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:

_____ _____

Ruizhe Wu Date

Risk Stratification by Nomogram for Stage II Colon Cancer Patients and Impacts from Adjuvant Chemotherapy on Overall Survival Based on NCDB

By

Ruizhe Wu

Master of Public Health

Biostatistics and Bioinformatics

_____

Yuan Liu, PhD

(Thesis Advisor)

_____

Paul Weiss, MS

(Reader)

Risk Stratification by Nomogram for Stage II Colon Cancer Patients and Impacts from Adjuvant Chemotherapy on Overall Survival Based on NCDB

By

Ruizhe Wu

B.S.

University of Washington

2016

Thesis Committee Chair: Yuan Liu, PhD

Reader: Paul Weiss, MS

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 Public Health

in Biostatistics and Bioinformatics

2018

Abstract

Risk Stratification by Nomogram for Stage II Colon Cancer Patients and Impacts from Adjuvant Chemotherapy on Overall Survival Based on NCDB

By Ruizhe Wu

The benefits of adjuvant chemotherapy on overall survival (OS) for patients with stage II colon cancer is controversial. The major question is whether these patients should take adjuvant chemotherapy or not. This study aims to establish a risk stratification for stage II colon cancer patients under surgical resection and evaluate benefits of adjuvant chemotherapy on OS for these patients. Based on National Cancer Data Base (NCDB), we tried to build up two risk prediction models that aim to classify patients into three risk groups. Through conducting single-variable and multivariable survival analysis stratified by the risk groups, we aimed to examine whether the impact of adjuvant chemotherapy on OS would be different among the three risk groups for the target stage II colon cancer patients.

Risk Stratification by Nomogram for Stage II Colon Cancer Patients and Impacts from Adjuvant Chemotherapy on Overall Survival Based on NCDB

By

Ruizhe Wu

B.S.

University of Washington

2016

Thesis Committee Chair: Yuan Liu, PhD

Reader: Paul Weiss, MS

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 Public Health

in Biostatistics and Bioinformatics

2018

Table of Contents

Risk Stratification by Nomogram for Stage II Colon Cancer Patients and Impacts from Adjuvant Chemotherapy on Overall Survival Based on NCDB

Introduction

Colon cancer, the second leading cause of cancer death in the United States, is the development of cancer in the colon or rectum[1]. It is also the third most common cancer in both men and women. According to a report from National Cancer Institute, the risk of getting colon cancer increases with age and people aged 75-84 have the highest colon cancer death rate. The number of new cases of colon cancer was 40.1 per 100,000 men and women per year and the death rate was 1.48 per 100,000 for men and women per year[2]. For a diagnosed patient, we use cancer staging to describe the severity of the patient's cancer based on the magnitude of the tumor and the extent that the cancer has developed by growing and spreading in the body. Overall Stage Grouping uses numerals I, II, III, and IV (plus the 0). Stage 0 colon cancer means abnormal cells are present but the cells have not grown beyond the inner lining of the colon and they may become cancer. Stage I, stage II, and stage III colon cancer mean cancers are localized in the layers of the colon wall. The higher the stage, the more the tumor has spread into nearby tissues and the larger the cancer tumor. Stage IV colon cancer means the cancer has spread to other parts of the body[4].

Since stage 0 colon cancers have not spread to nearby tissue, the only treatment needed is surgery to take out the abnormal cells. In stage I colon cancer, the widely accepted

standard treatment is surgical resection alone and the standard treatment in stage III colon cancer is adjuvant chemotherapy (chemotherapy after surgery)[5]. However, the standard treatment for patients with stage II colon cancer remains a controversial area. Surgical resection of the colon containing the cancer along nearby tissues may be the only treatment needed, but some doctors may recommend chemotherapy after surgery if the patient has a higher risk of recurring. Nevertheless, considering the side effect of chemotherapy and unclear benefits of chemotherapy to overall survival (OS), not all doctors agree on such treatment[6].

Trials including International Multicenter Pooled Analysis of B2 Colon Cancer Trials (IMPACT B2 Analysis), Intergroup Analysis, Cancer Care Ontario Program, The Leucovorin and Fluorouracil Compared with Observation in Treating Patients with Colorectal Cancer that has been Surgically Removed (QUASAR), Multicenter International Study of Oxaliplatin/5-Fluorouracil/Leucovorin in the Adjuvant Treatment of Colon Cancer (MOSAIC), National Surgical Breast and Bowel Project (NSABP), and Adjuvant Colon Cancer Endpoints (ACCENT), evaluated the benefits of chemotherapy after surgery to OS. Some of them did not acquire supportive results and some of them include patients with both stage II and stage III colon cancers so that evaluating the true benefits of chemotherapy for patients with stage II cancer becomes difficult[7]. In general, benefits of chemotherapy after surgery to OS for patients with stage II colon cancer are still being questioned. In this case, we considered it possible that benefits for patients with stage II colon cancer are different among different risk groups and that is the reason why accessing benefits for stage II patients are difficult. So far, there are some

researchers defining characteristics of a high-risk stage II colon cancer patient but there are no clear guidelines to build a stage II colon cancer patient's specific risk score to create a comprehensive patient profile. Thus, in this article, I will establish a risk stratification for stage II colon cancer patients under surgical resection and evaluate benefits of adjuvant chemotherapy on OS for stage II colon cancer patients.

Method

Patient Selection

The data come from the nationally recognized National Cancer Data Base (NCDB), a joint program of Commission on Cancer (CoC) of the American College of Surgeons and the American Cancer Society, a nationwide oncology outcomes database from more than 1,500 Commission-accredited cancer programs in the United States and Puerto Rico[8]. Patients who satisfy the following conditions are included: 1) patients had surgery; 2) patients with stage 2 colon cancer; 3) patients with invasive tumors. In other words, following patients are excluded: 1) patients with radiation; 2) patients had chemotherapy before surgery; 3) patients whose colon cancer was in another stage besides stage 2; 4) patients whose AJCC Pathologic N was in (1, 1A, 1B, 1C, 2, 2A, 2B); 5) patients whose Regional Lymph Nodes were positive; 6) patients with distant metastasis. Since we focused on exploring the effect of chemotherapy, patients whose chemotherapy situation was unknown or neoadjuvant chemotherapy were excluded. Patients who have missing values in main predictors (Lymph-vascular invasion, Surgical Margins Status at any CoC Facility, CS SSF8 - Perineural Invasion, AJCC Pathologic T) were excluded[9].

For survival analysis, based on the population mentioned above, additional exclusions were applied: 1) patients whose current cancer diagnosis was not the first one or the only one on the sequence; 2) patients who were diagnosed in 2014 due to unavailable vital status.

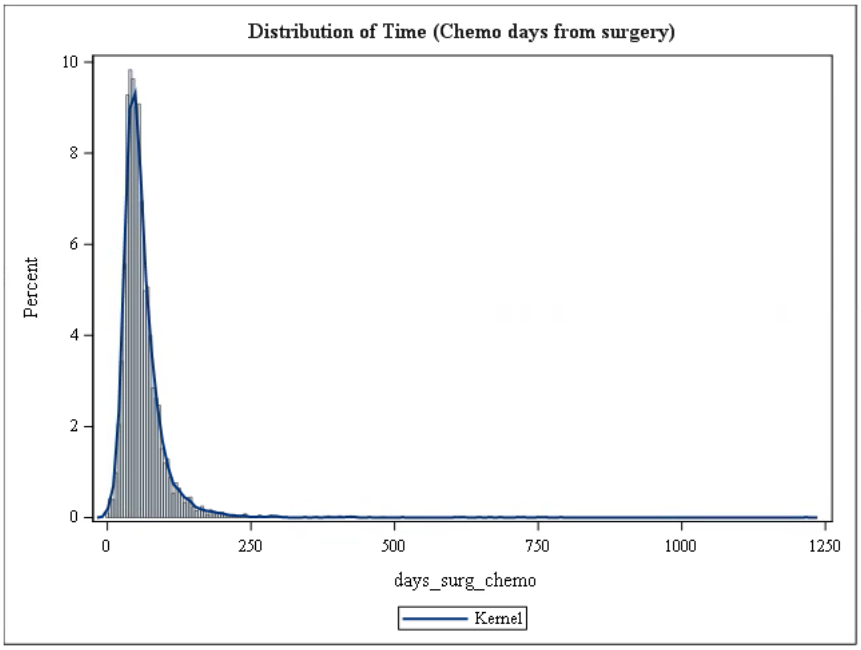


Figure 1. Distribution of days between Chemotherapy and surgery

Before defining outcome OS, we need to reduce guarantee time bias (GTB) first. GTB occurs when an analysis that is timed from enrollment or random assignment, comparing across groups defined by a classifying event. GTB does not always arise but occurs sometime during follow-up[10]. In this case, GTB means that patients in the chemotherapy group were guaranteed to have survived at least until an adjuvant chemotherapy was available. However, the non-chemotherapy group included all patients who had died before an adjuvant chemotherapy was available. Following that, for the chemotherapy group, survival estimates would be exaggerated so that we would acquire a conclusion

that favors the chemotherapy group more than the non-chemotherapy group[10]. Thus, to reduce GTB, finding out the guarantee time and extracting that time from the original survival time were necessary, and this related statistical strategy refers to landmark analysis[11]. To find out the proper landmark, the guarantee time point, we drew distribution of days between chemotherapy and surgery (Figure 1). In this case, the 95th percentile was chosen - 123 days (about 4 months) and was set as the landmark. This means within 4 months after the surgery most patients had chemotherapy if chemotherapy was part of the first course of the treatment plan. Thus, patients who died or were lost follow up within 4 months after surgery were excluded.

Nomogram and Model Validation

Nomogram and corresponding model validations were conducted using R Version 3.4.3. To set up the risk-score criteria, we built nomograms. Nomograms from Cox model is a popular visual plot that is widely used for cancer prognosis to show the predicted probabilities of an event for decision support. With a nomogram, we could reduce statistical predictive models and acquire a single numerical estimate of the probability of an event[12]. That is, the nomogram will produce a criterion assigning points for each value/level of each predictor and each scale of total points corresponds to a numerical estimate of the probability of the event. Then, after summing up points of their individual properties, every patient would have a total point and an estimated probability of the event. In this case, we used nomograms to predict 5-Year survival probability in patients with stage II colon cancer. Then, following our resulting nomogram, doctors could

acquire a total point and a 5-Year survival probability for each patient and decide which risk rank the patient is in immediately.

An R package 'hdnom' was mainly used to select variables, build the Cox model, plot the nomogram, and validate the resulting model[13]. With the function hdcox.aenet within hdnom package, we fitted a penalized Cox model by adaptive elastic-net regularization. A survival object, being needed as the response matrix for fitting, was built by survival package. Appropriate predictors were atomically selected from all variables we were interested in. Then, after generating the nomogram objects with hdnom.nomogram function, the nomograms were plotted. Considering the decision of "take chemotherapy or not" might be affected by doctors' subjective judgements, we set up two versions of nomograms (risk-score criteria): One was based on our whole study population while the other one included non-chemo or chemo-naive patients in our study population only.

Based on each version of risk-score criteria resulted by hdnom package, we acquired a risk-score for each subject by applying the criteria to the study population. According to the risk-score, we stratified patients into three groups: low-risk group, intermediate-risk group, high-risk group. The 33th percentile and the 66th percentile were chosen as stratification points.

Model validation was necessary to validate the predictive performance of a penalized Cox model. The hdnom package supports both internal validation and external model validation by resampling methods. In this case, we focused on internal model validation.

The function hdnom.validate within the hdnom package was used to estimate our adaptive elastic-net model performance internally through time-dependent area under the ROC curve (AUC) with bootstrap resampling method[12]. From the first year to the fifth year, the model was validated at every half year. After the validation, the model validation result was plotted.

Statistical Analysis

Statistical analysis was conducted using SAS Version 9.4. We defined months from 4-month post-surgery to death or last follow up as the OS. Single-variable and multivariable Cox proportional hazard model for OS were performed to identify the independent variables with a significant effect on patients' survival. Backward selection with an alpha level of removal of .10 was used. Kaplan-Meier survival curves were constructed; differences between chemo group and non-chemo group, and differences between each risk-group were tested by the log-rank test. The same analyses were applied with each version of risk-stratification. Single-variable survival analysis for chemotherapy was also performed within the whole population and each risk group of two criteria. Performing these analyses, the following variables were considered: Lymph-vascular invasion, Surgical Margins Status at any CoC Facility, CS SSF8 - Perineural Invasion, AJCC Pathologic T, Age at diagnosis, Charlson-Deyo Score, Grade, Facility Type, Facility Location, Primary Payer, Median Income Quartiles 2000, Race, Spanish Hispanic Origin, Percent No High School Degree Quartiles 2000, Year of diagnosis.

Results

Patient characteristics

31375 patients, including patients who were diagnosed at age 18 to 90, were eligible for analysis with the median age 71, 83.9% were white, and 47.4% were male. Among them, 3956 (12.6%) patients took chemotherapy after surgery while 27419 (87.4%) patients did not take chemotherapy after surgery. 3967 (12.6%) patients' Lymph-vascular invasion were present or identified and 27408 (87.4%) patients' Lymph-vascular invasion were absent or not identified. 491 (1.6%) patients had residual tumor and 30884 (98.4%) patients had no residual tumor (Table 1).

Nomograms and Model Validation

Based on the whole study population, we acquired a nomogram which selected five variables including Lymph-vascular invasion, Surgical Margins Status at any CoC Facility, CS SSF8 - Perineural Invasion, Age at diagnosis, and Charlson-Deyo Score (Figure 2). Following the nomogram which has a total point score of 160, each patient got a risk-score (Table 2). Setting 33th and 66th percentiles as stratification points, we had 0-76 points as the low-risk group, 77-99 points as the intermediate-risk group, 100-160 points as the high-risk group. Following this criterion, we had 11524 patients in low-risk group, 8903 patients in intermediate-risk group, and 10948 patients in high-risk group (Table 3).

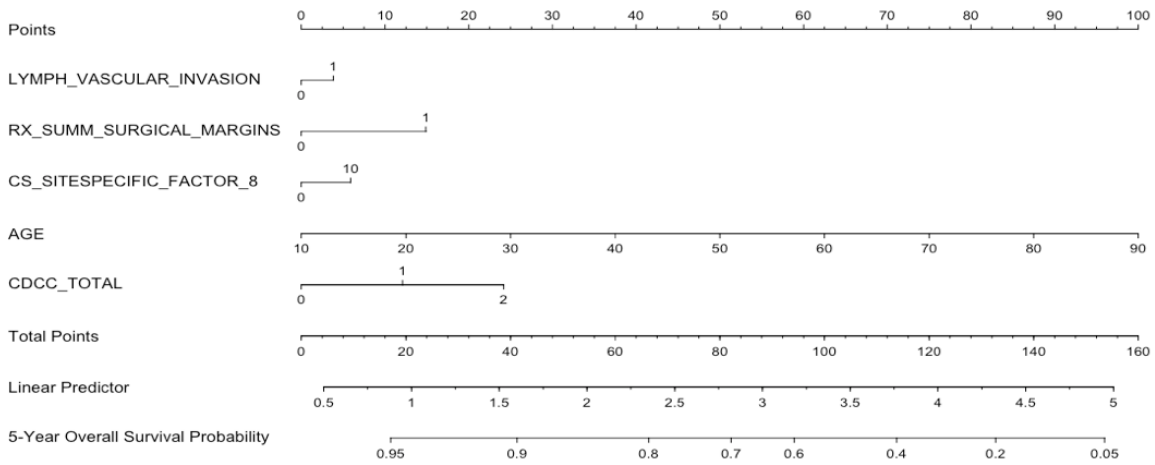


Figure 2. Nomogram based on whole study population

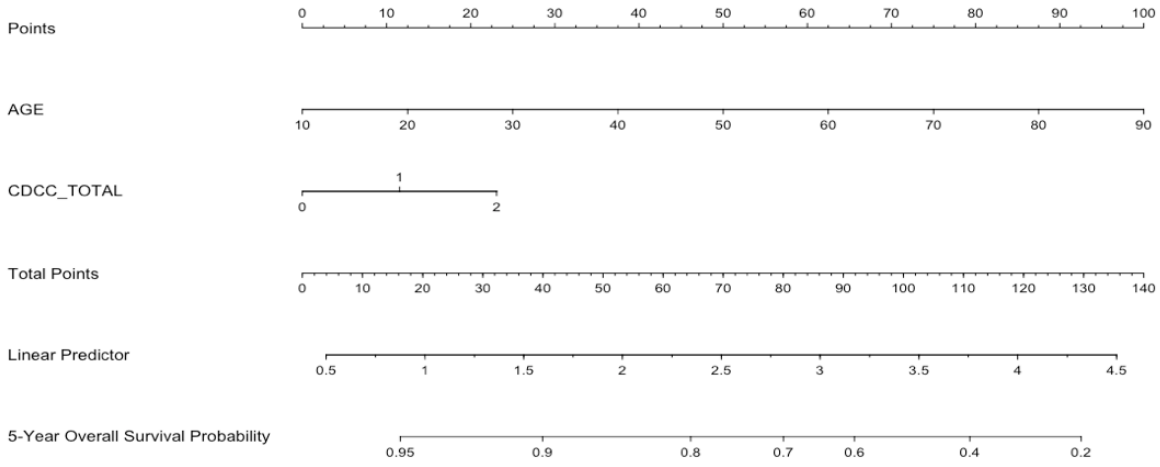


Figure 3. Nomogram based on non-Chemo patients only

Based on the patients who did not take chemotherapy after surgery, we acquired a nomogram including only two variables, Age at diagnosis and Charlson-Deyo Score (Figure 3). The total point of the nomogram is 140 (Table 4). Following the same 33th-66th-quantile stratification, we had 0-75 points as the low-risk group, 76-98 points as the intermediate-risk group, 99-140 as the high-risk group. With such criterion, there were

12585 patients in low-risk group, 8109 patients in intermediate-risk group, and 10681 patients in high-risk group (Table 5).

Results of internal validation for the two nomograms were plotted (Figure 4-5). The trends of the two results were similar. They both acquired fairly high AUC values that followed a great performance within the first four years. Overall, the AUC values of the model based on whole population within the first four years are higher than that of the model based on non-Chemo patients only. However, the AUC values of both models decreased significantly at the fifth year. In general, the performance of the two models were similar.

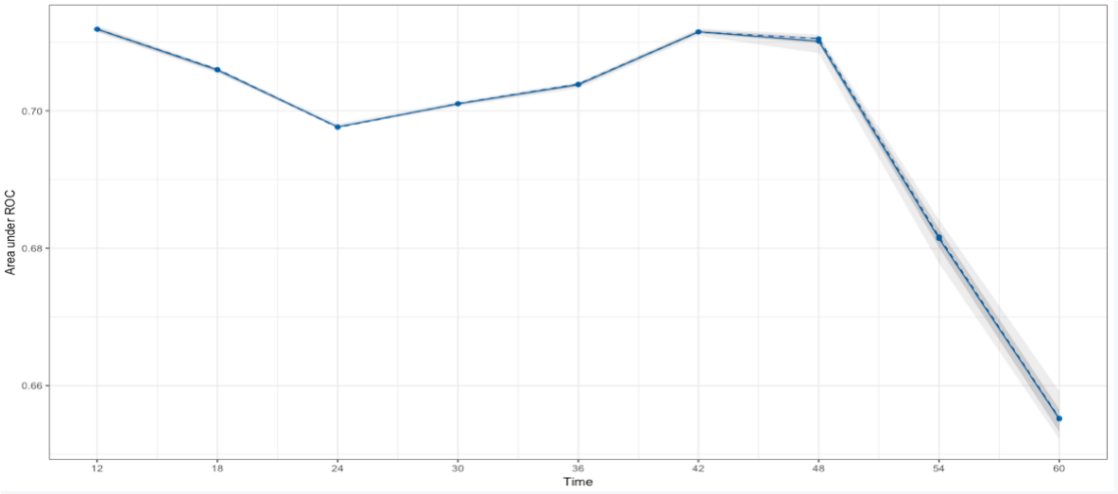


Figure 4. Interval validation for nomogram, based on whole population

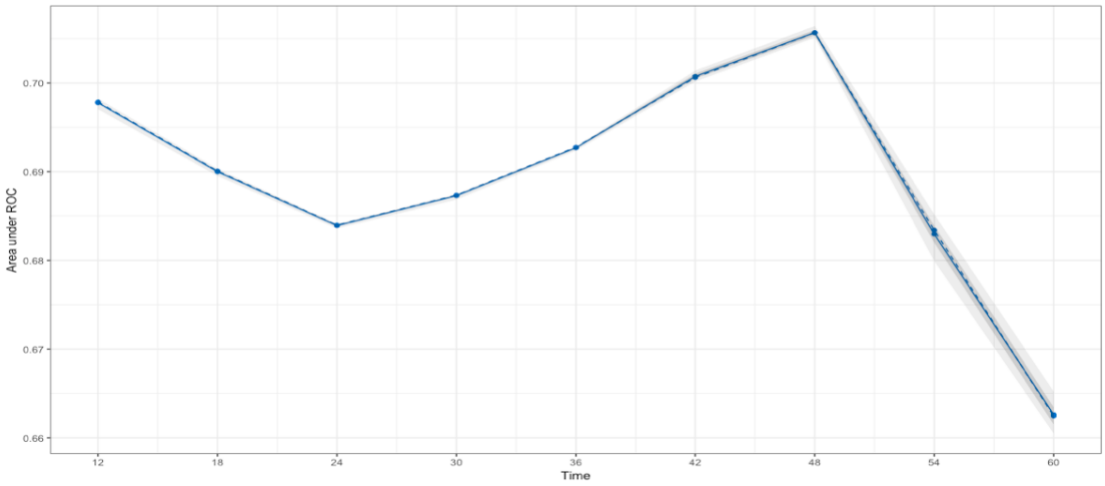
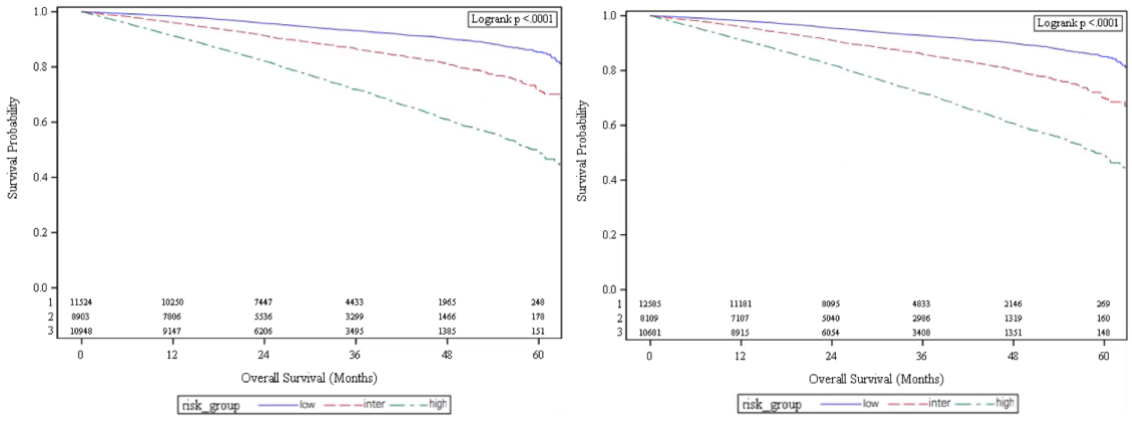


Figure 5. Interval validation for nomogram, based on non-Chemo patients only

Kaplan-Meier Curves

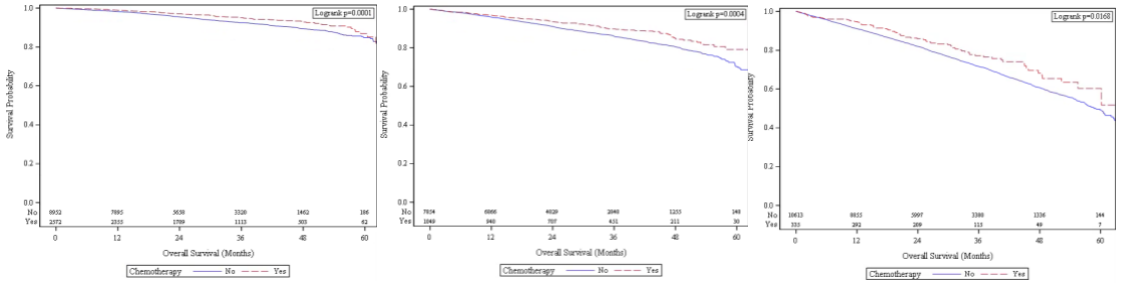
Kaplan-Meier survival curves which showed the association between risk-group and patients' overall survival (OS) were constructed within each risk-score criterion (Figure 6, Figure 7). Under both criteria, differences between each risk-group were statistically significant with p-value less than 0.0001. The survival probability of patients in the low-risk group was the highest among patients of the three groups and the survival probability of patients in the intermediate-risk group was higher than that of patients in the high-risk group. Kaplan-Meier survival curves that drew the association between chemotherapy and patients' OS were constructed within each risk group (Figure 8, Figure 9).



Kaplan-Meier survival curve for overall survival

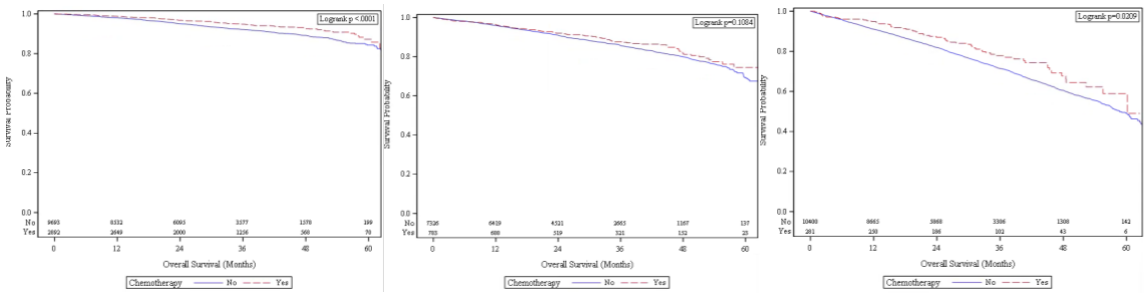
Figure 6. Based on whole-population criterion

Figure 7. Based on non-Chemo criterion



Low-risk group Intermediate-risk group High-risk group

Figure 8. Kaplan-Meier survival curve for overall survival within each risk group (Based on whole-population criterion)

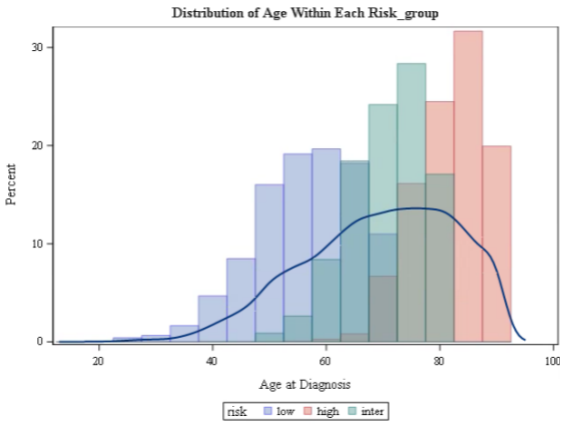


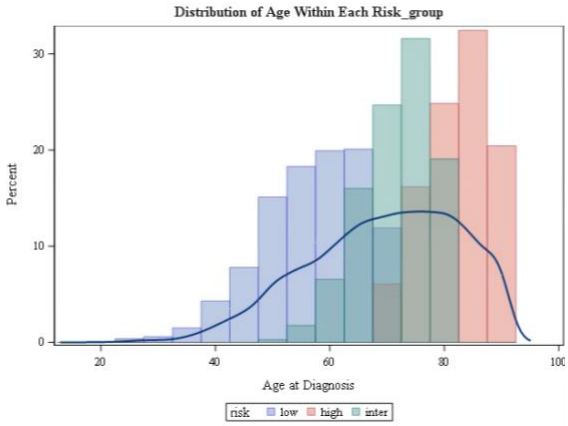
Low-risk group Intermediate-risk group High-risk group

Figure 9. Kaplan-Meier survival curve for overall survival within each risk group (Based on non-chemo criterion)

Under the whole-population criterion, differences between chemotherapy and non-chemotherapy were statistically significant with p-value less than 0.05 within each risk-group. Within each risk-group, the survival probability of patients who took chemotherapy was higher than that of non-chemotherapy patients. Under the non-chemo-population criterion, the differences were statistically significant with p-value less than 0.05 within the low-risk and high-risk groups while there was no statistically significant difference within the intermediate-risk group. Within the low and high-risk groups, patients who took chemotherapy acquired higher survival probability than non-chemotherapy patients did. Within the intermediate-risk group, the survival probability of patients who took chemotherapy and that of non-chemotherapy patients were similar. Corresponding patients' chemotherapy information within each risk group are shown in Table 6 and Table 7. That revealed that fewer patients took chemotherapy in high-risk groups. Distribution of age at diagnosis is shown in Figure 10 and Figure 11. The figures of distribution were quite similar under both criteria. There were some overlaps between the risk-groups especially between the intermediate-risk group and the high-risk group

and the age of most patients in low-risk group was less than that of patients in other two groups.





Age distribution within each risk group

Figure 10. Based on whole-population criterion **Figure 11. Based on non-chemo criterion**

Single-variable Survival Analysis

Single-variable survival analysis for OS showed that non-Chemo was significantly associated with OS compared to taking-Chemo (with non-Chemo criterion: P<0.001, HR=2.34, 95% CI=2.09-2.63; with whole-population criterion: P<0.001, HR=2.34, 95% CI= 2.09-2.63). The hazard of dying for patients who did not take chemotherapy was more than two times that of patients who took chemotherapy (Table 8). We acquired the same significant association between chemotherapy and OS within all risk groups except the intermediate-risk group based on the non-chemotherapy criterion. The results under the two criteria were similar: the hazard of dying for patients within the high-risk group was approximately five times that of patients within the low-risk group and two times that of patients within the intermediate-risk group. The hazard of dying for patients with a Charlson-Deyo score of two or more was more than 2 times that of patients with a 0

score. Under both criteria, the following variables also contributed to OS of patients: Risk group, Lymph-vascular invasion, Surgical Margins Status at any CoC Facility, CS SSF8 - Perineural Invasion, Charlson-Deyo Score, Grade, Facility Type, Facility Location, Primary Payer, Median Income Quartiles 2000, Race, Spanish Hispanic Origin, Year of diagnosis, Age at diagnosis, and Total points (Table 8). The results we had within every risk groups were quite similar. Patients who were within the high-risk group, or had a Charlson-Deyo score of two or more faced higher risk of death.

Under both criteria, patients who had present Lymph Vascular invasion, had a residual tumor, present CS SSF8-Perineural invasion, or a zero in AJCC Pathologic T were more likely to take chemotherapy (Table 9).

Multivariable Survival Analysis

Multivariable survival analysis for OS indicated that there existed a statistically significant association between chemotherapy and OS (HR = 1.23, 95%CI = 1.08, 1.39) after controlling for other selected covariates in the model (Table 10). Using backward selection with an alpha level of removal of .10, following variables were selected: chemotherapy, Lymph-vascular invasion, Surgical Margins Status at any CoC Facility, CS SSF8 - Perineural Invasion, Charlson-Deyo Score, Facility Type, Primary Payer, Median Income Quartiles 2000, Race, Percent No High School Degree Quartiles 2000, Year of diagnosis, and Age at diagnosis. Non-Chemo, present Lymph-vascular invasion, present residual tumor, present CS SSF8 - Perineural Invasion, two or larger than two Charlson-Deyo Score, being treated in Community Cancer Program/Other Facility Type,

Medicaid insured, living in area where residents' median household income is lower than $30,000, black, living in area where percent of residents who do not graduate from high school is 20-28.9, diagnosed in 2013, old age were all contribute factors to shorter OS for patients. The most important factors that are needed to predict the patient's survival are chemotherapy, CS SSF8 - Perineural Invasion, Race, and Age at Diagnosis. Under both criteria, senior patients had a higher risk of death. Patients who did not take chemotherapy, had present CS SSF8-Perineural invasion, or were black faced a higher risk of death.

Multivariable survival analysis for OS stratified by risk group specifying effect of chemotherapy was performed under both criteria. The estimated treatment effect was controlled by baseline covariates that had not been used for defining the risk group. Under non-Chemo criterion, CS SSF8 - Perineural Invasion, Facility Type, Lymph Vascular Invasion, Median Income Quartiles 2000, Percent No High School Degree Quartiles 2000, Primary Payor, Race, Spanish Hispanic Origin, Surgical Margins Status at any CoC Facility, and Year of Diagnosis were used to control the estimated treatment effect. Under whole-population criterion, Facility Type, Median Income Quartiles 2000, Percent No High School Degree Quartiles 2000, Primary Payor, Race, Spanish Hispanic Origin, Year of Diagnosis were used. Based on whole-population criterion, taking chemotherapy was statistically significantly associated with OS within all risk groups: low-risk group (P-value: 0.001, HR: 1.38, 95% CI: 1.13-1.68); intermediate-risk group (P-value: 0.005, HR: 1.34, 95% CI: 1.09-1.63); high-risk group (P-value: 0.019, HR: 1.32, 95% CI: 1.05-1.67) Based on non-chemotherapy criteria, the significant association

between taking chemotherapy and OS only appeared in the low-risk and high-risk groups: low-risk group (P-value: <.001, HR: 1.48; 95% CI: 1.23-1.78); high-risk group (P-value: 0.003, HR: 1.46, 95% CI: 1.14-1.88) (Table 11-12). Under both criteria, non-Chemo was a contributing factor to patients' shorter OS regardless of patients' risk level. Patients within the low-risk group benefited more from chemotherapy than patients within the other two risk groups. The non-significant p-values for interaction terms between the chemo and risk groups may indicate that the treatment effect of chemo on OS may not differ by patient's risk category, however it may be due to the fact that a small number of patients had been treated by chemo and we usually need a larger sample size to detect the significance for interactions. Overall, we can see the significant benefit of adjuvant chemotherapy in most of risk groups, which may suggest that utilization of adjuvant chemotherapy should be considered in stage II colon cancer patients in general.

Discussion

According to our results, whether we use all-patients risk stratification or non-chemotherapy risk stratification, patients in low-risk group increased their survival probability more by chemotherapy than those in the intermediate or high risk group. However, bias that resulted from age of diagnosis should be considered. We built two nomograms including one based on all patients and the other based on non-chemotherapy patients only. In both nomograms, age covered a high portion of calculating a patient's risk score (Figure 2, Figure 3) and in multivariable analysis, age was one of the variables that was selected within all risk groups. Such results also existed in the distribution of age within each risk group (Figure 10, Figure 11). Although there was some overlap between

the groups, we could see that most patients in the low-risk group were younger than most patients in the other two groups and most patients in the high-risk group were older than most patients in the other two groups. In other words, age played a very important role in risk stratification. But, from patients' chemotherapy information (Table 6, Table 7), whether we use all-patients risk stratification or non-chemotherapy risk stratification, there were less patients who took chemotherapy in high-risk group than that in other two risk groups. The reason was that doctors tended to assign chemotherapy to young patients since doctors might think senior patients may not tolerate the side effect of chemotherapy as well compared to their younger counterpart. However, such assignment based on doctors' subjective judgements may introduce substantial selection bias when we try to draw conclusions about the effect of chemo on OS for this study population. Even we observed an overall benefit of chemotherapy, the usage of chemo in high risk patients is still less than other lower risk groups in practice. Limited by the nature of the observational study design of this study, the findings may still be biased by unobserved confounders and will warrant validation by future prospective studies.

References

1. *Colon Cancer Treatment (PDQ®).* NCI. May 12, 2014.

2. *Cancer Stat Facts: Colorectal Cancer*. March 24, 2018, Retrieved from https://seer.cancer.gov/statfacts/html/colorect.html

3. *What is Cancer Staging*. AJCC. Retrieved from https://cancerstaging.org/references-tools/Pages/What-is-Cancer-Staging.aspx

4. *Cancer Staging*. NIH. March 9, 2015. Retrieved from https://www.cancer.gov/about-cancer/diagnosis-staging/staging

5. Carolyn Vachani and Bruce Giantonio, *Stage II Colon Cancer: To Treat or Not to Treat*. April 3, 2013. Retrieved from https://www.oncolink.org/cancers/gastrointestinal/colon-cancer/treatments/stage-ii-colon-cancer-to-treat-or-not-to-treat

6. The American Cancer Society medical and editorial content team. *Treatment of Colon Cancer, by Stage*. February 21, 2018. Retrieved from https://www.cancer.org/cancer/colon-rectal-cancer/treating/by-stage-colon.html

7. Joseph Kannarkatt, Joe Joseph, Peter C. Kurniali, Anas Al-Janadi, and Borys Hrinczen. *Adjuvant Chemotherapy for Stage II Colon Cancer: A Clinical Dilemma*. Journal of Oncology Practice 13, no. 4 (April 2017) 233-241.

8. *National Cancer Database*. Retrieved from https://www.facs.org/quality-programs/cancer/ncdb

9. *The CoC/NCDB PUF Data Dictionary Items*. Retrieved from http://ncdbpuf.facs.org/node/259?q=print-pdf-all

10. Giobbie-Hurder, Anita, Richard D. Gelber, and Meredith M. Regan. *Challenges of Guarantee-Time Bias*. Journal of Clinical Oncology 31.23 (2013): 2963–2969. PMC. Web. 28 Mar. 2018.

11. Urania Dafni. *Landmark Analysis at the 25-Year Landmark Point*. Cardiovascular Quality and Outcomes. 2011;4:363-371, originally published May 17, 2011

12. Alexia Iasonos, Deborah Schrag, Ganesh V. Raj, Katherine S. Panageas. *How To Build and Interpret a Nomogram for Cancer Prognosis*. Journal of Clinical Oncology 26, no. 8 (March 2008)

13. Miaozhu Li, Nan Xiao, 2017. *An Introduction to hdnom*. Retrieved from https://cran.r-project.org/web/packages/hdnom/vignettes/hdnom.html#introduction

Appendix

Table 1. Summary table for all variables of interest

Variable	Level	N (%) = 31375
Chemotherapy	No	27419 (87.4)
	Yes	3956 (12.6)
Lymph Vascular Invasion	Not present	27408 (87.4)
	present	3967 (12.6)
Surgical Margins Status at any CoC Facility	no	30884 (98.4)
	yes	491 (1.6)
CS SSF8 - Perineural Invasion	Not present	29555 (94.2)
	Present	1820 (5.8)
AJCC Pathologic T	0	241 (0.8)
	1	31134 (99.2)
Charlson-Deyo Score	0	20820 (66.4)
	1	7510 (23.9)
	2+	3045 (9.7)

Variable	Level	N (%) = 31375
Grade	Well differentiated, differentiated, NOS	3058 (9.7)
	Moderately differentiated, moderately well differentiated, intermediate differentiation	22726 (72.4)
	Poorly differentiated	4165 (13.3)
	Undifferentiated, anaplastic	788 (2.5)
	Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia	638 (2.0)

Variable	Level	N (%) = 31375
Facility Type	Community Cancer Program/Other	4514 (14.6)
	Comprehensive Community Cancer Program	15298 (49.5)
	Academic/Research Program	7596 (24.6)
	Integrated Network Cancer Program	3489 (11.3)
	Missing	478
Facility Location	Northeast	6052 (19.6)
	South	11986 (38.8)
	Midwest	8469 (27.4)
	West	4390 (14.2)
	Missing	478

Variable	Level	N (%) = 31375
Primary Payor	Not Insured	1201 (3.8)
	Private Insurance	9331 (29.7)
	Medicaid	1279 (4.1)
	Medicare	18942 (60.4)
	Other Government	239 (0.8)
	Insurance Status Unknown	383 (1.2)
Median Income Quartiles 2000	Not Available	826
	< $30,000	4148 (13.6)
	$30,000 - $35,999	5569 (18.2)
	$36,000 - $45,999	8703 (28.5)
	$46,000 +	12129 (39.7)
Race	White	26320 (83.9)
	Black	3681 (11.7)
	Other	1374 (4.4)

Variable	Level	N (%) = 31375
Spanish Hispanic Origin	Non-Spanish; non-Hispanic	28679 (91.4)
	Spanish or Hispanic	1545 (4.9)
	Unknown	1151 (3.7)
Percent No High School Degree Quartiles 2000	Not Available	828
	>=29%	5197 (17.0)
	20-28.9%	7289 (23.9)
	14-19.9%	7301 (23.9)
	< 14%	10760 (35.2)
Year of Diagnosis	2010	7438 (23.7)
	2011	7975 (25.4)
	2012	8045 (25.6)
	2013	7917 (25.2)
Sex	Male	14882 (47.4)
	Female	16493 (52.6)

Variable	Level	N (%) = 31375
Age at Diagnosis	Mean	69.41
	Median	71.00
	Minimum	18.00
	Maximum	90.00
	Std Dev	13.11
	Missing	0.00

Table 2. Risk score criteria (based on the whole study population)

LYMPH_VASCULAR_INVASION	Points
0	0
1	4
RX_SUMM_SURGICAL_MARGINS	Points
0	0
1	15
CS_SITESPECIFIC_FACTOR_8	Points
0	0
10	6
AGE	Points
10-19	0

20-29	12
30-39	25
40-49	38
50-59	50
60-69	62
70-79	75
80-89	88
90-90+	100
CDCC_TOTAL (Charlson-Deyo Score)	Points
0	0
1	12
2	24
Total Points	5-Year Overall Survival Probability
154-160	0.05
133-153	0.20
114-132	0.40
94-113	0.60
82-93	0.70
66-81	0.80
41-65	0.90
17-40	0.95

Table 3. Patients within each risk group (Based on whole-population criterion)

risk_group	No. of Subject	Event	Censored	Median Survival (95% CI)	12 Mo Survival	60 Mo Survival
Low	11524	736 (6%)	10788 (94%)	NA (NA, NA)	98.4% (98.1%, 98.6%)	85.4% (83.5%, 87.0%)
Intermediate	8903	1147 (13%)	7756 (87%)	NA (NA, NA)	96.0% (95.6%, 96.4%)	71.5% (68.8%, 74.1%)
High	10948	2909 (27%)	8039 (73%)	59.8 (57.3, 60.8)	91.2% (90.6%, 91.7%)	49.7% (47.3%, 52.0%)

Table 4. Risk score criteria (based on non-Chemo patients only)

AGE	Points
10-19	0
20-29	12
30-39	25
40-49	38
50-59	50

60-69	62
70-79	75
80-89	88
90-90+	100
CDCC_TOTAL (Charlson-Deyo Score)	Points
0	0
1	12
2	23
Total Points	5-Year Overall Survival Probability
130-140	0.20
111-129	0.40
92-110	0.60
80-91	0.70
65-79	0.80
40-64	0.90
16-39	0.95

Table 5. Patients within each risk group (Based on non-Chemo criterion)

risk_group	**No. of Subject**	**Event**	**Censored**	**Median Survival (95% CI)**	**12 Mo Survival**	**60 Mo Survival**

Low	12585	842 (7%)	11743 (93%)	NA (NA, NA)	98.2% (97.9%, 98.4%)	85.1% (83.3%, 86.6%)
Intermediate	8109	1093 (13%)	7016 (87%)	NA (NA, NA)	95.9% (95.5%, 96.3%)	70.0% (67.0%, 72.8%)
High	10681	2857 (27%)	7824 (73%)	58.7 (57.2, 60.6)	91.1% (90.5%, 91.6%)	49.4% (47.0%, 51.8%)

Table 6. Patients' Chemo information within each risk group (Based on whole-population criterion)

Group	Chemotherapy	No. of Subject	Event	Censored	12 Mo Survival	60 Mo Survival
Low	No	8952	605 (7%)	8347 (93%)	98.2% (97.9%, 98.5%)	84.9% (82.9%, 86.7%)

	Yes	2572	131 (5%)	2441 (95%)	98.9% (98.4%, 99.2%)	86.9% (82.3%, 90.3%)
Intermediate	No	7854	1039 (13%)	6815 (87%)	95.9% (95.4%, 96.3%)	70.3% (67.2%, 73.2%)
	Yes	1049	108 (10%)	941 (90%)	96.9% (95.6%, 97.8%)	79.2% (73.5%, 83.8%)
High	No	10613	2836 (27%)	7777 (73%)	91.1% (90.5%, 91.6%)	49.3% (46.9%, 51.7%)
	Yes	335	73 (22%)	262 (78%)	94.8% (91.7%, 96.7%)	60.3% (49.4%, 69.7%)

Table 7. Patients' Chemo information within each risk group (Based on non-chemo criterion)

	Chemotherapy	**No. of Subject**	**Event**	**Censored**	**12 Mo Survival**	**60 Mo Survival**

Group						
Low	No	9693	691 (7%)	9002 (93%)	98.0% (97.7%, 98.3%)	84.4% (82.4%, 86.1%)
	Yes	2892	151 (5%)	2741 (95%)	98.7% (98.3%, 99.1%)	87.2% (83.1%, 90.4%)
Intermediate	No	7326	995 (14%)	6331 (86%)	95.9% (95.4%, 96.3%)	69.4% (66.1%, 72.4%)
	Yes	783	98 (13%)	685 (87%)	96.4% (94.8%, 97.5%)	74.5% (67.5%, 80.2%)
High	No	10400	2794 (27%)	7606 (73%)	91.0% (90.4%, 91.6%)	49.1% (46.7%, 51.5%)
	Yes	281	63 (22%)	218 (78%)	94.9% (91.5%, 96.9%)	58.8% (46.8%, 69.0%)

Table 8. Single-variable analysis of overall survival

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Chemotherapy	No	27419	2.34 (2.09-2.63)	**<.001**	**<.001**
	Yes	3956	-	-	
risk_group (Non-Chemo criterion)*	Low	12585	0.22 (0.20-0.24)	**<.001**	**<.001**
	Intermediate	8109	0.46 (0.43-0.49)	**<.001**	
	High	10681	-	-	
risk_group (Whole-population criterion)*	Low	11524	0.21 (0.20-0.23)	**<.001**	**<.001**

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
	Intermediate	8903	0.44 (0.41-0.47)	**<.001**	
	High	10948	-	-	
Lymph Vascular Invasion	Not present	27408	0.83 (0.77-0.90)	**<.001**	**<.001**
	present	3967	-	-	
Surgical Margins Status at any CoC Facility	no	30884	0.64 (0.53-0.77)	**<.001**	**<.001**
	yes	491	-	-	
CS SSF8 - Perineural Invasion	Not present	29555	0.73 (0.66-0.81)	**<.001**	**<.001**
	Present	1820	-	-	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
AJCC Pathologic T	0	241	0.73 (0.51-1.05)	0.092	0.089
	1	31134	-	-	
Charlson-Deyo Score	0	20820	0.39 (0.36-0.42)	**<.001**	**<.001**
	1	7510	0.56 (0.51-0.61)	**<.001**	
	2+	3045	-	-	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Grade	Well differentiated, differentiated, NOS	3058	1.13 (0.89-1.44)	0.320	**0.018**
	Moderately differentiated, moderately well differentiated, intermediate differentiation	22726	1.14 (0.91-1.43)	0.269	
	Poorly differentiated	4165	1.29 (1.02-1.63)	**0.036**	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
	Undifferentiated, anaplastic	788	1.25 (0.94-1.65)	0.129	
	Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia	638	-	-	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Facility Type	Community Cancer Program/Other	4514	1.17 (1.05-1.31)	**0.006**	**<.001**
	Comprehensive Community Cancer Program	15298	1.13 (1.03-1.24)	**0.013**	
	Academic/Research Program	7596	0.91 (0.82-1.01)	0.082	
	Integrated Network Cancer Program	3489	-	-	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Facility Location	Northeast	6052	1.06 (0.96-1.17)	0.273	**0.004**
	South	11986	1.00 (0.92-1.10)	0.945	
	Midwest	8469	1.13 (1.03-1.24)	**0.010**	
	West	4390	-	-	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Primary Payor	Not Insured	1201	0.56 (0.40-0.78)	**<.001**	**<.001**
	Private Insurance	9331	0.45 (0.34-0.59)	**<.001**	
	Medicaid	1279	0.85 (0.63-1.17)	0.319	
	Medicare	18942	1.37 (1.05-1.80)	**0.022**	
	Other Government	239	0.62 (0.38-1.01)	0.055	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
	Insurance Status Unknown	383	-	-	
Median Income Quartiles 2000	< $30,000	4148	1.11 (1.02-1.22)	**0.018**	**0.025**
	$30,000 - $35,999	5569	1.08 (0.99-1.17)	0.072	
	$36,000 - $45,999	8703	0.99 (0.92-1.06)	0.788	
	$46,000 +	12129	-	-	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Race	White	26320	1.89 (1.56-2.28)	**<.001**	**<.001**
	Black	3681	1.57 (1.28-1.93)	**<.001**	
	Other	1374	-	-	
Spanish Hispanic Origin	Non-Spanish; non-Hispanic	28679	1.03 (0.89-1.20)	0.661	**<.001**
	Spanish or Hispanic	1545	0.68 (0.55-0.85)	**<.001**	
	Unknown	1151	-	-	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Percent No High School Degree Quartiles 2000	>=29%	5197	1.03 (0.95-1.12)	0.504	0.063
	20-28.9%	7289	1.07 (0.99-1.16)	0.072	
	14-19.9%	7301	1.10 (1.02-1.19)	**0.012**	
	< 14%	10760	-	-	

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Year of Diagnosis	2010	7438	0.86 (0.77-0.95)	**0.003**	**0.002**
	2011	7975	0.85 (0.77-0.94)	**0.002**	
	2012	8045	0.95 (0.86-1.05)	0.323	
	2013	7917	-	-	
Age at Diagnosis		31375	1.06 (1.06-1.06)	**<.001**	-
Total_points (Non-Chemo criterion)*		31375	1.04 (1.04-1.04)	**<.001**	-

			Overall Survival (Months)		
Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Log-rank P-value**
Total_points (Whole-population criterion)*		31375	1.04 (1.04-1.04)	**<.001**	-

*: The analysis tables under the two criteria only different in risk-group and total_points. So, the two tables were combined for convenience.

Table 9. Single-variable analysis for Chemotherapy

		Chemotherapy		
Covariate	**Level**	**No N=27419**	**Yes N=3956**	**Parametric P-value***
Lymph Vascular Invasion	Not present	24301 (88.63)	3107 (78.54)	**<.001**
	present	3118 (11.37)	849 (21.46)	
Surgical Margins Status at any CoC Facility	no	27057 (98.68)	3827 (96.74)	**<.001**
	yes	362 (1.32)	129 (3.26)	

		Chemotherapy		
Covariate	**Level**	**No N=27419**	**Yes N=3956**	**Parametric P-value***
CS SSF8 - Perineural Invasion	Not present	25999 (94.82)	3556 (89.89)	**<.001**
	Present	1420 (5.18)	400 (10.11)	
AJCC Pathologic T	0	232 (0.85)	9 (0.23)	**<.001**
	1	27187 (99.15)	3947 (99.77)	
Charlson-Deyo Score	0	17782 (64.85)	3038 (76.79)	**<.001**
	1	6775 (24.71)	735 (18.58)	
	2+	2862 (10.44)	183 (4.63)	
Grade	Well differentiated, differentiated, NOS	2739 (9.99)	319 (8.06)	**<.001**
	Moderately differentiated, moderately well differentiated, intermediate differentiation	20034 (73.07)	2692 (68.05)	
	Poorly differentiated	3444 (12.56)	721 (18.23)	
	Undifferentiated, anaplastic	668 (2.44)	120 (3.03)	
	Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia	534 (1.95)	104 (2.63)	

		Chemotherapy		
Covariate	**Level**	**No N=27419**	**Yes N=3956**	**Parametric P-value***
Facility Type	Community Cancer Program/Other	3932 (14.49)	582 (15.44)	**0.006**
	Comprehensive Community Cancer Program	13495 (49.75)	1803 (47.84)	
	Academic/Research Program	6605 (24.35)	991 (26.29)	
	Integrated Network Cancer Program	3096 (11.41)	393 (10.43)	
Facility Location	Northeast	5301 (19.54)	751 (19.93)	**<.001**
	South	10423 (38.42)	1563 (41.47)	
	Midwest	7498 (27.64)	971 (25.76)	
	West	3906 (14.4)	484 (12.84)	
Primary Payor	Not Insured	931 (3.4)	270 (6.83)	**<.001**
	Private Insurance	7308 (26.65)	2023 (51.14)	
	Medicaid	1002 (3.65)	277 (7)	
	Medicare	17653 (64.38)	1289 (32.58)	
	Other Government	198 (0.72)	41 (1.04)	
	Insurance Status Unknown	327 (1.19)	56 (1.42)	

		Chemotherapy		
Covariate	**Level**	**No N=27419**	**Yes N=3956**	**Parametric P-value***
Median Income Quartiles 2000	< $30,000	3565 (13.35)	583 (15.17)	**0.014**
	$30,000 - $35,999	4894 (18.33)	675 (17.56)	
	$36,000 - $45,999	7644 (28.62)	1059 (27.56)	
	$46,000 +	10603 (39.7)	1526 (39.71)	
Race	White	23134 (84.37)	3186 (80.54)	**<.001**
	Black	3123 (11.39)	558 (14.11)	
	Other	1162 (4.24)	212 (5.36)	
Spanish Hispanic Origin	Non-Spanish; non-Hispanic	25162 (91.77)	3517 (88.9)	**<.001**
	Spanish or Hispanic	1273 (4.64)	272 (6.88)	
	Unknown	984 (3.59)	167 (4.22)	
Percent No High School Degree Quartiles 2000	>=29%	4485 (16.8)	712 (18.53)	**<.001**
	20-28.9%	6326 (23.69)	963 (25.06)	
	14-19.9%	6391 (23.93)	910 (23.68)	
	< 14%	9502 (35.58)	1258 (32.73)	

		Chemotherapy		
Covariate	**Level**	**No N=27419**	**Yes N=3956**	**Parametric P-value***
Year of Diagnosis	2010	6454 (23.54)	984 (24.87)	**<.001**
	2011	6888 (25.12)	1087 (27.48)	
	2012	7036 (25.66)	1009 (25.51)	
	2013	7041 (25.68)	876 (22.14)	
risk_group (Non-Chemo criterion)[1]	Low	9693 (35.35)	2892 (73.1)	**<.001**
	Intermediate	7326 (26.72)	783 (19.79)	
	High	10400 (37.93)	281 (7.1)	
risk_group (Whole-population criterion)[1]	Low	8952 (32.65)	2572 (65.02)	**<.001**
	Intermediate	7854 (28.64)	1049 (26.52)	
	High	10613 (38.71)	335 (8.47)	
Age at Diagnosis	N	27419	3956	**<.001**
	Mean	70.91	58.96	
	Median	72	59	
Overall Survival (Months)	N	27419	3956	**<.001**
	Mean	29.64	32.86	
	Median	28.66	32.7	

		Chemotherapy		
Covariate	**Level**	**No N=27419**	**Yes N=3956**	**Parametric P-value***
Total_points (Non-Chemo criterion)[1]	N	27419	3956	**<.001**
	Mean	86.93	70.01	
	Median	88	74	
Total_points (Whole-population criterion)[1]	N	27419	3956	**<.001**
	Mean	88	72.01	
	Median	88	74	

* The parametric p-value is calculated by ANOVA for numerical covariates and chi-square test for categorical covariates.

1: The analysis tables under the two criteria only different in risk-group and total_points. So, the two tables were combined for convenience.

Table 10. Multivariable analysis of overall survival

		Overall Survival (Months)		
Covariate	**Level**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Chemotherapy	No	1.23 (1.08-1.39)	**0.001**	**0.001**
	Yes	-	-	
Lymph Vascular Invasion	Not present	0.86 (0.79-0.94)	**<.001**	**<.001**
	present	-	-	
Surgical Margins Status at any CoC Facility	no	0.62 (0.51-0.75)	**<.001**	**<.001**
	yes	-	-	
CS SSF8 - Perineural Invasion	Not present	0.72 (0.64-0.80)	**<.001**	**<.001**
	Present	-	-	

		Overall Survival (Months)		
Covariate	**Level**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Charlson-Deyo Score	0	0.49 (0.45-0.53)	**<.001**	**<.001**
	1	0.62 (0.57-0.68)	**<.001**	
	2+	-	-	

		Overall Survival (Months)		
Covariate	**Level**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Facility Type	Community Cancer Program/Other	1.13 (1.01-1.27)	**0.036**	**<.001**
	Comprehensive Community Cancer Program	1.11 (1.00-1.22)	**0.041**	
	Academic/Research Program	0.96 (0.86-1.07)	0.480	
	Integrated Network Cancer Program	-	-	

		Overall Survival (Months)		
Covariate	**Level**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Primary Payor	Not Insured	1.08 (0.77-1.53)	0.651	**<.001**
	Private Insurance	0.72 (0.54-0.95)	**0.022**	
	Medicaid	1.26 (0.92-1.74)	0.152	
	Medicare	0.89 (0.67-1.17)	0.389	
	Other Government	0.74 (0.45-1.22)	0.242	
	Insurance Status Unknown	-	-	

		Overall Survival (Months)		
Covariate	**Level**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Median Income Quartiles 2000	< $30,000	1.10 (0.98-1.24)	0.110	**0.008**
	$30,000 - $35,999	0.99 (0.89-1.09)	0.798	
	$36,000 - $45,999	0.92 (0.85-1.00)	0.051	
	$46,000 +	-	-	
Race	White	1.56 (1.28-1.89)	**<.001**	**<.001**
	Black	1.60 (1.30-1.97)	**<.001**	
	Other	-	-	

		Overall Survival (Months)		
Covariate	**Level**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Percent No High School Degree Quartiles 2000	>=29%	1.09 (0.97-1.23)	0.156	**0.017**
	20-28.9%	1.15 (1.04-1.26)	**0.005**	
	14-19.9%	1.12 (1.03-1.22)	**0.007**	
	< 14%	-	-	
Year of Diagnosis	2010	0.84 (0.76-0.94)	**0.001**	**<.001**
	2011	0.84 (0.76-0.94)	**0.001**	
	2012	0.97 (0.88-1.08)	0.583	
	2013	-	-	
Age at Diagnosis		1.05 (1.05-1.06)	**<.001**	**<.001**

* Number of observations in the original data set = 31375. Number of observations used = 30082.

** Backward selection with an alpha level of removal of .10 was used. The following variables were removed from the

model: Facility Location, Grade, Spanish Hispanic Origin, AJCC Pathologic T, and Total _points.

Table 11. Multivariable analysis with interaction (Based on non-chemo criterion)

			Overall Survival (Months)		

Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Comparisons Stratified by risk_group :	**Chemotherapy :**		-	-	0.297
Low	No vs. Yes	9121 vs. 2630	1.48 (1.23-1.78)	**<.001**	-
Intermediate	No vs. Yes	7134 vs. 759	1.20 (0.97-1.48)	0.088	-
High	No vs. Yes	10162 vs. 276	1.46 (1.14-1.88)	**0.003**	-

			Overall Survival (Months)		

Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
* Number of observations in the original data set = 31375. Number of observations used = 30082. ** Backward selection with an alpha level of removal of .10 was used. The following variables were removed from the model: Grade. *** The estimated stratified treatement effect was controlled by: CS SSF8 - Perineural Invasion, Facility Type, Lymph Vascular Invasion, Median Income Quartiles 2000, Percent No High School Degree Quartiles 2000, Primary Payor, Race, Spanish Hispanic Origin, Surgical Margins Status at any CoC Facility, Year of Diagnosis					

Table 12. Multivariable analysis with interaction (Based on whole-population criterion)

			Overall Survival (Months)		

Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Comparisons Stratified by risk_group :	**Chemotherapy :**		-	-	0.959

			Overall Survival (Months)		

Covariate	**Level**	**N**	**Hazard Ratio (95% CI)**	**HR P-value**	**Type3 P-value**
Low	No vs. Yes	8393 vs. 2318	1.38 (1.13-1.68)	**0.001**	-
Intermediate	No vs. Yes	7655 vs. 1022	1.34 (1.09-1.63)	**0.005**	-
High	No vs. Yes	10369 vs. 325	1.32 (1.05-1.67)	**0.019**	-

* Number of observations in the original data set = 31375. Number of observations used = 30082.

** Backward selection with an alpha level of removal of .10 was used. The following variables were removed from the model: Grade.

*** The estimated stratified treatement effect was controlled by: Facility Type, Median Income Quartiles 2000, Percent No

High School Degree Quartiles 2000, Primary Payor, Race, Spanish Hispanic Origin, Year of Diagnosis