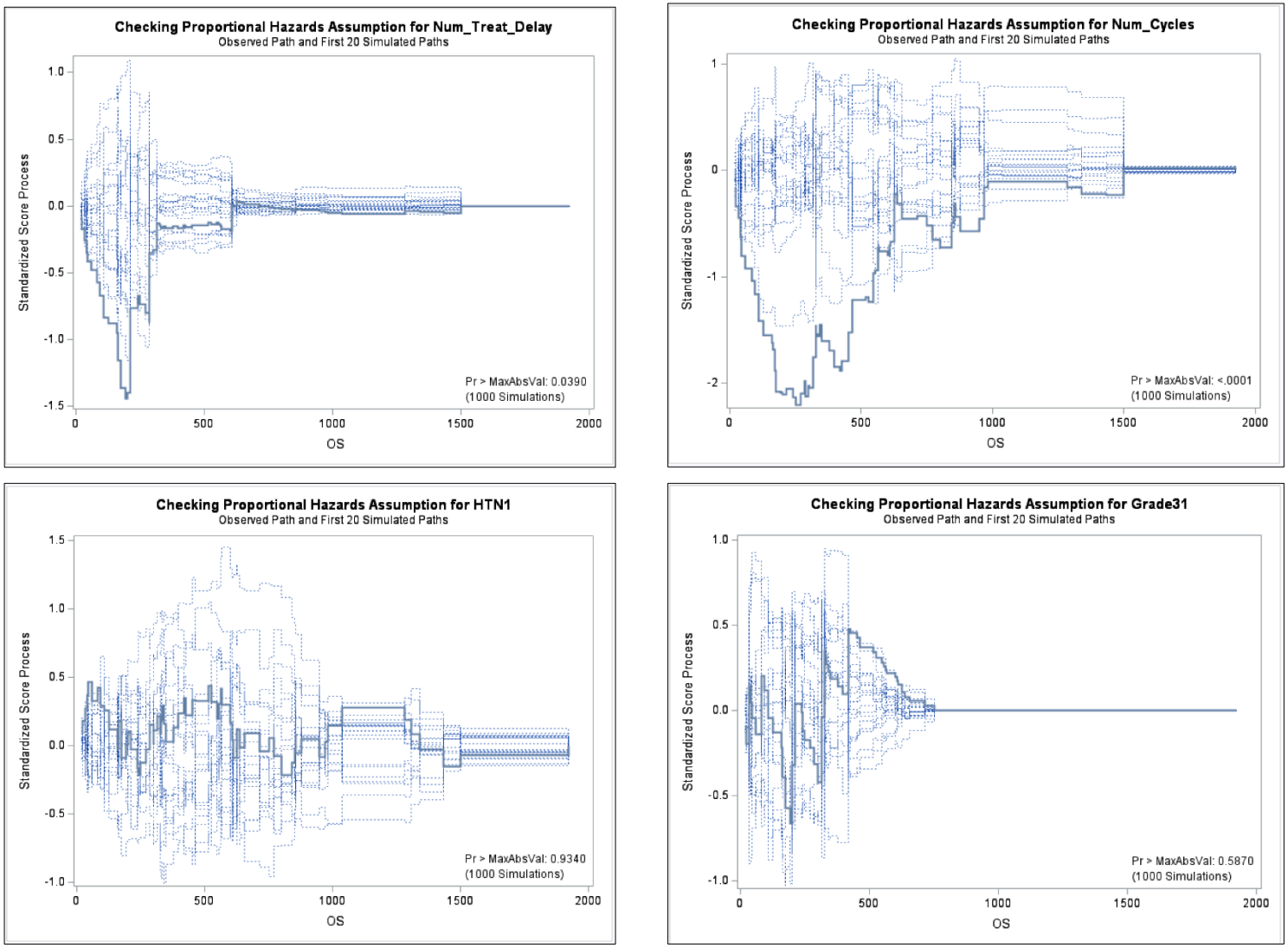
* p-values associated with the hazard ratio estimate for one-unit increase in the variable.
Each variable was fitted in the Cox Model with the original variable and the time-dependent variable. The significance level of time-dependent variable is assessed, and the proportional hazard assumption is violated if p-value for time-dependent variable is less than 0.05. In this model, no variable violated the assumption.

Figure 6: Cox-Snell Residual Plot to assess fit of final model for PFS.



The Cumulative hazard for Cox-Snell Residuals approximately follows the 45° line, indicating relatively good fit of the model to the data. The deviation from the line near tail of the plot could be due to the small number of patients left near end of the study, thus larger variability involved in estimating β coefficients.

Figure 7: Mallow's C(p) plot to assess proportional hazard assumption for variables in Model 2 (OS).



The curves for all variables except Number of Completed Cycles fluctuate around 0 within – 1.0 to 1.0 range. Therefore, the proportional hazard assumption is violated

Table 13: Check for time-dependent covariate in Model 2.

Variable	**Hazard Ratio (95% CI)**	**Hazard Ratio p-value***
Number of Treatment Delays · log(OS)	1.152 (0.998, 1.331)	0.054
HTN · log(OS)	0.801 (0.491, 1.307)	0.375
Number of Completed Cycles · log(OS)†	1.513 (1.266, 1.807)	<.0001
Grade 3 ANC · log(OS)	0.939 (0.396, 2.225)	0.886

* p-values associated with the hazard ratio estimate for one-unit increase in the variable.

† The proportional hazard assumption is violated for the highlighted variable at $\alpha = 0.05$.

Each variable was fitted in the Cox Model with the original variable and the time-dependent variable. The significance level of time-dependent variable is assessed, and the proportional hazard assumption is violated if p-value for time-dependent variable is less than 0.05. In this model, no variable violated the assumption.