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:

_____ _____

Ruowen Qi Date

Timing of neoadjuvant or adjuvant chemotherapy to surgery and its impact on overall survival among breast cancer patients

By

Ruowen Qi

Master of Science in Public Health

Department of Biostatistics and Bioinformatics

_____

Yuan Liu, Ph.D.

Thesis Advisor

_____

Yi-An Ko, Ph.D.

Thesis Reader

Timing of neoadjuvant or adjuvant chemotherapy to surgery and its impact on overall survival among breast cancer patients

By

Ruowen Qi

B.A.

Southwestern University of Finance and Economics

2018

Thesis Committee Chair: Yuan Liu, Ph.D.

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

2020

Abstract

Timing of neoadjuvant or adjuvant chemotherapy to surgery and its impact on overall survival among breast cancer patients

By Ruowen Qi

Background: Breast cancer is the second most diagnosed cancer in women, and one in eight women will live with invasive breast cancer during the period of their lifetime in the U.S. according to the research of American Cancer Society. Chemotherapy plus surgical procedures are effective treatments for breast cancer, but the optimal time interval between chemotherapy and surgery for overall survival is still unclear.

Objective: Identify the impact of time of neoadjuvant or adjuvant chemotherapy to surgery on overall survival among female breast cancer patients.

Methods: We divided the study population into two cohorts, neoadjuvant chemotherapy group and adjuvant chemotherapy group, and all the analyses were conducted for the two cohorts separately. To study the impact of the time lag between chemo and surgery on overall survival (OS), we categorized time into three-level time intervals to maximize the discrimination power to predict OS in Cox proportional hazards (PH) model. Univariate, multivariate and propensity score weighting methods based on Cox PH model were used to study the associations between the selected variables and the three-level time intervals, and the associations between variables and overall survival. Hazard ratio with 95% confidence interval and type-3 p-value were calculated to show the association.

Results: Sufficient imbalance in covariates distribution among the three intervals was detected for both cohorts. Multivariate analyses got similar results with propensity score weighting method. After balancing the covariates, it shown the start of surgery after 12 weeks from the start of chemo effectively benefits the overall survival in neoadjuvant chemo group; the start of chemo within 9 weeks from the surgery most contributing to prolonged survival, while receiving chemotherapy later than 12 weeks after the surgery apparently lower the survival rate in adjuvant chemo group.

Conclusion: These results make the importance of timing between chemotherapy and surgery to the overall survival of breast cancer patients clearer. More extensive studies on this time interval are needed for the well-being of breast cancer patients.

Timing of neoadjuvant or adjuvant chemotherapy to surgery and its impact on overall survival among breast cancer patients

By

Ruowen Qi

B.A.

Southwestern University of Finance and Economics

2018

Thesis Committee Chair: Yuan Liu, Ph.D.

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

2020

Table of Contents

1. Introduction

The population that is suffering from breast cancer in the United States keeps increasing, and it is the second most diagnosed cancer in women until today [1].

One in eight women, which is about twelve percent of the whole population in the U.S., will live with invasive breast cancer during the period of their lifetime, and it is estimated that 276,480 new cases of this cancer are expected to be diagnosed in women in the U.S. until 2020 from the research of American Cancer Society [2].

Chemotherapy is a treatment choice that uses drugs and chemicals to destroy cancer cells in the body [3]. The potential gain of chemotherapy includes the growth of the possibility of a cure, the decline in the risk rate of cancer returning, the symptom remission from cancer, and a better future life for the cancer survivors [3]. The chemotherapy given before the surgery is called neoadjuvant therapy, and chemotherapy used after surgery is called adjuvant chemotherapy. Chemotherapy is regularly used accompanied by other treatments such as surgery, hormone therapy, or radiation [3,4].

Neoadjuvant systemic therapy (NST) was mostly applied for patients with inoperable locally advanced breast cancer from the 1980s [5]. NST has become a more popular treatment option for patients with early-stage breast cancer (BC) in recent decades. Some theories suggested that NSC may result in more accelerated eradication of micrometastasis of cancer cells, enhanced body image compared with complete breast removal and increased overall survival (OS) [6-8]. Subsequent randomized clinical trials proved that NST could reduce the extent of breast surgery and allow the higher possibility of breast-conserving surgery (BCS) without imperiling long-term survival [9].

Further, NSC can help provide convenience for researchers to obtain information about the impact of systemic therapies on cancer biology [6]. Nowadays NSC is being applied more constantly based on the receptor status in patient bodies and the nodal burden of cancer [7].

The surgical approach to the breast after NST is essential, but the optimal time interval between the end of NST and definitive surgery for overall survival is still unclear [6,7]. In some clinical practice, primary tumor removal is commonly implemented within a few weeks from the finalization of chemotherapy. Still it is undiscovered if a delay in surgery would harm the benefit of past systemic treatment [6]. From the results in a study published in Annals of Surgical Oncology journal in 2015, patients who had surgery after neoadjuvant therapy more than eight weeks held obviously worse overall survival [10].

In opposition to the neoadjuvant chemotherapeutic series, the solid data on optimal time interval after surgery for adjuvant chemotherapy is available [10]. Although the exact optimal interval between surgery and adjuvant chemotherapy is still unclear, there are accessible data displaying an important reduction in adjuvant therapy effectiveness when it is executed more than twelve weeks after the tumor removal surgery [6]. The European Society of Medical Oncology advised in its guidelines that systemic adjuvant treatment needs to be given in preference within two to six weeks after surgery [6].

In this study, we aim to figure out the impact of time of neoadjuvant or adjuvant chemotherapy to surgery on overall survival among female breast cancer patients. All the data was collected from the National Cancer Database (NCDB) for breast cancer, which represents almost 70% of new cancer diagnoses in the United States. Two cohorts

consisted of the study population, patients with neoadjuvant chemotherapy and patients with adjuvant chemotherapy. All analyses were conducted separately for two cohorts.

2. Methods

2.1 Study design and target population

All the data for this project comes from the National Cancer Database (NCDB) for breast cancer. The National Cancer Database, a joint project between the American College of Surgeons and American Cancer Society that provides de-identified data from over 1500 hospitals affiliated with the Commission on Cancer program, which represents approximately 70% of new cancer diagnoses in the United States. The Participant User File (PUF) was a publicly available file consisting of the NCDB dataset for each cancer type [11].

The criteria for selecting the study population are listed in Table 1. Patients were included only if they had a single malignant primary and excluded if the patients had previous or concurrent malignancy. Patients were excluded if they got cancer diagnosis at the reporting facility but not treated there. Patients were included if they had confirmed invasive breast cancer. All male patients were excluded. Patients were included if they got histology diagnostic confirmation. Patients were included if they had AJCC clinical stage T1-2N0-1M0 with the exclusion of metastatic cases. Patients were included when they had chemotherapy and surgical procedure of primary site at any CoC facility. Cases with missing value in relevant covariates (e.g., facility type, income, education, radiation status) or outcome were excluded. The original NCDB breast cancer database contains

2,696,734 cases from 2004-2016, and after sequential selection and exclusion, the target population includes 320,900 cases from 2004-2015.

We divided the study population into two cohorts based on the relative time point of chemotherapy to surgery, such as patients who received chemotherapy before surgery (neoadjuvant chemotherapy, N=70,731) and who received chemotherapy after surgery (adjuvant chemotherapy, N=250,169). The impact of the time lag between chemo and surgery on overall survival was examined for the two cohorts separately.

2.2 Definition of exploratory and outcome variables

For the cohort with neoadjuvant chemotherapy, we defined three-time intervals: T1 is the weeks from cancer diagnosis to the start of chemotherapy, T2 is the weeks from cancer diagnosis to definitive surgical procedure, and T3 = T1-T2 is the weeks between the start of chemotherapy and definitive surgical procedure. Because chemotherapy was performed before the surgical procedure, all values of T3 were negative. We used the absolute value of T3 in the analysis. The range of T2 was set from 8 weeks to 36 weeks, since the shortest cycle of chemotherapy is 8 weeks, so surgery could only be performed at least 8 weeks after diagnosis. The range of T3 was larger than 8 weeks for the same reason. Meanwhile, only 1.68% of patients received surgical procedures after 36 weeks of diagnosis; we excluded these extreme values for more stable results of analysis. The overall survival (OS) was defined as the months from definitive surgical procedure to last contact or death for this population.

For the cohort with adjuvant chemotherapy, the same three-time interval variables were created. All values of T3 were positive in this cohort as the chemotherapy was applied after surgery. The range of T1 was set from 0 to 36 weeks since 99.19% patients started

chemotherapy within 36 weeks from diagnosis. The range of T2 was set from 0 to 24 weeks for 99.43% patients got surgical procedure within 24 weeks from diagnosis. All extreme values were removed for more explicable results. The overall survival (OS) was defined as the months from the date of chemotherapy stated to the last contact or death for this cohort.

For the weeks between the start of chemotherapy and definitive surgical procedure, defined as T3 for both neoadjuvant and adjuvant settings, we categorized it into a three-level time intervals in a way to maximize the discrimination power to predict OS in Cox proportional hazards model. The R package 'CatPredi' was applied to find two optimal cut-points of T3 for both populations [12].

2.3 Definition of covariates

For the cohort with neoadjuvant chemotherapy, we selected 18 baseline clinical characteristics: race, age at diagnosis, median household income at patient residence area, educational attainment at patient residence area, Charlson-Deyo score, reporting facility type, patient's primary insurance carrier, year of cancer diagnosis, tumor's resemblance to normal tissue, ER, PR, HER2, clinically-determined size or extension of the primary tumor, clinically-determined absence or presence of regional lymph node metastasis, clinically-determined absence or presence of distant metastasis, whether patient received radiation therapy, final status of the surgical margins after resection of the primary tumor, response to neoadjuvant chemotherapy.

For the cohort with adjuvant chemotherapy, we selected 17 baseline covariates. The covariates are the same as those for population one except response to neoadjuvant chemotherapy, which is only available for cases with neoadjuvant chemotherapy.

2.4 Statistical analysis

All the analyses were performed in SAS - through a series of generic SAS macros [13]. SAS macro %DESCRIPTIVE produced frequencies and percentage, including the number of missing values, for categorical variables, and summary statistics, including mean, median, min, max, standard deviation, and the number of missingness for numerical variables. To identify the optimal cut-points for the time interval T3, we used R package 'CatPredi' to find two optimal cut-points based on a Cox proportional hazards model to maximize the discrimination power of predicting OS for the two cohorts separately. The analytic process includes five steps. The first step was to generate Kaplan Meier plots and summary statistic tables, which including median survival, log-rank p-value, and accrual survival rate for this time-to-event data set with SAS macro %KM_PLOT [13]. The second step was the descriptive analysis, which reassesses the composition of the whole target population and the boundaries that apply to the conclusion [13]. Also, descriptive analyses assisted in checking missing data, possible data entry errors, or outliers [13]. The SAS macro %DESCRIPTIVE helped yield summary statistics such as frequencies, percentages and means for all variables of interest and create a basic description of target population characteristics. The third step was to examine univariate associations between multiple variables and the three-level T3, the explanatory variable. SAS macro %UNI_CAT, column percentage for contingency tables was reported, along with p-value, from Pearson chi-square test for categorical covariates, and for numeric covariates, ANOVA. The fourth step was to examine univariate associations for each variable with overall survival. SAS macro %UNI_PHREG conducted this examination with Cox proportional hazards (PH) models for reporting

frequency, hazard ratio with 95% confidence interval, type-3 p-value and log-rank p-value of categorical covariates. The fifth step was conducting backward selection on a Cox PH model at the 0.05 significance level. SAS macro %PHREG_SEL was used to report frequency, hazard ratio with 95% confidence interval, type-3 p-value, and HR p-value of categorical covariates. To further control for potential confounding effects due to covariate imbalance, we created a propensity score weighted sample. The propensity score was calculated in SAS macro %CALC_PS, where the 3-level time interval was treated as the outcome and all covariates as predictors in a multinomial logistic regression. The absolute standardized difference (ASD) to check the covariate balance [14]. SAS macro %STD_DIFF was used to calculate the ASD for each covariate, and a benchmark to claim a sufficient balance is ASD < 0.1. We chose the average effect of treatment after stabilized inverse probability treatment weighting (IPTW) [15]. After weighting, we examined univariate associations for the time interval T3 and each covariate with overall survival based on the Cox PH model in the final weighted sample.

3. Results

Table 1 shows a target study population selection and exclusion. In the neoadjuvant chemo group, the total case number is 70,731, and in adjuvant chemo group, the number is 250,169. The two optimal cut points for T3 in the neoadjuvant cohort are 11.96 and 17.62 weeks, and we rounded them to 12 and 18 weeks; for the adjuvant chemo group, they are 8.69 and 11.23 weeks, and we rounded them to 9 and 12 weeks.

3.1 Neoadjuvant chemotherapy population

Figure 1 is the KM plot for the neoadjuvant chemo population which is divided into three time-interval groups based on the optimal cut-points before propensity score weighting. The survival curve decreases the fastest when patients received surgical procedure within 12 weeks from the start of chemotherapy, and the flattest downtrend of the curve appears when patients got surgery more than 18 weeks after the start of chemotherapy. The 5-year and 10-year survival rates for the time interval "< 12 weeks" are 78.8% (95% CI: 77.1%-80.4%) and 66.5% (95% CI: 64.0%-68.9%); for the time interval "12-18 weeks" are 82.1% (95% CI: 81.3%-82.8%) and 69.4% (95% CI: 68.0%-70.7%); for the time interval ">= 18 weeks" are 83.3% (95% CI: 82.9%-83.7%) and 71.0% (95% CI: 70.1%-71.9%).

Figure 2 is the KM plot for neoadjuvant chemo population after propensity score weighting. It can be seen the distance among three curves become closer, especially between time intervals "12-18 weeks" and ">18 weeks". This time the 5-year and 10-year survival rates for the time interval "< 12 weeks" are 80.2% (95% CI: 78.2%-82.0%) and 67.7% (95% CI: 64.9%-70.3%); for the time interval "12-18 weeks" are 82.4% (95% CI: 81.6%-83.2%) and 70.0% (95% CI: 68.6%-71.3%); for the time interval ">= 18 weeks" are 83.0% (95% CI: 82.5%-83.4%) and 70.5% (95% CI: 69.6%-71.4%).

Table 2 shows the univariate association between the clinical characteristics and the 3-level time interval for patients who got neoadjuvant chemo. The ASDs before and after the propensity score weighting is included in this table. In this population, 3,235 (4.6%) patients got surgery within 12 weeks from the start of chemo, 14,665 (20.7%) patients received surgery between 12 to 18 weeks from the start of chemo, and 52,831 (74.7%) patients received surgery after 18 weeks. Before IPTW adjustment, the ASDs of 9

covariates have values larger than 0.1, showing a sufficient imbalance in covariates distribution among the three intervals. The unbalanced covariates are type of the reporting facility, patient's primary insurance carrier, median household income at patient residence area, educational attainment at patient residence area, HER2, clinically-determined size or extension of the primary tumor, clinically-determined absence or presence of regional lymph node metastasis, year of cancer diagnosis and response to neoadjuvant therapy. After propensity score weighting, all the ASDs are smaller than 0.1, which shows balance among all covariates.

For the clinical characteristics in Table 2, an academic facility is more likely to delay the surgery (29.2% vs. 34.2% vs. 35.8%), while a comprehensive community cancer program may tend to start surgery earlier (47.3% vs. 43.1% vs. 40.4%). Patients with private insurance are more possible to have delayed surgery (64.1% vs. 65.0% vs. 70.15%), while patients with Medicare insurance are likely to have early surgery (21.2% vs. 20.6% vs. 14.4%). Patients who had higher median household income at their residence area waited more time before surgery (< $30,000: 14.0% vs. 11.8% vs. 10.7%; >=$46,000: 40.0% vs. 44.5% vs. 48.2%). Also, for patient who had HER2 positive (11.9% vs. 16.5%, vs. 27.4%) or AJCC N positive (43.7% vs. 44.2% vs. 52.3%) or positive response to neoadjuvant therapy (31.0% vs. 41.8% vs. 51.7%), there are bigger time gap in between neoadjuvant chemo and surgery. Over the years of diagnosis, there is a trend of surgery delay relative to neoadjuvant started (2004-2007: 28.5% vs. 18.7% vs. 11.1%; 2008-2011: 33.9% vs. 31.6% vs. 28.9%; 2012-2015: 37.6% vs. 49.7% vs. 60.0%).

Table 3 shows three association analyzes between overall survival and clinical characteristics based on Cox proportional hazard model: univariate association (UVA)

before weighting, univariate association (UVA) after weighting and multivariable association (MVA) before weighting. In MVA analysis, two variables were removed from the model: median household income at patient residence area and year of cancer diagnosis. The level with the smallest hazard was selected as the reference for each clinical characteristic. The hazard ratio with 95% confidence interval and p-value for each clinical characteristic were listed in the table for the neoadjuvant chemo population. Among all the three analyzes, the time interval with the smallest hazard is ">18 weeks", and the hazard for interval "< 12 weeks" is obviously larger than that for interval ">18 weeks" (HR=1.26 (95% CI: 1.16-1.36), p-value <0.001 for UVA before weighting; HR=1.17 (95% CI: 1.07-1.28), p-value <0.001 for UVA after weighting; HR=1.12 (95% CI: 1.04-1.21), p-value=0.005 for MVA before weighting), which indicates better overall survival for patients who received surgical procedure more than 18 weeks after the start of chemo compared to patients who received surgery less than 12 weeks from chemo start. For the interval "12-18 weeks", only UVA before weighting shows significant larger hazard than interval "< 12 weeks" (HR=1.08 (95% CI: 1.03-1.13), p-value=0.002), the other two analyzes show no difference in hazard (HR=1.02 (95% CI: 0.97-1.07), p-value=0.381 for UVA after weighting; HR=1.00 (95% CI: 0.96-1.05), p-value=0.930 for MVA before weighting). From the results, we can say 12 weeks after the start of chemo is an important time point for invasive breast cancer patients who received neoadjuvant chemotherapy. Operation after 12 weeks from the start of chemo can benefit the overall survival of patients. More specific hazard ratios and p-values can be seen in Table 3.

3.2 Adjuvant chemotherapy population

Figure 3 shows the KM plot for adjuvant chemo population which is separated by two optimal cut-points before propensity score weighting. The survival curve decreases the fastest when patients received chemotherapy more than 12 weeks after the surgery, and the flattest downtrend of the curve appears when patients got chemo within 9 weeks after the surgery. The 5-year and 10-year survival rates for the time interval "< 9 weeks" are 91.0% (95% CI: 90.8%-91.1%) and 80.2% (95% CI: 79.9%-80.5%); for the time interval "9-12 weeks" are 89.4% (95% CI: 88.9%-89.9%) and 77.5% (95% CI: 76.5%-78.5%); for the time interval ">= 12 weeks" are 86.7% (95% CI: 86.0%-87.4%) and 71.8% (95% CI: 70.4%-73.2%).

Figure 4 is the KM plot for adjuvant chemo population after propensity score weighting. It can be seen the distance among three curves become closer after weighting, but the difference is still obvious. The 5-year and 10-year survival rates for the time interval "< 9 weeks" are 90.9% (95% CI: 90.8%-91.0%) and 79.9% (95% CI: 79.6%-80.2%); for the time interval "9-12 weeks" are 89.7% (95% CI: 89.2%-90.2%) and 78.5% (95% CI: 77.4%-79.5%); for the time interval ">= 12 weeks" are 87.7% (95% CI: 87.0%-88.4%) and 74.3% (95% CI: 72.8%-75.7%).

Table 4 shows the frequency and percentage of patients in each level of each clinical characteristic and for patients in each separated interval of time between surgery and chemotherapy for population two. The ASDs before and after the propensity score weighting are also included in this table. In population two, 212,864 (85.1%) patients started chemo within 9 weeks after the surgery, 23,295 (9.3%) patients started chemo between 9 to 12 weeks from the end of surgery, and 14,010 (5.6%) patients received

chemo after 12 weeks from the surgical procedure. Before propensity score weighting with ATE_SW, the ASDs of 10 covariates have values larger than 0.1, showing insufficient balance of covariates among three intervals. The unbalanced covariates are type of the reporting facility, race, patient's primary insurance carrier, median household income at patient residence area, educational attainment at patient residence area, ER, HER2, Charlson-Deyo score, whether patient received radiation therapy, and year of cancer diagnosis. After propensity score weighting, all the ASDs are smaller than 0.1, which implies balance among all covariates.

For the clinical characteristics in Table 4, an academic program is more apt to choose the time interval “9-12 weeks” to start the chemotherapy (29.3% vs. 35.1% vs. 33.5%), while a comprehensive community cancer program may be more likely to start chemo earlier (46.7% vs. 41.4% vs. 42.2%). The white race tends to start chemo earlier (83.2% vs. 79.5% vs. 76.3%), while the black race is more possible to delay the chemotherapy (11.8% vs. 15.0% vs. 17.9%). Patients with private insurance are more possible to start chemo earlier (70.0% vs. 59.4% vs. 54.5%), and patients with Medicare insurance are likely to have postponed chemo (22.2% vs. 25.8% vs. 27.8%). Patients who had higher median household income at their residence area waited less time before chemo (< $30,000: 10.8% vs. 13.5% vs. 15.7%; >=$46,000: 46.7% vs. 43.2% vs. 40.1%). Patients who had higher educational attainment at their residence area also waited less time to start chemo (No high school degree >=29%: 13.6% vs. 17.1% vs. 20.4%; No high school degree <14%: 41.5% vs. 36.8% vs. 33.1%). Also, for patient who had ER positive (69.3% vs. 74.9%, vs. 73.5%) or HER2 negative (48.0% vs. 49.3% vs. 44.0%), there are bigger time gap in between surgery and adjuvant chemo. Patients who had radiation got

higher probability to start chemo earlier (Yes: 70.5% vs. 63.7% vs. 63.3%; No: 29.5% vs. 36.3% vs. 36.7%).

Table 5 shows three associations between overall survival and clinical characteristics based on Cox proportional hazard model: univariate association (UVA) before weighting, univariate association (UVA) after weighting and multivariate association (MVA) before weighting. In MVA analysis, no variables were removed from the model. The level with the smallest hazard was selected as the reference for each clinical characteristic. The hazard ratio with 95% confidence interval and p-value for each clinical characteristic were listed in the table for the adjuvant chemo population. Among all the three analyzes, the time interval with the smallest hazard is "<9 weeks", and the hazard for interval "9-12 weeks" is significantly larger than that for interval "<9 weeks" (HR=1.20 (95% CI: 1.15-1.25), p-value <0.001 for UVA before weighting; HR=1.13 (95% CI: 1.09-1.18), p-value <0.001 for UVA after weighting; HR=1.14 (95% CI: 1.09-1.19), p-value <0.001 for MVA before weighting), also the hazard for interval ">12 weeks" is significantly larger than that for interval "<9 weeks" (HR=1.53 (95% CI: 1.46-1.60), p-value <0.001 for UVA before weighting; HR=1.38 (95% CI: 1.31-1.44), p-value <0.001 for UVA after weighting; HR=1.39 (95% CI: 1.33-1.46), p-value <0.001 for MVA before weighting), which shows better overall survival for patients who started chemotherapy within 9 weeks after the surgery compared to patients who started chemo between 9-12 weeks from surgery, and to patients who started chemo more than 12 weeks from surgery. From the results, we can speculate that both 9 and 12 weeks after surgery are important time points for invasive breast cancer patients who received adjuvant chemotherapy. It seems

the earlier to start chemo, the better the overall survival, especially within 9 weeks from surgery. More specific hazard ratios and p-values can be seen in Table 5.

4. Discussion

Because this study is based on the observational database, the allocation of treatment, which is the time interval selection between chemotherapy and surgery here, is not randomized and can be influenced by the subject characteristics [16]. The systematic difference in baseline characteristics for different subjects should be considered in the analysis to avoid biased average treatment effect. We chose two methods: propensity score weighting and regression adjustment to eliminate the impact of potential confounding [16]. For the propensity score method, weighting by the inverse probability of treatment was used because this way does not result in much loss in the sample size as the propensity score matching does. After weighting, a synthetic sample was created and the distribution of the baseline characteristics was independent of the timing assignment, so the effect of timing between chemo and surgery on the overall survival can be estimated directly [16]. For the regression adjustment method, this is the most frequently used way for estimating the treatment effect in the observational study [16]. Both of methods effectively reduced the influence of confounding, and the results from these two methods are quite similar, which indicates the Cox PH model was correctly specified [16].

In the neoadjuvant chemotherapy group, the sample size (70,731) is much smaller than the adjuvant chemotherapy group (250,169), which is because adjuvant chemo is the

mainstream in invasive breast cancer treatment. The result indicates the start of surgery after 12 weeks from the start of chemo effectively benefits the overall survival of patients, and 18 weeks after the start of chemo can also count as an important time point. However, a limitation is that the end time of chemotherapy for each patient is unavailable from the database; we could only use the start time. The duration of chemotherapy varies by different agents used, so calculating time from the start of chemo may not explain the association between timing and survival credibly enough. Omarini (2017) demonstrated in his article that 3 weeks after the end of chemo was an important turning point, for the overall survival of patients who got surgery within 3 weeks from the end of chemo was significantly better than that of other patients (HR=3.1, p-value=0.030) [6]. Our sample size was much larger than that in Omarini's study (70,731 vs. 319), which means more credible results, but the uncertainty of the chemo duration impeded the further precision of our study.

In the adjuvant chemotherapy group, the result is more convincing. Starting chemo within 9 weeks from the surgery most contributing to prolonged survival, while receiving the therapy later than 12 weeks after the surgery apparently lower the survival rate. More attention and study are needed for patients who receive neoadjuvant chemotherapy as breast cancer treatment.

5. Conclusion

The importance of timing between chemotherapy and surgical procedures to the overall survival among breast cancer patients is quite obviously nowadays. For the neoadjuvant

chemo population, patients treated later than 12 weeks from the start of chemo experienced better overall survival independent from other clinical characteristics. More efforts should be made to study the impact of this time interval to overall survival for this population in order to maximize the benefits of previous treatments. For the adjuvant chemo population, the highest survival rate appeared when patients got treated within 9 weeks after surgery. Longer waiting time to chemo may reduce the probability of survival, so the postponed chemotherapy should be avoided to improve the treatment effect. Further studies are needed in this area for the well-being of breast cancer patients.

Reference

1.Breast Cancer. Mayo Clinic.

https://www.mayoclinic.org/diseases-conditions/breast-cancer/symptoms-causes/syc-20352470. Accessed January, 2020

2. How Common Is Breast Cancer? American Cancer Society.

https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html. Accessed January, 2020

3. Chemotherapy for breast cancer. Mayo Clinic.

https://www.mayoclinic.org/tests-procedures/chemotherapy-for-breast-cancer/about/pac-20384931. Accessed January, 2020

4. Chemotherapy for breast cancer. American Cancer Society.

http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-treating-chemotherapy. Accessed January, 2020.

5. Kaufmann, M., et al. (2012). "Recommendations from an International Consensus Conference on the Current Status and Future of Neoadjuvant Systemic Therapy in Primary Breast Cancer." Annals of Surgical Oncology 19(5): 1508-1516.

6. Omarini, C., et al. (2017). "Impact of time to surgery after neoadjuvant chemotherapy in operable breast cancer patients." European Journal of Surgical Oncology 43(4): 613-618.

7. Arciero, C., et al. "Time from Completion of Neo-adjuvant Chemotherapy to Surgery: Effects on Outcomes in Breast Cancer Patients." The Breast Journal

8. Mieog, J. S. D., et al. (2007). "Neoadjuvant chemotherapy for operable breast cancer." BJS (British Journal of Surgery) 94(10): 1189-1200.

9. Kühn, T., et al. (2018). "Current Status and Future Perspectives of Axillary Management in the Neoadjuvant Setting." Breast care (Basel, Switzerland) 13(5): 337-341.

10. Farazmand, B., et al. (2017). "Optimal Timing of Surgery after Neoadjuvant Chemotherapy of Breast Cancer." Archives of Breast Cancer 4: 1-3.

11. Boffa DJ, Rosen JE, Mallin K, et al. Using the National Cancer Database for Outcomes Research: A Review. JAMA Oncol. 2017;3(12):1722-1728.

12. Irantzu Barrio and Maria Xose Rodriguez-Alvarez (2017). CatPredi: Optimal Categorisation of Continuous Variables in Prediction Models. R package version 1.1. https://CRAN.R-project.org/package=CatPredi

13. Liu, Y., et al. (2018). "Carrying out streamlined routine data analyses with reports for observational studies: introduction to a series of generic SAS (®) macros." F1000Research 7: 1955-1955.

14. Peter C. Austin (2009) Using the Standardized Difference to Compare the Prevalence of a Binary Variable Between Two Groups in Observational Research, Communications in Statistics - Simulation and Computation, 38:6, 1228-1234, DOI: 10.1080/03610910902859574

15. Ruth C, Brownell M, Isbister J, MacWilliam L, Gammon H, Singal D, Soodeen R, McGowan K, Kulbaba C, Boriskewich E. Long-Term Outcomes of Manitoba's Insight

Mentoring Program: A Comparative Statistical Analysis. Winnipeg, MB: Manitoba Centre for Health Policy, 2015.

16. Austin, P. C. (2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." Multivariate behavioral research 46(3): 399-424.

FIGURES AND TABLES

Table1. Target population selection

Selection and Exclusion Criteria	Sample Size	Excluded
NCDB Breast 2016 PUF Cancer Cases	2696734	-
Exclude previous or concurrent malignancy	2224731	472003
Exclude cases were not treated at reporting facility	2134136	90595
Include Behavior = Invasive	1727949	406187
Exclude Male Patients	1711594	16355
Include cases with positive histology diagnostic confirmation	1699888	11706
Include Patients with Clinical Stage T1-3N0-1M0	952998	746890
Include patients had both Surgery and Chemotherapy with known sequence	372863	580135
Include patients had adjuvant radiation, if any	368671	4192
Exclude missing facility type	331566	37105
Exclude missing median income quartiles	320984	10582
Exclude missing percent no high school degree quartiles	320947	37
Exclude missing radiation therapy	320937	10
Exclude missing outcome	320900	37

Table 2. Frequencies, percentages and absolute standardized differences in each level of each clinical characteristic for neoadjuvant chemotherapy group

			time interval between surgery and chemotherapy				
Clinical Characteristics	**Level**	**Total N (%) 70731(100.0)**	**< 12 weeks N=3235**	**12-18 weeks N=14665**	**> 18 weeks N=52831**	**ASD before**	**ASD after**
Chemo-surgery sequence	Neoadjuvant Chemo	70731 (100.0)				-	-
Facility Type	Community or Integrated Network Cancer Program	16648 (23.5)	760 (23.49)	3330 (22.71)	12558 (23.77)	0.025	0.024
	Comprehensive Community Cancer Program	29184 (41.3)	1531 (47.33)	6317 (43.08)	21336 (40.39)	**0.14**	0.012
	Academic/Research Program	24899 (35.2)	944 (29.18)	5018 (34.22)	18937 (35.84)	**0.143**	0.03
Race	White	55603 (78.6)	2571 (79.47)	11680 (79.65)	41352 (78.27)	0.034	0.023
	Black	11081 (15.7)	485 (14.99)	2232 (15.22)	8364 (15.83)	0.023	0.019
	Others	4047 (5.7)	179 (5.53)	753 (5.13)	3115 (5.9)	0.034	0.014
Primary Payor	Not Insured/Unknown/ Medicaid/Other Government	10747 (15.2)	477 (14.74)	2115 (14.42)	8155 (15.44)	0.029	0.004
	Private	48664 (68.8)	2072 (64.05)	9533 (65.01)	37059 (70.15)	**0.13**	0.013
	Medicare	11320 (16.0)	686 (21.21)	3017 (20.57)	7617 (14.42)	**0.178**	0.013

Clinical Characteristics	Level	Total N (%) 70731(100.0)	time interval between surgery and chemotherapy			ASD before	ASD after
			< 12 weeks N=3235	12-18 weeks N=14665	> 18 weeks N=52831		
Median Income Quartiles 2000	< $30,000	7811 (11.0)	454 (14.03)	1725 (11.76)	5632 (10.66)	**0.103**	0.014
	$30,000 - $34,999	10817 (15.3)	558 (17.25)	2385 (16.26)	7874 (14.9)	0.064	0.014
	$35,000 - $45,999	18852 (26.7)	929 (28.72)	4034 (27.51)	13889 (26.29)	0.054	0.006
	>=$46,000	33251 (47.0)	1294 (40)	6521 (44.47)	25436 (48.15)	**0.165**	0.02
Percent No High School Degree Quartiles 2000	>=29%	10568 (14.9)	575 (17.77)	2259 (15.4)	7734 (14.64)	0.085	0.033
	20-28.9%	15118 (21.4)	761 (23.52)	3279 (22.36)	11078 (20.97)	0.061	0.003
	14-19.9%	15713 (22.2)	704 (21.76)	3209 (21.88)	11800 (22.34)	0.014	0.003
	< 14%	29332 (41.5)	1195 (36.94)	5918 (40.35)	22219 (42.06)	**0.105**	0.02
Grade	Well Differentiated	3938 (5.6)	210 (6.49)	1022 (6.97)	2706 (5.12)	0.078	0.012
	Moderately Differentiated	23118 (32.7)	988 (30.54)	4964 (33.85)	17166 (32.49)	0.071	0.003
	Poorly Differentiated/Undifferentiated	38525 (54.5)	1837 (56.79)	7758 (52.9)	28930 (54.76)	0.078	0.011
	Cell Type Not Determined	5150 (7.3)	200 (6.18)	921 (6.28)	4029 (7.63)	0.057	0.009
ER	Negative	29259 (41.4)	1380 (42.66)	5773 (39.37)	22106 (41.84)	0.067	0.013
	Positive	40747 (57.6)	1797 (55.55)	8719 (59.45)	30231 (57.22)	0.079	0.016
	Unknown	725 (1.0)	58 (1.79)	173 (1.18)	494 (0.94)	0.073	0.012

Clinical Characteristics	Level	Total N (%) 70731(100.0)	time interval between surgery and chemotherapy			ASD before	ASD after
			< 12 weeks N=3235	12-18 weeks N=14665	> 18 weeks N=52831		
PR	Negative	36520 (51.6)	1672 (51.68)	7308 (49.83)	27540 (52.13)	0.046	0.016
	Positive	33360 (47.2)	1504 (46.49)	7142 (48.7)	24714 (46.78)	0.044	0.021
	Unknown	851 (1.2)	59 (1.82)	215 (1.47)	577 (1.09)	0.061	0.02
HER2	Negative	33096 (46.8)	1332 (41.17)	7050 (48.07)	24714 (46.78)	**0.139**	0.007
	Positive	17351 (24.5)	385 (11.9)	2418 (16.49)	14548 (27.54)	**0.401**	0.01
	Unknown	20284 (28.7)	1518 (46.92)	5197 (35.44)	13569 (25.68)	**0.453**	0.012
Charlson-Deyo Score	0	62338 (88.1)	2788 (86.18)	12745 (86.91)	46805 (88.59)	0.073	0.022
	1	7044 (10.0)	369 (11.41)	1535 (10.47)	5140 (9.73)	0.055	0.017
	>=2	1349 (1.9)	78 (2.41)	385 (2.63)	886 (1.68)	0.065	0.014
TNM_CLIN_T_1[1]	c1	14149 (20.0)	589 (18.21)	3321 (22.65)	10239 (19.38)	**0.11**	0.005
	c2	40399 (57.1)	1858 (57.43)	8219 (56.05)	30322 (57.39)	0.028	0.015
	c3	16183 (22.9)	788 (24.36)	3125 (21.31)	12270 (23.23)	0.073	0.013
TNM_CLIN_N_1[2]	c0	35195 (49.8)	1821 (56.29)	8190 (55.85)	25184 (47.67)	**0.173**	0.028
	c1	35536 (50.2)	1414 (43.71)	6475 (44.15)	27647 (52.33)	**0.173**	0.028
Radiation	No	19119 (27.0)	932 (28.81)	4239 (28.91)	13948 (26.4)	0.056	0.003
	Yes	51612 (73.0)	2303 (71.19)	10426 (71.09)	38883 (73.6)	0.056	0.003

[1] Clinically determined size or extension of the primary tumor
[2] Clinically determined absence or presence of regional lymph node metastasis

Clinical Characteristics	Level	Total N (%) 70731(100.0)	time interval between surgery and chemotherapy			ASD before	ASD after
			< 12 weeks N=3235	12-18 weeks N=14665	> 18 weeks N=52831		
Surgical Margin	Negative	66825 (94.5)	3056 (94.47)	13750 (93.76)	50019 (94.68)	0.039	0.011
	Positive	2955 (4.2)	145 (4.48)	715 (4.88)	2095 (3.97)	0.044	0.012
	Unknown	951 (1.3)	34 (1.05)	200 (1.36)	717 (1.36)	0.028	0.002
Year of Diagnosis	2004-2007	9542 (13.5)	922 (28.5)	2746 (18.72)	5874 (11.12)	**0.447**	0.005
	2008-2011	21000 (29.7)	1097 (33.91)	4627 (31.55)	15276 (28.91)	**0.108**	0.003
	2012-2015	40189 (56.8)	1216 (37.59)	7292 (49.72)	31681 (59.97)	**0.459**	0.006
Response to Neoadjuvant Therapy	Yes	34462 (48.7)	1002 (30.97)	6130 (41.8)	27330 (51.73)	**0.431**	0.001
	No	2398 (3.4)	187 (5.78)	672 (4.58)	1539 (2.91)	**0.141**	0.007
	Missing	18389 (26.0)	1454 (44.95)	4817 (32.85)	12118 (22.94)	**0.478**	0.01
	Unknown	15482 (21.9)	592 (18.3)	3046 (20.77)	11844 (22.42)	**0.102**	0.009
Weeks in between Chemo and Surgery	Mean	20.59				-	-
	Median	20.57				-	-
	Min	8.00				-	-
	Max	36.00				-	-
	Std Dev	4.76				-	-
Age at Diagnosis	Mean	54.32	55.49	55.52	53.91	-	-
	Median	53.00	54	54	53	-	-
	Min	40.00	40	40	40	-	-
	Max	90.00	90	90	90	-	-
	Std Dev	9.36	10.2	10.17	9.03	-	-

Table 3. Hazard ratios with 95% CI and p-values from UVA before weighting, UVA after weighting and MVA based on Cox PH model for neoadjuvant chemotherapy group

		UVA before weighting		**UVA after weighting**		**MVA before weighting**	
Clinical Characteristics	**Level**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**
two optimal cut points	< 12 weeks	1.26 (1.16-1.36)	**<.001**	1.17 (1.07-1.28)	**<.001**	1.12 (1.04-1.21)	**0.005**
	12-18 weeks	1.08 (1.03-1.13)	**0.002**	1.02 (0.97-1.07)	0.381	1.00 (0.96-1.05)	0.930
	>= 18 weeks	-	-	-	-	-	-
Facility Type	Community or Integrated Network Cancer Program	1.03 (0.97-1.08)	0.346	1.02 (0.97-1.08)	0.405	1.04 (0.99-1.10)	0.128
	Comprehensive Community Cancer Program	1.08 (1.03-1.13)	**<.001**	1.08 (1.03-1.13)	**0.001**	1.12 (1.07-1.18)	**<.001**
	Academic/ Research Program	-	-	-	-	-	-
Race	Black	1.51 (1.43-1.58)	**<.001**	1.51 (1.44-1.59)	**<.001**	1.21 (1.15-1.28)	**<.001**
	Others	0.80 (0.72-0.88)	**<.001**	0.78 (0.70-0.86)	**<.001**	0.82 (0.74-0.91)	**<.001**
	White	-	-	-	-	-	-

		UVA before weighting		UVA after weighting		MVA before weighting	
Clinical Characteristics	**Level**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**
Primary Payor	Not Insured/Unknown/Medicaid/Other Government	1.47 (1.40-1.55)	**<.001**	1.46 (1.39-1.54)	**<.001**	1.29 (1.22-1.36)	**<.001**
	Medicare	1.85 (1.76-1.95)	**<.001**	1.85 (1.76-1.94)	**<.001**	1.25 (1.17-1.33)	**<.001**
	Private	-	-	-	-	-	-
Median Income Quartiles 2000	< $30,000	1.40 (1.31-1.49)	**<.001**	1.41 (1.32-1.50)	**<.001**		-
	$30,000 - $34,999	1.29 (1.22-1.36)	**<.001**	1.28 (1.21-1.36)	**<.001**		-
	$35,000 - $45,999	1.20 (1.15-1.26)	**<.001**	1.22 (1.16-1.28)	**<.001**		-
	>=$46,000	-	-	-	-		-
Percent No High School Degree Quartiles 2000	>=29%	1.39 (1.31-1.48)	**<.001**	1.40 (1.32-1.48)	**<.001**	1.13 (1.06-1.20)	**<.001**
	20-28.9%	1.35 (1.28-1.42)	**<.001**	1.36 (1.29-1.43)	**<.001**	1.17 (1.11-1.24)	**<.001**
	14-19.9%	1.16 (1.10-1.22)	**<.001**	1.16 (1.10-1.22)	**<.001**	1.07 (1.01-1.13)	**0.014**
	< 14%	-	-	-	-	-	-
Grade	Moderately Differentiated	1.36 (1.22-1.51)	**<.001**	1.31 (1.18-1.45)	**<.001**	1.30 (1.17-1.45)	**<.001**
	Poorly Differentiated/Undifferentiated	2.02 (1.82-2.25)	**<.001**	1.95 (1.76-2.15)	**<.001**	1.64 (1.47-1.83)	**<.001**
	Cell Type Not Determined	1.46 (1.28-1.65)	**<.001**	1.39 (1.23-1.58)	**<.001**	1.28 (1.12-1.45)	**<.001**
	Well Differentiated	-	-	-	-	-	-

		UVA before weighting		UVA after weighting		MVA before weighting	
Clinical Characteristics	**Level**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**
ER	Negative	1.65 (1.59-1.72)	**<.001**	1.65 (1.59-1.72)	**<.001**	1.20 (1.13-1.27)	**<.001**
	Unknown	1.22 (1.04-1.43)	**0.017**	1.19 (1.01-1.40)	**0.037**	1.06 (0.80-1.42)	0.682
	Positive	-	-	-	-	-	-
PR	Negative	1.72 (1.65-1.79)	**<.001**	1.72 (1.65-1.79)	**<.001**	1.39 (1.31-1.48)	**<.001**
	Unknown	1.30 (1.12-1.51)	**<.001**	1.30 (1.12-1.51)	**<.001**	1.23 (0.94-1.61)	0.134
	Positive	-	-	-	-	-	-
HER2	Negative	2.16 (2.01-2.32)	**<.001**	2.12 (1.97-2.27)	**<.001**	2.05 (1.90-2.20)	**<.001**
	Unknown	1.82 (1.69-1.96)	**<.001**	1.78 (1.65-1.91)	**<.001**	1.57 (1.35-1.83)	**<.001**
	Positive	-	-	-	-	-	-
Charlson-Deyo Score	1	1.51 (1.42-1.60)	**<.001**	1.50 (1.42-1.60)	**<.001**	1.28 (1.20-1.36)	**<.001**
	>=2	2.33 (2.08-2.61)	**<.001**	2.32 (2.07-2.59)	**<.001**	1.73 (1.54-1.94)	**<.001**
	0	-	-	-	-	-	-
TNM_CLIN_T_1[1]	c2	1.34 (1.26-1.43)	**<.001**	1.31 (1.24-1.39)	**<.001**	1.25 (1.18-1.33)	**<.001**
	c3	2.00 (1.88-2.13)	**<.001**	1.96 (1.84-2.09)	**<.001**	1.77 (1.66-1.89)	**<.001**
	c1	-	-	-	-	-	-
TNM_CLIN_N_1[2]	c1	1.66 (1.60-1.73)	**<.001**	1.65 (1.58-1.72)	**<.001**	1.62 (1.55-1.69)	**<.001**
	c0	-	-	-	-	-	-

[1] Clinically determined size or extension of the primary tumor
[2] Clinically determined absence or presence of regional lymph node metastasis

		UVA before weighting		UVA after weighting		MVA before weighting	
Clinical Characteristics	**Level**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**
Radiation	Yes	1.03 (0.99-1.08)	0.164	1.04 (0.99-1.08)	0.133	0.91 (0.87-0.95)	**<.001**
	No	-	-	-	-	-	-
Surgical Margin	Positive	1.51 (1.39-1.63)	**<.001**	1.49 (1.37-1.61)	**<.001**	1.52 (1.40-1.65)	**<.001**
	Unknown	1.15 (0.98-1.34)	0.078	1.16 (1.00-1.35)	**0.049**	1.12 (0.96-1.30)	0.156
	Negative	-	-	-	-	-	-
Year of Diagnosis	2004-2007	1.05 (0.99-1.11)	0.123	1.04 (0.98-1.10)	0.176		-
	2008-2011	1.01 (0.97-1.07)	0.568	1.02 (0.97-1.07)	0.534		-
	2012-2015	-	-	-	-		-
Response to Neoadjuvant Therapy	No	3.10 (2.84-3.39)	**<.001**	3.17 (2.91-3.46)	**<.001**	2.84 (2.60-3.11)	**<.001**
	Missing	1.13 (1.07-1.19)	**<.001**	1.12 (1.07-1.18)	**<.001**	1.27 (1.10-1.46)	**<.001**
	Unknown	1.05 (0.98-1.11)	0.142	1.05 (0.99-1.11)	0.139	1.10 (1.03-1.17)	**0.003**
	Yes	-	-	-	-	-	-
Age at Diagnosis		1.03 (1.03-1.03)	**<.001**	1.03 (1.03-1.03)	**<.001**	1.02 (1.02-1.02)	**<.001**

Table 4. Frequencies, percentages and absolute standardized differences in each level of each clinical characteristic for adjuvant chemotherapy group

			time interval between surgery and chemotherapy				
Clinical Characteristics	**Level**	**Total N (%) 250169(100.0)**	**< 9 weeks N=212864**	**9-12 weeks N=23295**	**> 12 weeks N=14010**	**ASD before**	**ASD after**
Chemo-surgery sequence	Adjuvant Chemo	250169 (100.0)				-	-
Facility Type	Community or Integrated Network Cancer Program	60041 (24.0)	51155 (24.03)	5480 (23.52)	3406 (24.31)	0.019	0.007
	Comprehensive Community Cancer Program	114963 (46.0)	99408 (46.7)	9647 (41.41)	5908 (42.17)	**0.107**	0.004
	Academic/Research Program	75165 (30.0)	62301 (29.27)	8168 (35.06)	4696 (33.52)	**0.124**	0.01
Race	White	206312 (82.5)	177106 (83.2)	18515 (79.48)	10691 (76.31)	**0.172**	0.008
	Black	31175 (12.5)	25180 (11.83)	3494 (15)	2501 (17.85)	**0.17**	0.01
	Others	12682 (5.1)	10578 (4.97)	1286 (5.52)	818 (5.84)	0.038	0.003
Primary Payor	Not Insured/Unknown/Medicaid/Other Government	28945 (11.6)	23023 (10.82)	3445 (14.79)	2477 (17.68)	**0.197**	0.008
	Private	164006 (65.6)	142533 (66.96)	13832 (59.38)	7641 (54.54)	**0.256**	0.015
	Medicare	57218 (22.9)	47308 (22.22)	6018 (25.83)	3892 (27.78)	**0.129**	0.012

Clinical Characteristics	Level	Total N (%) 250169(100.0)	time interval between surgery and chemotherapy			ASD before	ASD after
			< 9 weeks N=212864	9-12 weeks N=23295	> 12 weeks N=14010		
Median Income Quartiles 2000	< $30,000	28297 (11.3)	22953 (10.78)	3152 (13.53)	2192 (15.65)	**0.144**	0.011
	$30,000 - $34,999	39982 (16.0)	33634 (15.8)	3875 (16.63)	2473 (17.65)	0.05	0.007
	$35,000 - $45,999	66730 (26.7)	56803 (26.69)	6204 (26.63)	3723 (26.57)	0.003	0.008
	>=$46,000	115160 (46.0)	99474 (46.73)	10064 (43.2)	5622 (40.13)	**0.133**	0.005
Percent No High School Degree Quartiles 2000	>=29%	35876 (14.3)	29042 (13.64)	3982 (17.09)	2852 (20.36)	**0.18**	0.014
	20-28.9%	53777 (21.5)	45038 (21.16)	5339 (22.92)	3400 (24.27)	0.074	0.004
	14-19.9%	58926 (23.6)	50399 (23.68)	5404 (23.2)	3123 (22.29)	0.033	0.004
	< 14%	101590 (40.6)	88385 (41.52)	8570 (36.79)	4635 (33.08)	**0.175**	0.01
Grade	Well Differentiated	23909 (9.6)	19855 (9.33)	2496 (10.71)	1558 (11.12)	0.059	0.007
	Moderately Differentiated	95962 (38.4)	80938 (38.02)	9462 (40.62)	5562 (39.7)	0.053	0.007
	Poorly Differentiated/Undifferentiated	120480 (48.2)	103918 (48.82)	10341 (44.39)	6221 (44.4)	0.089	0.012
	Cell Type Not Determined	9818 (3.9)	8153 (3.83)	996 (4.28)	669 (4.78)	0.047	0.004
ER	Negative	72528 (29.0)	63393 (29.78)	5625 (24.15)	3510 (25.05)	**0.127**	0.009
	Positive	175281 (70.1)	147528 (69.31)	17457 (74.94)	10296 (73.49)	**0.126**	0.008
	Unknown	2360 (0.9)	1943 (0.91)	213 (0.91)	204 (1.46)	0.051	0.008

Clinical Characteristics	Level	Total N (%) 250169(100.0)	time interval between surgery and chemotherapy			ASD before	ASD after
			< 9 weeks N=212864	9-12 weeks N=23295	> 12 weeks N=14010		
PR	Negative	99135 (39.6)	85410 (40.12)	8507 (36.52)	5218 (37.24)	0.074	0.008
	Positive	148018 (59.2)	124990 (58.72)	14513 (62.3)	8515 (60.78)	0.073	0.007
	Unknown	3016 (1.2)	2464 (1.16)	275 (1.18)	277 (1.98)	0.066	0.01
HER2	Negative	119805 (47.9)	102174 (48)	11474 (49.26)	6157 (43.95)	**0.107**	0.01
	Positive	35473 (14.2)	30724 (14.43)	2949 (12.66)	1800 (12.85)	0.052	0.006
	Unknown	94891 (37.9)	79966 (37.57)	8872 (38.09)	6053 (43.2)	**0.115**	0.012
Charlson-Deyo Score	0	212476 (84.9)	181891 (85.45)	19251 (82.64)	11334 (80.9)	**0.122**	0.007
	1	31378 (12.5)	25987 (12.21)	3264 (14.01)	2127 (15.18)	0.086	0.005
	>=2	6315 (2.5)	4986 (2.34)	780 (3.35)	549 (3.92)	0.091	0.004
TNM_CLIN_T_1[1]	c1	148955 (59.5)	126440 (59.4)	14097 (60.52)	8418 (60.09)	0.023	0.006
	c2	91833 (36.7)	78364 (36.81)	8405 (36.08)	5064 (36.15)	0.015	0.006
	c3	9381 (3.7)	8060 (3.79)	793 (3.4)	528 (3.77)	0.021	0.004
TNM_CLIN_N_1[2]	c0	208868 (83.5)	177032 (83.17)	19918 (85.5)	11918 (85.07)	0.064	0.004
	c1	41301 (16.5)	35832 (16.83)	3377 (14.5)	2092 (14.93)	0.064	0.004
Radiation	No	76325 (30.5)	62725 (29.47)	8456 (36.3)	5144 (36.72)	**0.155**	0.016
	Yes	173844 (69.5)	150139 (70.53)	14839 (63.7)	8866 (63.28)	**0.155**	0.016

[1] Clinically determined size or extension of the primary tumor
[2] Clinically determined absence or presence of regional lymph node metastasis

Clinical Characteristics	Level	Total N (%) 250169(100.0)	time interval between surgery and chemotherapy			ASD before	ASD after
			< 9 weeks N=212864	9-12 weeks N=23295	> 12 weeks N=14010		
Surgical Margin	Negative	238128 (95.2)	202826 (95.28)	22168 (95.16)	13134 (93.75)	0.067	0.006
	Positive	9961 (4.0)	8310 (3.9)	938 (4.03)	713 (5.09)	0.057	0.005
	Unknown	2080 (0.8)	1728 (0.81)	189 (0.81)	163 (1.16)	0.035	0.017
Year of Diagnosis	2004-2007	41802 (16.7)	35552 (16.7)	3610 (15.5)	2640 (18.84)	0.089	0.01
	2008-2011	100733 (40.3)	85116 (39.99)	9617 (41.28)	6000 (42.83)	0.058	0.004
	2012-2015	107634 (43.0)	92196 (43.31)	10068 (43.22)	5370 (38.33)	**0.101**	0.011
Weeks in between Chemo and Surgery	Mean	6.32				-	-
	Median	5.71				-	-
	Min	0.00				-	-
	Max	35.86				-	-
	Std Dev	3.28				-	-
Age at Diagnosis	Mean	56.79	56.61	57.66	58.05	-	-
	Median	56.00	56	57	58	-	-
	Min	40.00	40	40	40	-	-
	Max	90.00	90	90	90	-	-
	Std Dev	9.80	9.77	9.78	10.09	-	-

Table 5. Hazard ratios with 95% CI and p-values from UVA before weighting, UVA after weighting and MVA based on Cox PH model for adjuvant chemotherapy group

		UVA before weighting		**UVA after weighting**		**MVA before weighting**	
Clinical Characteristics	**Level**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**
two optimal cut points	>= 12 weeks	1.53 (1.46-1.60)	**<.001**	1.38 (1.31-1.44)	**<.001**	1.39 (1.33-1.46)	**<.001**
	9-12 weeks	1.20 (1.15-1.25)	**<.001**	1.13 (1.09-1.18)	**<.001**	1.14 (1.09-1.19)	**<.001**
	< 9 weeks	-	-	-	-	-	-
Facility Type	Community or Integrated Network Cancer Program	1.20 (1.16-1.24)	**<.001**	1.19 (1.15-1.24)	**<.001**	1.13 (1.09-1.17)	**<.001**
	Comprehensive Community Cancer Program	1.16 (1.12-1.19)	**<.001**	1.16 (1.13-1.20)	**<.001**	1.11 (1.08-1.14)	**<.001**
	Academic/Research Program	-	-	-	-	-	-
Race	Black	1.42 (1.37-1.47)	**<.001**	1.41 (1.37-1.46)	**<.001**	1.16 (1.12-1.21)	**<.001**
	Others	0.68 (0.64-0.73)	**<.001**	0.68 (0.63-0.73)	**<.001**	0.74 (0.69-0.80)	**<.001**
	White	-	-	-	-	-	-
Primary Payor	Not Insured/Unknown/Medicaid/Other Government	1.67 (1.61-1.74)	**<.001**	1.65 (1.58-1.71)	**<.001**	1.38 (1.32-1.43)	**<.001**
	Medicare	2.67 (2.60-2.74)	**<.001**	2.65 (2.58-2.72)	**<.001**	1.37 (1.33-1.42)	**<.001**
	Private	-	-	-	-	-	-

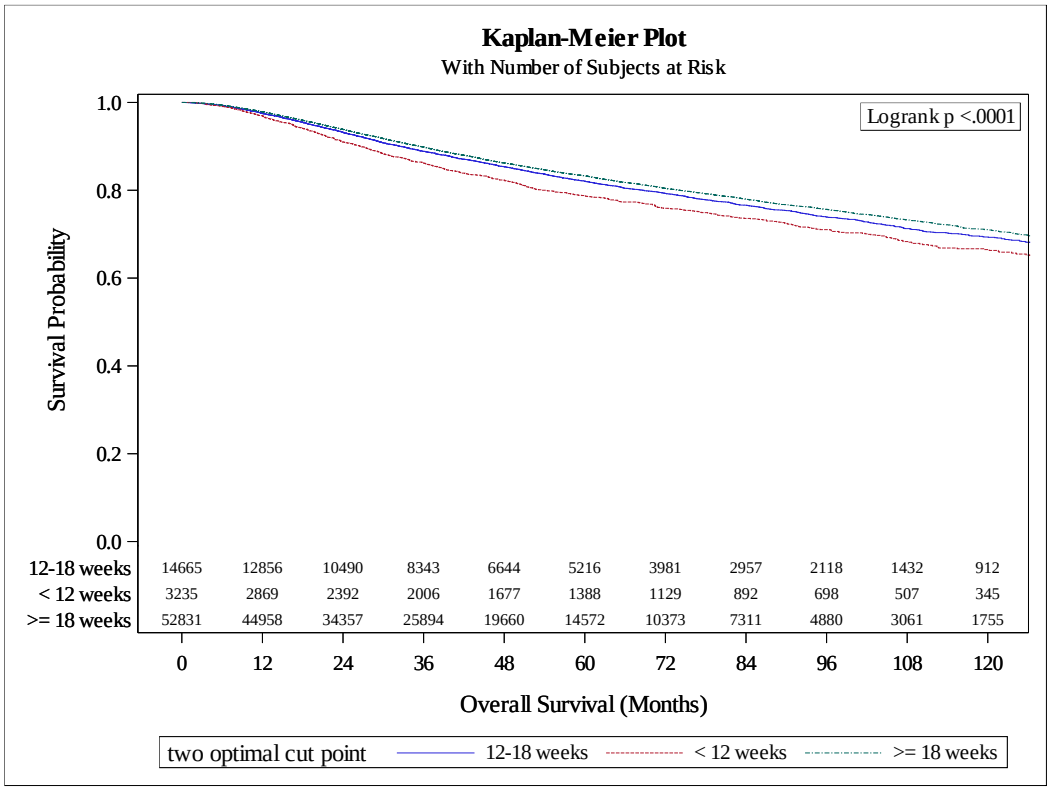
		UVA before weighting		UVA after weighting		MVA before weighting	
Clinical Characteristics	**Level**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**
Median Income Quartiles 2000	< $30,000	1.63 (1.57-1.70)	**<.001**	1.62 (1.56-1.68)	**<.001**	1.18 (1.13-1.25)	**<.001**
	$30,000 - $34,999	1.48 (1.43-1.53)	**<.001**	1.47 (1.42-1.52)	**<.001**	1.16 (1.11-1.21)	**<.001**
	$35,000 - $45,999	1.29 (1.25-1.33)	**<.001**	1.29 (1.25-1.33)	**<.001**	1.08 (1.05-1.12)	**<.001**
	>=$46,000	-	-	-	-	-	-
Percent No High School Degree Quartiles 2000	>=29%	1.55 (1.49-1.60)	**<.001**	1.53 (1.48-1.59)	**<.001**	1.05 (1.00-1.11)	**0.040**
	20-28.9%	1.42 (1.37-1.47)	**<.001**	1.41 (1.37-1.46)	**<.001**	1.09 (1.05-1.14)	**<.001**
	14-19.9%	1.30 (1.26-1.34)	**<.001**	1.30 (1.25-1.34)	**<.001**	1.11 (1.07-1.15)	**<.001**
	< 14%	-	-	-	-	-	-
Grade	Moderately Differentiated	1.34 (1.27-1.41)	**<.001**	1.34 (1.27-1.41)	**<.001**	1.21 (1.14-1.27)	**<.001**
	Poorly Differentiated/Undifferentiated	2.10 (2.00-2.21)	**<.001**	2.11 (2.00-2.22)	**<.001**	1.57 (1.49-1.66)	**<.001**
	Cell Type Not Determined	1.72 (1.59-1.86)	**<.001**	1.72 (1.59-1.86)	**<.001**	1.38 (1.28-1.50)	**<.001**
	Well Differentiated	-	-	-	-	-	-
ER	Negative	1.69 (1.65-1.73)	**<.001**	1.70 (1.66-1.74)	**<.001**	1.18 (1.14-1.22)	**<.001**
	Unknown	1.45 (1.32-1.60)	**<.001**	1.43 (1.30-1.58)	**<.001**	1.08 (0.92-1.27)	0.337
	Positive	-	-	-	-	-	-

		UVA before weighting		UVA after weighting		MVA before weighting	
Clinical Characteristics	**Level**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**
PR	Negative	1.71 (1.67-1.75)	**<.001**	1.72 (1.68-1.76)	**<.001**	1.23 (1.19-1.28)	**<.001**
	Unknown	1.52 (1.40-1.66)	**<.001**	1.50 (1.37-1.64)	**<.001**	1.17 (1.02-1.35)	**0.028**
	Positive	-	-	-	-	-	-
HER2	Negative	1.28 (1.22-1.34)	**<.001**	1.29 (1.22-1.35)	**<.001**	1.39 (1.32-1.46)	**<.001**
	Unknown	1.16 (1.11-1.22)	**<.001**	1.16 (1.11-1.22)	**<.001**	1.32 (1.25-1.39)	**<.001**
	Positive	-	-	-	-	-	-
Charlson-Deyo Score	1	1.81 (1.75-1.87)	**<.001**	1.81 (1.75-1.87)	**<.001**	1.41 (1.36-1.45)	**<.001**
	>=2	3.24 (3.07-3.41)	**<.001**	3.20 (3.03-3.37)	**<.001**	2.13 (2.01-2.24)	**<.001**
	0	-	-	-	-	-	-
TNM_CLIN_T_1[1]	c2	1.82 (1.77-1.86)	**<.001**	1.83 (1.78-1.88)	**<.001**	1.57 (1.53-1.61)	**<.001**
	c3	2.76 (2.63-2.91)	**<.001**	2.78 (2.65-2.93)	**<.001**	2.31 (2.20-2.43)	**<.001**
	c1	-	-	-	-	-	-
TNM_CLIN_N_1[2]	c1	1.64 (1.59-1.68)	**<.001**	1.65 (1.60-1.69)	**<.001**	1.43 (1.39-1.47)	**<.001**
	c0	-	-	-	-	-	-
Radiation	No	1.32 (1.29-1.36)	**<.001**	1.31 (1.28-1.34)	**<.001**	1.22 (1.19-1.25)	**<.001**
	Yes	-	-	-	-	-	-

[1] Clinically determined size or extension of the primary tumor
[2] Clinically determined absence or presence of regional lymph node metastasis

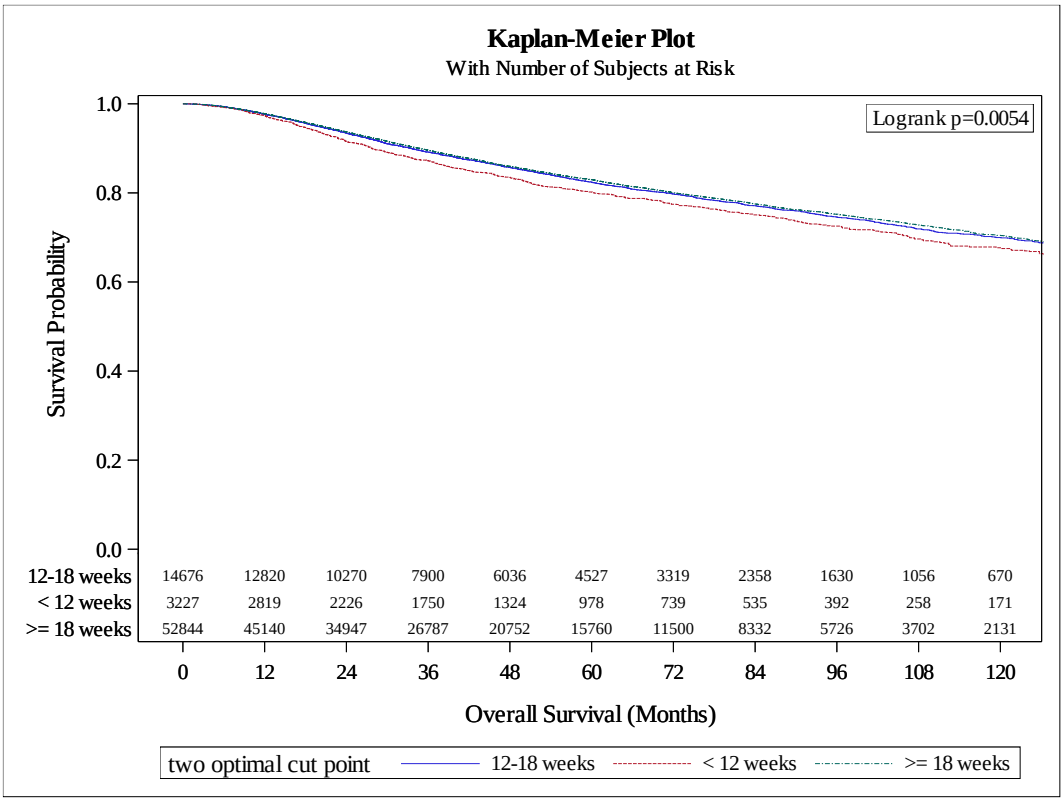
		UVA before weighting		UVA after weighting		MVA before weighting	
Clinical Characteristics	**Level**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**	**Hazard Ratio (95% CI)**	**P-value**
Surgical Margin	Positive	1.40 (1.33-1.48)	**<.001**	1.40 (1.32-1.47)	**<.001**	1.39 (1.32-1.47)	**<.001**
	Unknown	1.17 (1.04-1.33)	**0.012**	1.20 (1.06-1.35)	**0.004**	1.13 (1.00-1.28)	0.053
	Negative	-	-	-	-	-	-
Year of Diagnosis	2012-2015	1.11 (1.07-1.16)	**<.001**	1.11 (1.07-1.16)	**<.001**	1.05 (1.00-1.10)	0.065
	2008-2011	1.04 (1.01-1.07)	**0.014**	1.03 (1.00-1.07)	**0.041**	0.99 (0.96-1.02)	0.520
	2004-2007	-	-	-	-	-	-
Age at Diagnosis		1.05 (1.05-1.05)	**<.001**	1.05 (1.05-1.05)	**<.001**	1.03 (1.03-1.04)	**<.001**

Figure 1. KM plot for neoadjuvant chemotherapy group before weighting



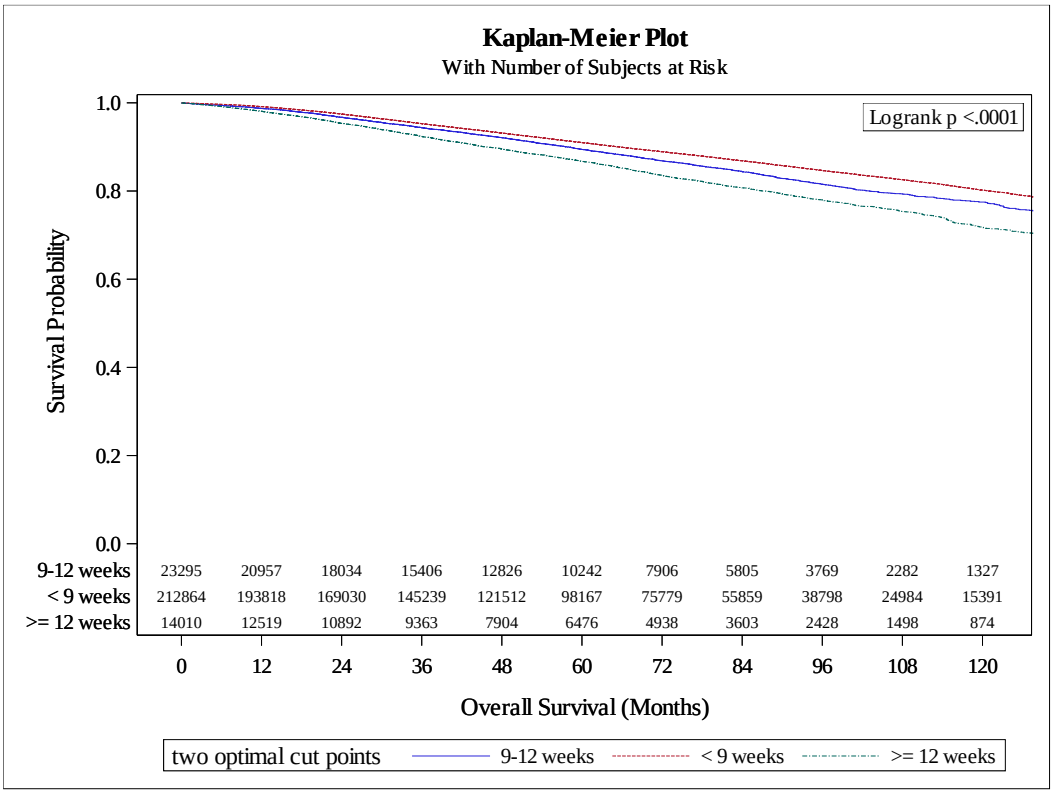
two optimal cut-points	No. of Subject	Event	Censored	5-year Survival	10-year Survival
< 12 weeks	3235	690 (21%)	2545 (79%)	78.8% (77.1%, 80.4%)	66.5% (64.0%, 68.9%)
12-18 weeks	14665	2354 (16%)	12311 (84%)	82.1% (81.3%, 82.8%)	69.4% (68.0%, 70.7%)
>= 18 weeks	52831	6720 (13%)	46111 (87%)	83.3% (82.9%, 83.7%)	71.0% (70.1%, 71.9%)

Figure 2. KM plot for neoadjuvant chemotherapy group after weighting



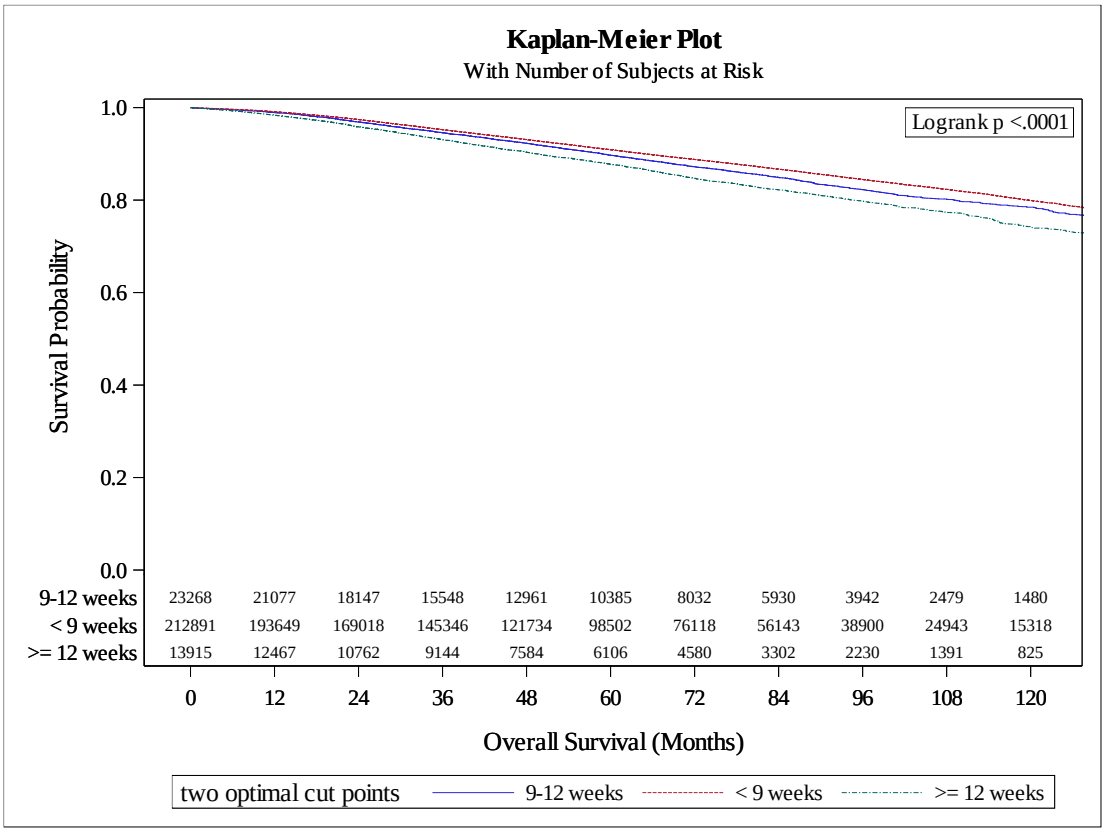
two optimal cut-points	No. of Subject	Event	Censored	5-year Survival	10-year Survival
< 12 weeks	3235	690 (21%)	2545 (79%)	80.2% (78.2%, 82.0%)	67.7% (64.9%, 70.3%)
12-18 weeks	14665	2354 (16%)	12311 (84%)	82.4% (81.6%, 83.2%)	70.0% (68.6%, 71.3%)
>= 18 weeks	52831	6720 (13%)	46111 (87%)	83.0% (82.5%, 83.4%)	70.5% (69.6%, 71.4%)

Figure 3. KM plot for adjuvant chemotherapy group before weighting



two optimal cut-points	No. of Subject	Event	Censored	5-year Survival	10-year Survival
< 9 weeks	212864	20802 (10%)	192062 (90%)	91.0% (90.8%, 91.1%)	80.2% (79.9%, 80.5%)
9-12 weeks	23295	2594 (11%)	20701 (89%)	89.4% (88.9%, 89.9%)	77.5% (76.5%, 78.5%)
>= 12 weeks	14010	2039 (15%)	11971 (85%)	86.7% (86.0%, 87.4%)	71.8% (70.4%, 73.2%)

Figure 4. KM plot for adjuvant chemotherapy group after weighting



two optimal cut-points	No. of Subject	Event	Censored	5-year Survival	10-year Survival
< 9 weeks	212864	20802 (10%)	192062 (90%)	90.9% (90.8%, 91.0%)	79.9% (79.6%, 80.2%)
9-12 weeks	23295	2594 (11%)	20701 (89%)	89.7% (89.2%, 90.2%)	78.5% (77.4%, 79.5%)
>= 12 weeks	14010	2039 (15%)	11971 (85%)	87.7% (87.0%, 88.4%)	74.3% (72.8%, 75.7%)