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:

_____ _____

Danyang Wang Date

Determinants and Prognostic Impact of Extent of Surgery for Thyroid Microcarcinoma

By

Danyang Wang
Master of Public Health

Epidemiology

_____

Kevin C. Ward, Ph.D., M.P.H.
Committee Chair

Determinants and Prognostic Impact of Extent of Surgery for Thyroid Microcarcinoma

By

Danyang Wang

B.S., Shanghai Jiao Tong University, 2015
M.S., Shanghai Jiao Tong University, 2018

Thesis Committee Chair: Kevin C. Ward, Ph.D., M.P.H.

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 Epidemiology
2021

Abstract

Determinants and Prognostic Impact of Extent of Surgery for Thyroid Microcarcinoma
By Danyang Wang

Background: Total thyroidectomy (TT) is the dominant surgical procedure in the management of thyroid microcarcinoma, despite the current guidelines suggesting that thyroid lobectomy (TL) alone is sufficient for localized microcarcinoma. This study primarily aimed to assess the possible demographic or clinical factors that may affect the decision on the extent of surgery for primary thyroid microcarcinoma.

Methods: Patients who were diagnosed with primary thyroid microcarcinoma between 2004 and 2016 were included from the U.S. National Cancer Institute Surveillance, Epidemiology, and End Results 18 registries database. Univariate and multivariate analyses using polytomous logistic regression were performed to analyze the association between demographic or clinical characteristics and the extent of surgery. The Kaplan-Meier method was used to estimate thyroid cancer related survival and the log-rank test was used to compare survival rates between groups. Multivariate analysis using Cox proportional hazards model was used to estimate the independent prognostic effect of surgery type on cause-specific survival (CSS).

Results: The cohort consisted of 45,495 patients. Overall, 76.8% of the patients underwent TT, 22.8% underwent TL, and 4.2% had no surgery. According to multivariate analysis, TL, compared to TT, was more frequently performed in patients with age $\geq$ 65 years (odds ratio [OR]=1.19, 95% confidence interval [CI] 1.11-1.27) and other non-Hispanic races (OR=1.18, 95% CI 1.09-1.28), and less likely to be performed in females (OR=0.73, 95% CI 0.68-0.77), non-Hispanic blacks (OR=0.81, 95% CI 0.74-0.89), and those with higher-stage cancers (OR=0.23, 95% CI 0.20-0.25) and multifocal tumors (OR=0.40, 95% CI 0.38-0.42). Excellent 10-year CSS was observed following both TT and TL in patients with early-stage thyroid microcarcinoma and no difference in CSS was found between patients who underwent TT vs. TL.

Conclusions: TT remains the predominant surgical method for treating primary thyroid microcarcinoma and this trend has increased in recent years, despite a lack of evidence of survival advantage offered by more extensive surgical procedures. In order to improve the quality of life of the patients, reduce healthcare costs, and prevent overtreatment, TT should be performed on a selected group of patients with a high risk of tumor recurrence in the management of thyroid microcarcinoma.

Determinants and Prognostic Impact of Extent of Surgery for Thyroid Microcarcinoma

By

Danyang Wang

B.S., Shanghai Jiao Tong University, 2015
M.S., Shanghai Jiao Tong University, 2018

Thesis Committee Chair: Kevin C. Ward, Ph.D., M.P.H.

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 Epidemiology
2021

Table of Contents

1. Introduction

The incidence of thyroid cancer has increased rapidly during the last three decades, [1,2] which largely reflects the detection of small papillary thyroid cancers or clinically occult tumors due to increased use of neck ultrasound. [3] Concerns of overdiagnosis have been raised, and the fact that mortality due to thyroid cancer remained stable despite the dramatic rise in incidence may be one reflection of the existing overdiagnosis. [4]

One of the most significant harms of overdiagnosis is that it can trigger overtreatment. Even though active surveillance has been proved to be an accepted and viable way for the control of thyroid microcarcinoma, some patients may have a strong will to have their cancer removed. [5] If surgery is chosen, thyroid lobectomy (TL) and total thyroidectomy (TT) are the two major surgical procedures for primary thyroid carcinoma, depending on the extent of disease. According to the latest 2015 American Thyroid Association (ATA) guidelines, TL alone should be sufficient for patients with thyroid microcarcinoma (tumor size $\leq$ 10 mm) when there is no extrathyroidal extension or lymph node involvement, even though this recommendation is mainly based on the evidence from studies of papillary thyroid microcarcinoma (PTMC). [6] A number of studies including systematic reviews have shown that both TL and TT are excellent and comparable surgery methods for the treatment of PTMC with regard to long-term survival. [7-10] Studies comparing the prognostic value of TL vs. TT for follicular thyroid microcarcinoma (FTMC) or Hürthle cell thyroid microcarcinoma (HCTMC) are relatively scarce. A recent study including 203 patients with FTMC indicated no difference in overall survival for the two surgical procedures. [11]

Lobectomy, compared to the more conventional total thyroidectomy, carries lower surgical risks of vocal cord palsy due to recurrent nerve damage and hypoparathyroidism, preserves functioning thyroid tissue, and reduces the need for permanent thyroid hormone replacement. However, despite the ATA recommendation and a lower incidence of surgical complications after TL, TT is still the dominant surgical procedure in the current clinical practice, even for small localized papillary carcinoma. It has been reported that about 80% of patients who have surgery for localized papillary thyroid cancer underwent total thyroidectomy during 1988-2014. [12,13]

Therefore, this study primarily aimed to assess the possible demographic or clinical factors that may affect the decision on the extent of surgery for primary thyroid microcarcinoma. The secondary goal of this study was to evaluate the impact of different extent of surgery on survival outcomes of patients with PTMC, FTMC and HCTMC.

2. Materials and Methods

2.1 Data Collection

2.1.1 Study Population

Cases that met the selection criteria were extracted from the Surveillance, Epidemiology and End Results (SEER) 18 registries database of National Cancer Institute. Patients who were diagnosed with a first primary differentiated thyroid cancer (DTC) (International Classification of Diseases for Oncology histology codes 8050, 8260, 8290, 8330- 8332, 8335, 8337, 8341-8344, 8350, and 8450) between 2004 and 2016 and had a tumor size of $\leq$ 1 cm were eligible for inclusion. Histologic subtypes of differentiated thyroid cancer

were identified using the International Classification of Diseases for Oncology, third edition (ICD-O-3). Papillary thyroid carcinoma (PTC) was defined as ICD-O-3 codes 8050, 8260, 8340-8344, and 8450, follicular thyroid carcinoma (FTC) as codes 8330-8332, 8335, 8337, and Hürthle cell thyroid carcinoma (HCTC) as code 8290. In addition, ICD-O-3 topography code C73.9 was applied to the patient selection to screen for primary cancer originated from the thyroid gland. For patients diagnosed between 2004 and 2015, Collaborative Stage (CS) codes were used to identify tumor size, and patients with CS tumor size (2004-2015) codes 001-010, which indicate a tumor size of 1 to 10 millimeters, and code 991, which indicates a tumor size of less than 10 millimeters, were included. For patients diagnosed in 2016, Tumor Size Summary (2016) codes were used to identify tumor size and patients with codes 001-010, which indicate a tumor size of 1 (or < 1) to 10 millimeters, were included. Cancer stage was classified as early-stage (i.e., localized) and late-stage (i.e., regional or distant) according to Summary Stage 2000 (1998+) and patients with unknown stage were excluded for the analysis. Surgery type was divided into three categories: 1) no surgery because surgery was not recommended; 2) TL, defined as removal of a lobe or less than a lobe or removal of a lobe and partial removal of the contralateral lobe, and 3) TT, defined as subtotal or near total or total thyroidectomy. Patients with unknown information regarding whether cancer-directed surgery was performed or unknown surgery type or those who did not receive cancer-directed surgery when surgery was recommended were excluded. Figure 1 showed the flow chart of patient selection process of this study.

2.1.2 Covariates

Data including the demographics of the patients (sex, age at diagnosis, race, and year of diagnosis), characteristics of the tumor (histology, tumor size, multifocality, and SEER stage), treatment information (type of surgery, reason for no surgery of the primary site), survival information (survival time, cause-specific death status), and the insurance status were collected from the database. Based on the age at diagnosis, patients were divided into four age groups: ≤ 44 years old, 45-54 years old, 55-64 years old, and ≥ 65 years old. Race was categorized as non-Hispanic white (NHW), non-Hispanic black (NHB), Hispanic, and other races. Tumor size was categorized into ≤ 6 mm and > 6 mm, since a tumor diameter > 6 mm was reported to be associated with less favorable disease-free survival among patients with PTMC (Besic et al., 2008). [14] With regard to insurance status, most patients who were 65 years old or older at the age of diagnosis that were classified as "Uninsured" or "Insurance status unknown" in the SEER database were Medicare eligible, therefore the insurance status was further stratified on age (< 65 years vs. ≥ 65 years). Data were obtained using SEER*Stat 8.3.9.

2.2 Data Organization

First, we checked each variable for missing values and set all the unknown or unspecified data inputs to missing. Second, we checked and cleaned survival related variables including survival time (in month), all cause death status and cause-specific death status. For the data cleaning of survival time, the variable survival months flag in the SEER database was used as an indicator of whether complete dates or sufficient follow-up time between the date of last contact and the date of diagnosis were available for the survival analysis. Patients with survival months flag code 0 (complete dates were available and

there were zero days of follow-up; N = 183) and 2 (incomplete dates are available and there could be zero days of follow-up; N = 22) were excluded for the survival analysis and their survival times were set to missing. In the remaining cohort, patients diagnosed in 2016 (N = 3864) and those diagnosed between 2004 and 2015 but with zero months of follow-up (N = 305) were also excluded for the survival analysis to maintain sufficient follow-up time of the cohort.

2.3 Statistical Methods

2.3.1 Descriptive Analysis

When reporting the descriptive statistics of each covariate, patients were divided into three surgery groups: no surgery, TL, TT. Means, standard deviations, medians, the first and the third quantiles (Q1 and Q3) and range were reported for continuous variables. Frequencies and proportions were reported for categorical variables.

2.3.2 Bivariate Associations Between Type of Surgery and Other Covariates

Univariate model analysis with type of surgery as the outcome and other categorical and continuous covariates as explanatory variables (i.e., exposures) separately was performed to evaluate the bivariate associations. Considering the primary outcome, type of surgery, has three levels, polytomous logistic regression was applied to construct the univariate models.

The detailed description of the models is as below:

Let $P(Y = j\,|X)$ be the probability that an individual falls into surgery type (Y) category j at a fixed setting X for explanatory variables and $\sum_j P(Y = j\,|X) = 1$. Logit models pair

each response category with the baseline category of the response variable Y. In our analysis, the baseline category is set to be thyroidectomies ($j = 0$), since this category has the largest number of patients among the three surgery type categories. The model is then:

$$\ln \frac{P(Y = j \mid X)}{P(Y = 0 \mid X)} = \alpha_j + \beta_j X$$

where $j = 1\ (TL), 2\ (no\ surgery)$, X is the explanatory variable selected, and α_j and β_j are the intercept and slope parameters for category j.

Considering that most patients who were 65 years old or older at the age of diagnosis were Medicare eligible but could be classified as "Uninsured" or "Insurance status unknown" in the SEER database, additional stratified analysis for this variable was performed, where patients were stratified by age (< 65 vs. ≥ 65 years of age).

2.3.3 Multivariate Analysis

Multivariate polytomous logistic regression was used to analyze the independent association between type of surgery and demographic or tumor characteristic variables. Since there could be potential misclassifications of insurance status regarding Medicare eligibility as mentioned above and univariate analysis indicated no significant correlation between type of surgery and insurance status, the covariate insurance status was not included in the multivariable models.

In order to further analyze the possible impact factors of the prevalence of TL vs. TT among low-risk patients, a subgroup analysis restricted to patients with localized thyroid microcarcinoma and treated with surgery was performed. Multivariate logistic regression

model was used to calculate the odds ratios (ORs) of TL over TT and the 95% confidence intervals (CIs).

2.3.4 Survival Analysis

The primary end-point of the survival model was cause-specific survival (CSS). The Kaplan-Meier method was used to estimate thyroid cancer related survival for patients diagnosed with thyroid microcarcinoma during 2004-2015 and the log-rank test was used to compare survival rates between groups for each covariate. Multivariate analysis using Cox proportional hazards model was used to estimate the prognostic effect of surgery type on CSS while controlling for factors including age, gender, race, year of diagnosis, histology of tumor, cancer stage, and tumor size. The proportional hazards (PH) assumption for the exposure variable, type of surgery, as well as for other covariates were evaluated comprehensively by three methods: graphical methods using log-log survivals curves, goodness of fit tests, and time-dependent variables. The PH assumption would be regarded as satisfied for a variable if at least two of these methods indicate no violation. As a result, the PH assumption is met for all the covariates except cancer stage, thus a stratified Cox procedure based on cancer stage was performed:

$$h_{j(t,X)} = h_{0j(t)} * e^{(\beta_{1j} SURGERY + \sum_{i=1}^{m} \gamma_{ij} X_i)}$$

where $j = 1$ (early-stage), 2 (late-stage). As shown in the model, the baseline hazard, $h_{0j(t)}$ could be varying with different stage category *j*, and so as the effect of surgery type on survival (i.e., hazard ratio [HR]), $e^{\beta_{1j}}$, which would allow for the analysis of the interaction between surgery type and cancer stage.

All the data organizations and analyses were performed using SAS 9.4 (SAS Institute, Cary, NC).

3. Results

3.1 Descriptive Statistics

A total of 45,495 patients diagnosed as DTC from 2004 to 2016 with tumor size $\leq$ 1 cm were included in the analysis. Among these patients, 76.8% (34,922/45,495) underwent TT, 22.8% (10,381/45,495) underwent TL, and 4.2% (192/45,495) did not had surgery for the primary site because surgery was not recommended by the provider (Table 1). In order to assess the distribution of demographic or tumor characteristic variables across different surgery groups, we grouped patients by the extent of surgery when reporting descriptive statistics.

As shown in Table 1, the TT group seemed to be comprised of more younger patients and females, compared to the no surgery group and the TL group. 69.8% of the TT group were NHW, 6.8% were NHB, 13.2% were Hispanic, and 9.4% were other non-Hispanic races. A similar distribution pattern of race was observed for the TL group, with 70.9% NHW, 6.5% NHB, 11.6% Hispanic, and 9.7% other non-Hispanic races, while the no surgery group tended to have less NHWs (65.1%) and NHBs (4.2%) and more patients of other non-Hispanic races (15.6%). Patients were stratified into two groups ($<$ 65 years vs. $\geq$ 65 years), when reporting the insurance status. Among patients who were younger than 65 years of age, 81.0% in the TT group, 76.8% in the TL group, and 82.1% in the no surgery group were insured. Among patients who were 65 years old or older, 85.1% in

the TT group, 81.9% in the TL group, and 77.8% in the no surgery group were insured. The majority of patients (63.0%) in the no surgery group were diagnosed between 2012 and 2016, while less proportion of patients in the TL group and the TT group (45.0% and 46.0%, respectively) were diagnosed in or after 2012. A predominant proportion of patients (99.0% in the no surgery group, 98.5% in the TL group, and 98.7% in the TT group, respectively) were with PTMC and no specific pattern was observed with regard to the difference in the distribution of histology subtype across comparison groups. The mean tumor size, as measured by the longest diameter of the tumor, of patients in the TL group was 4.2 mm (standard deviation [SD] 2.8 mm), which was lower than that of patients in the no surgery group (mean tumor size 6.9 mm, SD 2.9 mm) and the TT group (mean tumor size 5.8 mm, SD 3.0 mm). Less patients (22.8%) in the TL group had a tumor size larger than 6.0 mm, compared to 43.5% in the TT group and 63.5% in the no surgery group. Patients who underwent TT tended to have higher stage disease with higher proportion of patients in this group having regional or distant disease and lower proportion of patients having localized disease, compared to those in the other two groups.

3.2 Bivariate Associations Between Type of Surgery and Other Covariates

Bivariate associations between surgery type and demographic or tumor characteristics were assessed using polytomous logistic regression and the results including ORs and 95% CIs were listed in Table 2.

The TT group was regarded as the baseline category of the response variable and each of the remaining categories were paired to the baseline category when calculating odds. The odds of TL vs. TT was associated with age, gender, year of diagnosis, histology, tumor

size, cancer stage, and multifocality, while the odds of no surgery vs. TT was associated with gender, race, year of diagnosis, tumor size, cancer stage, and multifocality. Insurance status was shown to be not associated with surgery type.

Specifically, as shown in Table 2, older patients, compared to those who were 44 years old or younger, had significantly higher odds of TL vs. TT, while no difference was found in the odds of no surgery vs. TT across different age groups. Compared to males, females were more likely to receive TT rather than TL or no surgery for the primary site: the odds of TL vs. TT among females was 0.83 (95% CI 0.78-0.87) times the corresponding odds among males and the odds of no surgery vs. TT among females was 0.68 (95% CI 0.49-0.96) times the corresponding odds among males. Race was not associated with surgery types except that patients in other races seemed to have higher odds of receiving no surgery (vs. TT) than whites. The odds of TL vs. TT decreased over time with patients diagnosed in 2008 and after having lower odds of lobectomy vs. thyroidectomy than those diagnosed between 2004 and 2007 (OR = 0.84 for patients diagnosed during 2008-2011, 95% CI 0.79-0.89; OR = 0.87 for patients diagnosed during 2012-2016, 95% CI 0.82-0.92). On the contrary, the odds of no surgery vs. TT for patients diagnosed between 2012 and 2016 was 1.57 times higher than the corresponding odds for those diagnosed between 2004 and 2007 (OR = 2.57, 95% CI 1.63-4.04). The odds of TL vs. TT among patients with FTMC was 1.30 (95% CI 1.05-1.60) times the odds among patients with PTMC. As for tumor size, patients with a tumor size larger than 6 mm were more likely to receive no surgery rather than TT (OR = 2.21, 95% CI 1.64-2.97) and less likely to undergo TL rather than TT (OR = 0.39, 95% CI 0.37-0.41). The odds of TL vs. TT and the odds of no surgery vs. TT among patients with regional or

distant disease were both significantly lower than that among patients with localized disease (OR = 0.16, 95% CI 0.14-0.18 and OR = 0.40, 95% CI 0.24-0.68, respectively). In addition, multifocality seemed to be another important indicator of more aggressive surgery. The odds of TL vs. TT and the odds of no surgery vs. TT among patients with multifocal tumor were significantly lower than that among patients with unifocal tumor (OR = 0.33, 95% CI 0.31-0.35 and OR = 0.24, 95% CI 0.16-0.37, respectively).

3.3 Multivariate Analysis

After controlling for other covariates, age, gender, race, year of diagnosis, tumor size, cancer stage, and multifocality were all independently correlated with surgery type. As shown in Table 3, when adjusting for other demographic and clinical variables, the odds of TL vs. TT was 19% higher in patients aged 65 years or older (OR = 1.19, 95% CI 1.11-1.27), compared to the reference age group (≤ 44 years). The odds of TL vs. TT as well as the odds of no surgery vs. TT were similar between the reference age group and other age groups. As with in the univariate models, females were showed to have lower odds of receiving TL or no surgery vs. TT than males (OR = 0.73, 95% CI 0.68-0.77 and OR = 0.61, 95% CI 0.42-0.87, respectively). Compared to NHW, the odds of TL vs. TT was significantly lower in NHB (OR = 0.81, 95% CI 0.74-0.89) and higher in other non-Hispanic races (OR = 1.18, 95% CI 1.09-1.28. The odds of no surgery vs. TT was significantly higher in other non-Hispanic races than NHW (OR = 1.77, 95% CI 1.17-2.68), while no difference in the odds of no surgery vs. TT was found between NHWs or Hispanics vs. NHBs, after adjusting for other covariates. With regard to year of diagnosis, tumor size, cancer stage, and multifocality, multivariate analysis also generated similar results as in the univariate polytomous logistic regression models, while

histology was shown to be not associated with the extent of surgery in the multivariate model. The odds of TL vs. TT for patients diagnosed during 2008-2011 and patients diagnosed during 2012-2016 were 0.83 (95% CI 0.78-0.89) and 0.87 (95% CI 0.82-0.92) times, respectively, the corresponding odds for patients diagnosed during 2004-2007. The odds of no surgery vs. TT for patients diagnosed during 2008-2011 and 2012-2016 was 1.48 (95% CI 0.86-2.55) and 2.64 (95% CI 1.62-4.31) times, respectively, the corresponding odds for those diagnosed during 2004-2007. Patients with a relatively larger tumor (> 6 mm) were more likely to receive no surgery rather than TT (OR = 2.81, 95% CI 2.05-3.86) and less likely to undergo TL rather than TT (OR = 0.49, 95% CI 0.46-0.51). The odds of TL vs. TT and the odds of no surgery vs. TT among patients with regional or distant disease were 0.30 (95% CI 0.17-0.55) and 0.23 (95% CI 0.20-0.25) times, respectively, the odds among patients with localized disease. The odds of TL vs. TT and the odds of no surgery vs. TT among patients with multifocal tumor were significantly lower than that among patients with unifocal tumor (OR = 0.40, 95% CI 0.38-0.42 and OR = 0.22, 95% CI 0.14-0.34, respectively).

In the low-risk subgroup that comprised 37,425 patients with localized thyroid microcarcinoma who underwent surgery, factors including female gender, NHB race, more recent years of diagnosis, tumor size > 6 mm, and multifocal tumor were also found to be associated with higher prevalence of TT than TL, while age ≥ 65 years and other non-Hispanic races were associated with less aggressive surgery (Table 4).

3.4 Survival

The median follow-up time was 61 months. In general, patients with localized disease had excellent CSS. The10-year CSS among patients with localized PTMC was over 99%

and that among patients with localized FTMC or HCTMC was over 98% (Figure 2A & B). The log-rank test indicated no difference in CSS among different surgery groups for localized PTMC ($p = 0.76$) as well as for localized FTMC or HCTMC ($p = 0.96$). Among patients with regional/distant disease, however, patients who underwent no surgery had significantly lower 10-year CSS than those who underwent TL or TT (76% vs. 96% or 98%; log-rank $p < 0.001$). Moreover, Tukey-adjusted pairwise comparison showed that there was no difference in CSS between TL group and TT group ($p = 0.11$). Only 42 patients in our sample were with regional/distant FTMC or HCTMC, and among which, the six patients who underwent TL had a 10-year CSS of 100% and the 36 who underwent TT had a 10-year CSS of 83%. Log-rank test was not performed due to the sparse number of cases in this stratum.

The results of stratified Cox regression were shown in Table 5. Overall, after controlling for age, gender, race, year of diagnosis, tumor histology, tumor size, and multifocality and stratified on cancer stage, no significant difference was found in CSS between the TL group and the TT group. Specifically, among patients with localized disease, the hazard of death due to thyroid cancer for patients who underwent TL is 0.88 (95% CI 0.48-1.60) times the corresponding hazard for patients who underwent TT. Among patients with regional/distant disease, the hazard of death due to thyroid cancer for patients who underwent TL is 1.85 (95% CI 0.86-3.95) times the corresponding hazard for patients who underwent TT, while the hazard of death due to thyroid cancer for patients who had no surgery is 15.09 (95% CI 3.48-65.34) times the corresponding hazard for patients who underwent TT.

4. Discussion

TT still remains the predominant surgical method for treating primary thyroid microcarcinoma. Moreover, based on our results, TT has been performed more frequently compared to TL or active surveillance in recent years, despite the ATA guidelines recommending a less aggressive treatment method, especially for low-risk thyroid microcarcinomas.

Recommendations on the extent of surgery for thyroid microcarcinoma remain controversial according to the literature. A meta-analysis comprising 13,810 PTMC patients indicated that patients who underwent TL had an increased risk of recurrence but similar mortality rates, compared to those who underwent TT. [9] Another more recent meta-analysis of nine retrospective studies and 21,594 PTMC patients also confirmed that TT, in comparison to TL, was associated with a slightly better long-term recurrence-free survival (RFS), although excellent RFS was observed following both TT and TL (10-year RFS 95% vs. 92%). There was also evidence that TT and TL had comparative prognostic impacts among patients with FTMC and HCTMC. [11,15] However, it has also been demonstrated by several studies that more extensive surgery should be performed to a subset of patients with thyroid microcarcinoma with certain clinical features such as multifocality, extrathyroidal invasion, lymph node metastases, and distant metastases. [16-18] Furthermore, a retrospective cohort study comprising 18,445 cases of PTMC showed that the presence of two or more risk factors, including age > 45 years, male sex, African American or minority race, extrathyroidal invasion, lymph node metastases, and distant metastases, was strongly associated with cancer specific mortality. [19] Our study found

that patients who underwent TT and TL had comparative cancer related long-term mortality, even after holding cancer stage, multifocality, and histology subtype constant.

Thyroid microcarcinomas may have a wide spectrum of aggressiveness, which could possibly lead to varying prognosis and impact the choice of management approach. According to a system review, non-incidentally diagnosed thyroid microcarcinomas were found to have significantly higher rates of recurrence, multifocality and mortality than incidentally diagnosed thyroid microcarcinomas. [20] In addition, a retrospective cohort study including 243 PTMC patients showed that tumor size > 8 mm was correlated with more aggressive disease. [21] Another European study that comprised of 228 patients with PTMC also indicated that patients with tumor diameter ≤ 6 mm and no lymph-node metastases presented longer disease-free interval. [14] These findings were in consistent with our results that patients with larger and/or multifocal tumor and higher stage of disease tended to be treated with TT than TL.

However, we also observed a paradox of surgery selection that some of the patients who were usually considered as at higher risk of cancer related mortality and less favorable prognosis, such as older patients (≥ 65 years old) or those with FTMC or HCTMC, were actually less likely to receive more radical surgical procedures. FTMC and HCTMC, compared to PTMC, were reported to be associated with higher risk of distant metastasis. [15] Two major methods for distant metastasis surveillance in patients with thyroid carcinoma are serum thyroglobulin level monitoring and whole-body radioiodine scan, both of which are required to be performed under the premise of no remaining functioning thyroid lobe. [22] Thus, choosing to undergo TL may preclude the use of these

surveillance tools and impede the detection of potential distant metastasis, especially for high-risk patients.

4.1 Limitations

This study had several limitations. First, we were unable to obtain information regarding whether a case was incidentally diagnosed or not from the SEER database. It is possible that, for a certain proportion of patients, the tumor was incidentally diagnosed on surgical pathology, and this may serve as a viable explanation of high proportion of TT in our cohort. Besides, as mentioned above, non-incidentally diagnosed thyroid microcarcinoma cases were at higher risk of less favorable prognosis, compared to incidentally diagnosed cases, which might potentially confound the multivariate analysis when not being controlled for. In addition, data of other risk factors of thyroid cancer, such as history of head and neck irradiation and family history, as well as information on surgical complications were not available, which precluded us to decide what percentage of mortality was attributed to more aggressive surgical choice in the management of thyroid microcarcinoma. Lastly, the recurrence-free interval was not estimated in our analysis due to lack of data. Further prospective studies assessing the association between the extent of surgery and disease recurrence of thyroid microcarcinoma are needed.

4.2 Conclusions

TT remains the predominant surgical method for treating primary thyroid microcarcinoma and this trend has increased in recent years, despite a lack of evidence of survival advantage offered by more extensive surgical procedures. In order to improve the quality of life of the patients, reduce healthcare costs, and prevent overtreatment, TT

should be performed on a selected group of patients with a high risk of tumor recurrence in the management of thyroid microcarcinoma.

References:

1. Davies L, Welch HG. Current thyroid cancer trends in the United States. *JAMA otolaryngology–head & neck surgery.* 2014;140(4):317-322.
2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2015. *CA: a cancer journal for clinicians.* 2015;65(1):5-29.
3. Brito JP, Al Nofal A, Montori VM, Hay ID, Morris JC. The impact of subclinical disease and mechanism of detection on the rise in thyroid cancer incidence: a population-based study in Olmsted County, Minnesota during 1935 through 2012. *Thyroid.* 2015;25(9):999-1007.
4. Ahn HS, Kim HJ, Welch HG. Korea's thyroid-cancer "epidemic"—screening and overdiagnosis. *N Engl J Med.* 2014;371(19):1765-1767.
5. Leboulleux S, Tuttle RM, Pacini F, Schlumberger M. Papillary thyroid microcarcinoma: time to shift from surgery to active surveillance? *The lancet Diabetes & endocrinology.* 2016;4(11):933-942.
6. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid.* 2016;26(1):1-133.
7. Lin HW, Bhattacharyya N. Survival impact of treatment options for papillary microcarcinoma of the thyroid. *The Laryngoscope.* 2009;119(10):1983-1987.

8. Wang TS, Goffredo P, Sosa JA, Roman SA. Papillary thyroid microcarcinoma: an over-treated malignancy? *World journal of surgery.* 2014;38(9):2297-2303.

9. Zheng W, Li J, Lv P, Chen Z, Fan P. Treatment efficacy between total thyroidectomy and lobectomy for patients with papillary thyroid microcarcinoma: a systemic review and meta-analysis. *European Journal of Surgical Oncology.* 2018;44(11):1679-1684.

10. Vargas-Pinto S, Arenas MAR. Lobectomy compared to total thyroidectomy for low-risk papillary thyroid cancer: a systematic review. *Journal of Surgical Research.* 2019;242:244-251.

11. Megwalu UC, Green RW. Total thyroidectomy versus lobectomy for the treatment of follicular thyroid microcarcinoma. *Anticancer research.* 2016;36(6):2899-2902.

12. Esfandiari NH, Hughes DT, Reyes-Gastelum D, Ward KC, Hamilton AS, Haymart MR. Factors associated with diagnosis and treatment of thyroid microcarcinomas. *The Journal of Clinical Endocrinology & Metabolism.* 2019;104(12):6060-6068.

13. Welch HG, Doherty GM. Saving thyroids-overtreatment of small papillary cancers. *The New England journal of medicine.* 2018;379(4):310-312.

14. Besic N, Pilko G, Petric R, Hocevar M, Zgajnar J. Papillary thyroid microcarcinoma: prognostic factors and treatment. *Journal of surgical oncology.* 2008;97(3):221-225.

15. Kuo EJ, Roman SA, Sosa JA. Patients with follicular and Hurthle cell microcarcinomas have compromised survival: a population level study of 22,738 patients. *Surgery.* 2013;154(6):1246-1254.

16. Ross DS, Litofsky D, Ain KB, et al. Recurrence after treatment of micropapillary thyroid cancer. *Thyroid.* 2009;19(10):1043-1048.

17. Arora N, Turbendian HK, Kato MA, Moo TA, Zarnegar R, Fahey III TJ. Papillary thyroid carcinoma and microcarcinoma: is there a need to distinguish the two? *Thyroid.* 2009;19(5):473-477.

18. Malandrino P, Pellegriti G, Attard M, et al. Papillary thyroid microcarcinomas: a comparative study of the characteristics and risk factors at presentation in two cancer registries. *The Journal of Clinical Endocrinology & Metabolism.* 2013;98(4):1427-1434.

19. Yu X-M, Wan Y, Sippel RS, Chen H. Should all papillary thyroid microcarcinomas be aggressively treated?: an analysis of 18,445 cases. *Annals of surgery.* 2011;254(4):653-660.

20. Mehanna H, Al-Maqbili T, Carter B, et al. Differences in the recurrence and mortality outcomes rates of incidental and nonincidental papillary thyroid microcarcinoma: a systematic review and meta-analysis of 21 329 person-years of follow-up. *The Journal of Clinical Endocrinology & Metabolism.* 2014;99(8):2834-2843.

21. Roti E, Rossi R, Trasforini G, et al. Clinical and histological characteristics of papillary thyroid microcarcinoma: results of a retrospective study in 243 patients. *The Journal of Clinical Endocrinology & Metabolism.* 2006;91(6):2171-2178.

22. Gartland RM, Lubitz CC. Impact of extent of surgery on tumor recurrence and survival for papillary thyroid cancer patients. *Annals of surgical oncology*. 2018;25(9):2520-2525.

Tables & Figures

Table 1. Patient characteristics by surgery type

	No surgery (N = 192)		TL[a] (N = 10381)		TT[b] (N = 34922)	
	N	%	N	%	N	%
Age, years						
≤ 44	66	34.4	3281	31.6	12621	36.1
45-54	50	26.0	2796	26.9	9413	27.0
55-64	40	20.8	2404	23.2	7742	22.2
≥ 65	36	18.8	1900	18.3	5146	14.7
Gender						
Female	148	77.1	8332	80.3	29023	83.1
Male	44	22.9	2049	19.7	5899	16.9
Race						
NHW	125	65.1	7358	70.9	24364	69.8
NHB	8	4.2	676	6.5	2359	6.8
Hispanic	22	11.5	1200	11.6	4598	13.2
Other	30	15.6	1010	9.7	3264	9.4
Unknown	7	3.6	137	1.3	337	1.0
Insurance						
< 65 years						
Uninsured	6	3.8	157	1.9	575	1.9
Insured	128	82.1	6517	76.8	24111	81.0
Unknown	22	14.1	1807	21.3	5090	17.1
≥ 65 years						
Uninsured	0	0.0	9	0.5	24	0.5
Insured	28	77.8	1557	81.9	4377	85.1
Unknown	8	22.2	334	17.6	745	14.5
Year of diagnosis						
2004-2007	22	11.5	2513	24.2	7489	21.4
2008-2011	49	25.5	3196	30.8	11376	32.6
2012-2016	121	63.0	4672	45.0	16057	46.0
Histology						
PTMC	≥ 182	≥ 94.8	10226	98.5	34477	98.7
FTMC	≤ 5	≤ 2.6	123	1.2	312	0.9
HCTMC	≤ 5	≤ 2.6	35	0.3	133	0.4
Tumor size, mm						
Mean (SD)	6.9 (2.9)	-	4.2 (2.8)	-	5.8 (3.0)	-
≤ 6	70	36.5	7857	75.7	19405	55.6
> 6	122	63.5	2366	22.8	15185	43.5
≤ 10, NOS	0	0.0	158	1.5	332	0.9
SEER stage						
Localized	176	91.7	10021	96.5	28512	81.6
Regional/Distant	16	8.3	360	3.5	6410	18.4
Multifocal						

No	155	80.7	8446	81.4	20759	59.4
Yes	25	13.0	1850	17.8	13906	39.8
Unknown	12	6.3	85	0.8	257	0.7

Notes. TL = thyroid lobectomy; TT = total thyroidectomy; NHW = non-Hispanic white; NHB = non-Hispanic black; PTMC = papillary thyroid microcarcinoma; FTMC = follicular thyroid microcarcinoma; HCTMC = Hürthle cell thyroid microcarcinoma; SD = standard deviation; NOS = not otherwise specified

lobectomy, removal of less than a lobe, or removal of a lobe and partial removal of the contralateral lobe.

[b] The TT group included patients who underwent near total, subtotal, or total thyroidectomy.

Table 2. Odds ratio estimates for univariate polytomous logistic regressions

Variable	Levels	No surgery vs TT		TL vs TT	
		OR	95% CI	OR	95% CI
Age	≤ 44 years	Ref	-	Ref	-
	45-54 years	1.02	(0.70, 1.47)	1.14	(1.08, 1.21)
	55-64 years	0.99	(0.67, 1.47)	1.19	(1.13, 1.27)
	≥ 65 years	1.34	(0.89, 2.01)	1.42	(1.33, 1.52)
Gender	Male	Ref	-	Ref	-
	Female	0.68	(0.49, 0.96)	0.83	(0.78, 0.87)
Race	NHW	Ref	-	Ref	-
	NHB	0.66	(0.32, 1.35)	0.95	(0.87, 1.04)
	Hispanic	0.93	(0.59, 1.47)	0.96	(0.91, 1.03)
	Other	1.79	(1.20, 2.67)	1.03	(0.95, 1.11)
Insurance	< 65 years				
	Uninsured	Ref	-	Ref	-
	Insured	0.51	(0.22, 1.16)	0.99	(0.83, 1.18)
	≥ 65 years				
	Uninsured	Ref	-	Ref	-
	Insured	-[a]	-[a]	0.95	(0.44, 2.05)
Year of diagnosis	2004-2007	Ref	-	Ref	-
	2008-2011	1.47	(0.89, 2.43)	0.84	(0.79, 0.89)
	2012-2016	2.57	(1.63, 4.04)	0.87	(0.82, 0.92)
Histology	PTMC	Ref	-	Ref	-
	FTMC	0.59	(0.08, 4.16)	1.30	(1.05, 1.60)
	HCTMC	1.37	(0.19, 9.81)	0.89	(0.61, 1.29)
Tumor size	≤ 6 mm	Ref	-	Ref	-
	> 6 mm	2.21	(1.64, 2.97)	0.39	(0.37, 0.41)
Stage	Localized	Ref	-	Ref	-
	Reginal/Distant	0.40	(0.24, 0.68)	0.16	(0.14, 0.18)
Multifocal	No	Ref	-	Ref	-
	Yes	0.24	(0.16, 0.37)	0.33	(0.31, 0.35)

Notes. TL = thyroid lobectomy; TT = total thyroidectomy; OR = odds ratio; CI = confidence interval; NHW = non-Hispanic white; NHB = non-Hispanic black; PTMC = papillary thyroid microcarcinoma; FTMC = follicular thyroid microcarcinoma; HCTMC = Hürthle cell thyroid microcarcinoma

[a] The OR and corresponding 95% CI were not calculated due to the sparse number of cases.

Table 3. Odds ratio estimates for multivariate polytomous logistic regressions

Covariate	Levels	No surgery vs TT		TL vs TT	
		OR	95% CI	OR	95% CI
Age	≤ 44 years	Ref	-	Ref	-
	45-54 years	0.95	(0.64, 1.41)	1.04	(0.98, 1.11)
	55-64 years	0.92	(0.60, 1.40)	1.02	(0.96, 1.09)
	≥ 65 years	1.33	(0.87, 2.04)	1.19	(1.11, 1.27)
Gender	Male	Ref	-	Ref	-
	Female	0.61	(0.42, 0.87)	0.73	(0.68, 0.77)
Race	NHW	Ref	-	Ref	-
	NHB	0.48	(0.21, 1.10)	0.81	(0.74, 0.89)
	Hispanic	0.94	(0.59, 1.50)	0.99	(0.92, 1.07)
	Other	1.77	(1.17, 2.68)	1.18	(1.09, 1.28)
Year of diagnosis	2004-2007	Ref	-	Ref	-
	2008-2011	1.48	(0.86, 2.55)	0.83	(0.78, 0.89)
	2012-2016	2.64	(1.62, 4.31)	0.87	(0.82, 0.92)
Histology	PTMC	Ref	-	Ref	-
	FTMC	0.44	(0.06, 3.16)	1.19	(0.95, 1.50)
	HCTMC	0.97	(0.13, 7.08)	0.87	(0.58, 1.30)
Tumor size	≤ 6 mm	Ref	-	Ref	-
	> 6 mm	2.81	(2.05, 3.86)	0.49	(0.46, 0.51)
Stage	Localized	Ref	-	Ref	-
	Reginal/Distant	0.30	(0.17, 0.55)	0.23	(0.20, 0.25)
Multifocal	No	Ref	-	Ref	-
	Yes	0.22	(0.14, 0.34)	0.40	(0.38, 0.42)

Notes. TL = thyroid lobectomy; TT = total thyroidectomy; OR = odds ratio; CI = confidence interval; NHW = non-Hispanic white; NHB = non-Hispanic black; PTMC = papillary thyroid microcarcinoma; FTMC = follicular thyroid microcarcinoma; HCTMC = Hürthle cell thyroid microcarcinoma

Table 4. Odds ratio estimates for subgroup multivariate logistic regressions among patients with localized thyroid microcarcinoma and treated with surgery (N = 37,425)

Covariate	Levels	TL vs TT	
		OR	95% CI
Age	≤ 44 years	Ref	-
	45-54 years	1.02	(0.96, 1.09)
	55-64 years	1.00	(0.94, 1.07)
	≥ 65 years	1.15	(1.07, 1.24)
Gender	Male	Ref	-
	Female	0.71	(0.66, 0.75)
Race	NHW	Ref	-
	NHB	0.80	(0.73, 0.88)
	Hispanic	1.00	(0.92, 1.07)
	Other	1.18	(1.09, 1.28)
Year of diagnosis	2004-2007	Ref	-
	2008-2011	0.84	(0.79, 0.90)
	2012-2016	0.87	(0.82, 0.93)
Histology	PTMC	Ref	-
	FTMC	1.15	(0.91, 1.45)
	HCTMC	0.90	(0.60, 1.34)
Tumor size	≤ 6 mm	Ref	-
	> 6 mm	0.48	(0.45, 0.50)
Multifocal	No	Ref	-
	Yes	0.40	(0.37, 0.42)

Notes. TL = thyroid lobectomy; TT = total thyroidectomy; OR = odds ratio; CI = confidence interval; NHW = non-Hispanic white; NHB = non-Hispanic black; PTMC = papillary thyroid microcarcinoma; FTMC = follicular thyroid microcarcinoma; HCTMC = Hürthle cell thyroid microcarcinoma

Table 5. Hazard ratio estimates in Cox model based on death due to thyroid cancer by stage and surgery type

Stage	Surgery	HR	95% CI
Localized	TT	Ref.	-
	TL	0.88	(0.48, 1.60)
	No surgery	-	-
Regional/Distant	TT	Ref.	-
	TL	1.85	(0.86, 3.95)
	No surgery	15.09	(3.48, 65.34)

Notes. TL = thyroid lobectomy; TT = total thyroidectomy; CI = confidence interval

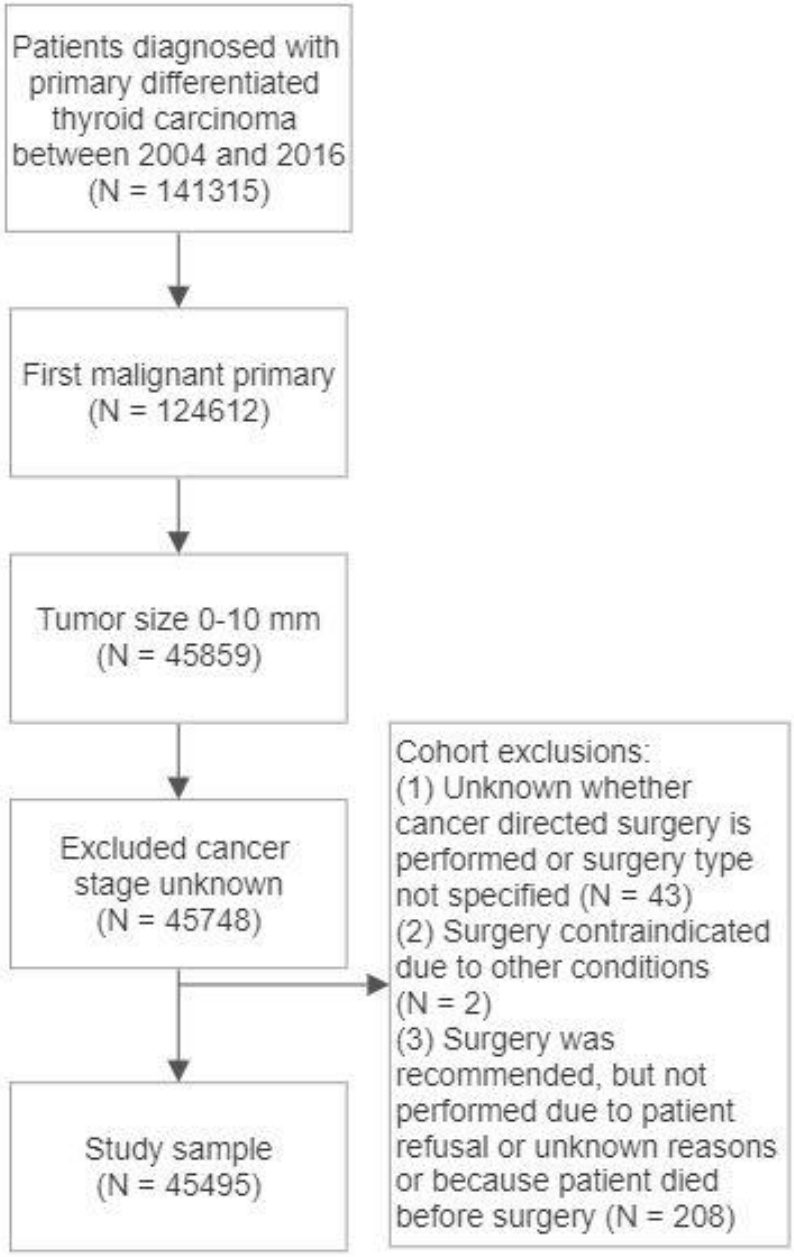

Figure 1. Flow chart of patient selection

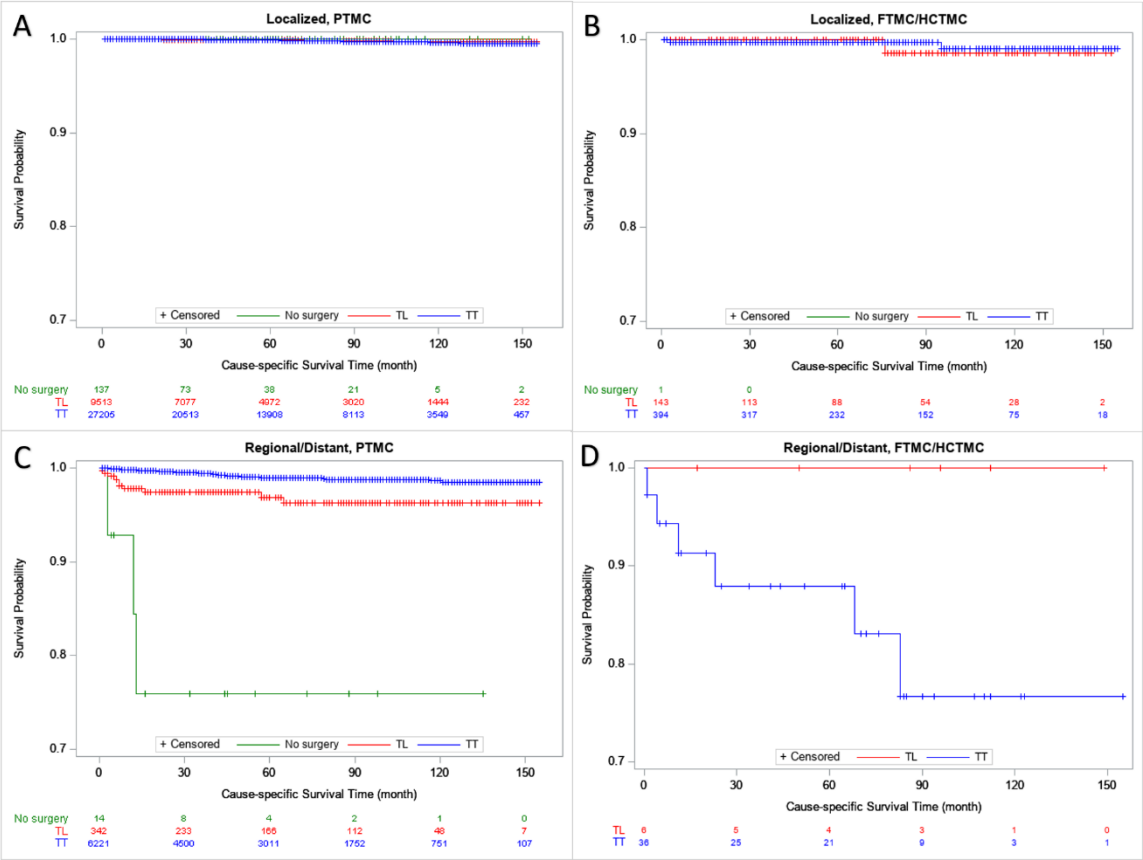

Figure 2. Cause-specific survival by surgery group for patients with (A) localized papillary thyroid microcarcinoma (PTMC), (B) localized follicular thyroid microcarcinoma (FTMC) or Hürthle cell thyroid microcarcinoma (HCTMC), (C) regional/distant PTMC, (D) regional/distant FTMC or HCTMC. For each plot, the blue line represents the thyroidectomy group, the red line represents the thyroid lobectomy group, and the green line represents the no surgery group. The table below each plot listed the number of people at risk at the corresponding time point. Survival was not estimated for patients with regional/distant FTMC or HCTMC who underwent no surgery due to lack of data.

Appendix

SAS Code for Data Organization

```
filename in1 'C:\Users\fishball\Desktop\DW\thesis\thesis3.txt';

proc format;
 value Age_recode_with_1_year_oldsf
  0 = "00 years"
  1 = "01-04 years"
  2 = "05-09 years"
  3 = "10-14 years"
  4 = "15-19 years"
  5 = "20-24 years"
  6 = "25-29 years"
  7 = "30-34 years"
  8 = "35-39 years"
  9 = "40-44 years"
  10 = "45-49 years"
  11 = "50-54 years"
  12 = "55-59 years"
  13 = "60-64 years"
  14 = "65-69 years"
  15 = "70-74 years"
  16 = "75-79 years"
  17 = "80-84 years"
  18 = "85+ years"
  29 = "Unknown"
  ;
 value Race_recode_White_Black_Otherf
  1 = "White"
  2 = "Black"
  3 = "Other (American Indian/AK Native, Asian/Pacific Islander)"
  9 = "Unknown"
  ;
 value RaceandoriginrecodeNHWNHBNHAIANf
  1 = "Non-Hispanic White"
  2 = "Non-Hispanic Black"
  3 = "Non-Hispanic American Indian/Alaska Native"
  4 = "Non-Hispanic Asian or Pacific Islander"
  5 = "Hispanic (All Races)"
  9 = "Non-Hispanic Unknown Race"
  ;
 value Sexf
  1 = "Male"
  2 = "Female"
  ;
 value Year_of_diagnosisf
  204 = "2004"
  205 = "2005"
  206 = "2006"
  207 = "2007"
  208 = "2008"
  209 = "2009"
```

```
 210 = "2010"
 211 = "2011"
 212 = "2012"
 213 = "2013"
 214 = "2014"
 215 = "2015"
 216 = "2016"
 ;
value Summary_stage_2000_1998f
 0 = "In situ"
 1 = "Localized"
 2 = "Regional"
 7 = "Distant"
 8 = "N/A"
 9 = "Unknown/unstaged"
 14 = "Blank(s)"
 ;
value SEERCombinedSummaryStage2000200f
 0 = "In situ"
 1 = "Localized only"
 2 = "Regional by direct extension only"
 3 = "Regional lymph nodes involved only"
 4 = "Regional by both direct extension and lymph node involvement"
 5 = "Regional, NOS"
 7 = "Distant site(s)/node(s) involved"
 8 = "Not applicable"
 9 = "Unknown/unstaged/unspecified/DCO"
 14 = "Blank(s)"
 ;
value CS_tumor_size_2004_2015f
 0-999 = "* 000-999"
 1022 = "Blank(s)"
 ;
value Tumor_Size_Summary_2016f
 0-990 = "* 000-990"
 998 = "998"
 999 = "999"
 1022 = "Blank(s)"
 ;
value RX_Summ_Surg_Prim_Site_1998f
 0-99 = "* 00-99"
 126 = "Blank(s)"
 ;
value Reasonnocancer_directed_surgeryf
 0 = "Surgery performed"
 1 = "Not recommended"
 2 = "Not recommended, contraindicated due to other cond; autopsy only (1973-2002)"
 5 = "Not performed, patient died prior to recommended surgery"
 6 = "Recommended but not performed, unknown reason"
 7 = "Recommended but not performed, patient refused"
 8 = "Recommended, unknown if performed"
 9 = "Unknown; death certificate; or autopsy only (2003+)"
 ;
```

```
value RX_Summ_Scope_Reg_LN_Sur_2003f
  0 = "None"
  1 = "Biopsy or aspiration of regional lymph node, NOS"
  2 = "Sentinel lymph node biopsy"
  3 = "Number of regional lymph nodes removed unknown"
  4 = "1 to 3 regional lymph nodes removed"
  5 = "4 or more regional lymph nodes removed"
  6 = "Sentinel node biopsy and lym nd removed same/unstated time"
  7 = "Sentinel node biopsy and lym nd removed different times"
  9 = "Unknown or not applicable"
  14 = "Blank(s)"
  ;
value RX_Summ_Surg_Oth_Reg_Dis_2003f
  0 = "None; diagnosed at autopsy"
  1 = "Non-primary surgical procedure performed"
  2 = "Non-primary surgical procedure to other regional sites"
  3 = "Non-primary surgical procedure to distant lymph node(s)"
  4 = "Non-primary surgical procedure to distant site"
  5 = "Any combo of sur proc to oth rg, dis lym nd, and/or dis site"
  9 = "Unknown; death certificate only"
  14 = "Blank(s)"
  ;
value Survival_monthsf
  0-503 = "* 0000-0503"
  9999 = "Unknown"
  ;
value Survival_months_flagf
  0 = "Complete dates are available and there are 0 days of survival"
  1 = "Complete dates are available and there are more than 0 days of survival"
  2 = "Incomplete dates are available and there could be zero days of follow-up"
  3 = "Incomplete dates are available and there cannot be zero days of follow-up"
  8 = "Not calculated because a Death Certificate Only or Autopsy Only case"
  ;
value Vitalstatusrecodestudycutoffusef
  1 = "Alive"
  0 = "Dead"
  ;
value SEERcausespecificdeathclassificf
  0 = "Alive or dead of other cause"
  1 = "Dead (attributable to this cancer dx)"
  8 = "Dead (missing/unknown COD)"
  9 = "N/A not first tumor"
  ;
value Insurance_Recode_2007f
  1 = "Uninsured"
  2 = "Any Medicaid"
  3 = "Insured"
  4 = "Insured/No specifics"
  5 = "Insurance status unknown"
  14 = "Blank(s)"
  ;
value Firstmalignantprimary_indicatorf
  0 = "No"
```

```
  1 = "Yes"
  ;
 value SEER_registryf
  1 = "San Francisco-Oakland SMSA - 1975+"
  2 = "Connecticut - 1975+"
  20 = "Detroit (Metropolitan) - 1975+"
  21 = "Hawaii - 1975+"
  22 = "Iowa - 1975+"
  23 = "New Mexico - 1975+"
  25 = "Seattle (Puget Sound) - 1975+"
  26 = "Utah - 1975+"
  27 = "Atlanta (Metropolitan) - 1975+"
  31 = "San Jose-Monterey - 1992+"
  35 = "Los Angeles - 1992+"
  29 = "Alaska Natives - 1992+"
  37 = "Rural Georgia - 1992+"
  41 = "California excluding SF/SJM/LA - 2000+"
  42 = "Kentucky - 2000+"
  43 = "Louisiana - 2000+"
  44 = "New Jersey - 2000+"
  47 = "Greater Georgia - 2000+"
  ;
        value surgeryf
  0 = "No surgery"
        1 = "TL"
        2 = "TT"
        ;
run;

data dtc;
 /*NOTE: The data file was created using the Windows format line delimiter.*/
 /*The TERMSTR=CRLF input option for reading the file in UNIX, requires SAS version 9.*/
 infile in1 LRECL = 32000 delimiter = ',' TERMSTR = CRLF;

 input Age_recode_with_1_year_olds
  Race_recode_White_Black_Other
  RaceandoriginrecodeNHWNHBNHAIAN
  Sex
  Year_of_diagnosis
  Histologic_Type_ICD_O_3
  Summary_stage_2000_1998
  SEERCombinedSummaryStage2000200
  CS_tumor_size_2004_2015
  Tumor_Size_Summary_2016
  RX_Summ_Surg_Prim_Site_1998
  Reasonnocancer_directed_surgery
  RX_Summ_Scope_Reg_LN_Sur_2003
  RX_Summ_Surg_Oth_Reg_Dis_2003
  Survival_months
  Survival_months_flag
  Vitalstatusrecodestudycutoffuse
  SEERcausespecificdeathclassific
  Insurance_Recode_2007
```

```
   Firstmalignantprimary_indicator
   CSsitespecificfactor12004varyin
   SEER_registry
   ;
  label Age_recode_with_1_year_olds = "Age recode with <1 year olds"
   Race_recode_White_Black_Other = "Race recode (White, Black, Other)"
   RaceandoriginrecodeNHWNHBNHAIAN = "Race and origin recode (NHW, NHB, NHAIAN, NHAPI,
Hispanic)"
   Sex = "Sex"
   Year_of_diagnosis = "Year of diagnosis"
   Histologic_Type_ICD_O_3 = "Histologic Type ICD-O-3"
   Summary_stage_2000_1998 = "Summary stage 2000 (1998+)"
   SEERCombinedSummaryStage2000200 = "SEER Combined Summary Stage 2000 (2004+)"
   CS_tumor_size_2004_2015 = "CS tumor size (2004-2015)"
   Tumor_Size_Summary_2016 = "Tumor Size Summary (2016+)"
   RX_Summ_Surg_Prim_Site_1998 = "RX Summ--Surg Prim Site (1998+)"
   Reasonnocancer_directed_surgery = "Reason no cancer-directed surgery"
   RX_Summ_Scope_Reg_LN_Sur_2003 = "RX Summ--Scope Reg LN Sur (2003+)"
   RX_Summ_Surg_Oth_Reg_Dis_2003 = "RX Summ--Surg Oth Reg/Dis (2003+)"
   Survival_months = "Survival months"
   Survival_months_flag = "Survival months flag"
   Vitalstatusrecodestudycutoffuse = "Vital status recode (study cutoff used)"
   SEERcausespecificdeathclassific = "SEER cause-specific death classification"
   Insurance_Recode_2007 = "Insurance Recode (2007+)"
   Firstmalignantprimary_indicator = "First malignant primary indicator"
   CSsitespecificfactor12004varyin = "CS site-specific factor 1 (2004+ varying by schema)"
   SEER_registry = "SEER registry"
   ;
  format Age_recode_with_1_year_olds Age_recode_with_1_year_oldsf.
   Race_recode_White_Black_Other Race_recode_White_Black_Otherf.
   RaceandoriginrecodeNHWNHBNHAIAN RaceandoriginrecodeNHWNHBNHAIANf.
   Sex Sexf.
   Year_of_diagnosis Year_of_diagnosisf.
   Summary_stage_2000_1998 Summary_stage_2000_1998f.
   SEERCombinedSummaryStage2000200 SEERCombinedSummaryStage2000200f.
   CS_tumor_size_2004_2015 CS_tumor_size_2004_2015f.
   Tumor_Size_Summary_2016 Tumor_Size_Summary_2016f.
   Reasonnocancer_directed_surgery Reasonnocancer_directed_surgeryf.
   RX_Summ_Scope_Reg_LN_Sur_2003 RX_Summ_Scope_Reg_LN_Sur_2003f.
   RX_Summ_Surg_Oth_Reg_Dis_2003 RX_Summ_Surg_Oth_Reg_Dis_2003f.
   Survival_months_flag Survival_months_flagf.
   Vitalstatusrecodestudycutoffuse Vitalstatusrecodestudycutoffusef.
   SEERcausespecificdeathclassific SEERcausespecificdeathclassificf.
   Insurance_Recode_2007 Insurance_Recode_2007f.
   Firstmalignantprimary_indicator Firstmalignantprimary_indicatorf.
   SEER_registry SEER_registryf.
   ;
run;

************************************
*Data Cleaning & Descriptive Analysis*
************************************
;
```

```
proc contents data=dtc varnum;
run;

*Patient selection;
data dtc1;
        set dtc;
        /*1. delete not first malignant primary*/
        if Firstmalignantprimary_indicator=1;
        /*Select patients with tumor size <= 1*/
        if 1 <= CS_tumor_size_2004_2015 <= 10
          or CS_tumor_size_2004_2015 in (990, 991)
          or 1 <= Tumor_Size_Summary_2016 <= 10
          or Tumor_Size_Summary_2016 = 990;
        /*2. delete cancer stage unknown*/
        if Summary_stage_2000_1998 in (0,1,2,7);
        /*3. delete patients wirh unrelated or unknown surgery codes*/
        if RX_Summ_Surg_Prim_Site_1998 in (90,99) then delete;
        /*4. delete patients with no surgery for reasons including:
          1) contraindicated due to other conditions;
          2) died before recommended surgery;
          3) patient or patient's guardian refused;
          4) unknown*/
        if Reasonnocancer_directed_surgery in (2,5,6,7,8,9) then delete;
run;

*Check missing values;
proc iml;
        use dtc1;
        read all var _NUM_ into x[colname=Names];
        n = countn(x,"col");
        nmiss = countmiss(x,"col");
        rNames = {"    Missing", "Not Missing"};
        cnt = (nmiss // n);
        print cnt[r=rNames c=Names label=""];

*Surgery;
data dtc1;
        set dtc1;
        /*no surgery*/
        if RX_Summ_Surg_Prim_Site_1998=0 then surgery1=0;
        /*lobectomy or less or one lobe+partial contralateral lobe*/
        if RX_Summ_Surg_Prim_Site_1998 in (13,20,21,22,23,25,26,27,30) then surgery1=1;
        /*near/sub/total thyroidectomy*/
        if RX_Summ_Surg_Prim_Site_1998 in (40,50,80) then surgery1=2;
        format surgery1 surgeryf.;
run;

*Age (4 groups);
proc freq data=dtc1;
        tables Age_recode_with_1_year_olds;
run;
data dtc2;
        set dtc1;
```

```
        if 0<Age_recode_with_1_year_olds<=9 then age4=1;
        if 9<Age_recode_with_1_year_olds<=11 then age4=2;
        if 11<Age_recode_with_1_year_olds<=13 then age4=3;
        if Age_recode_with_1_year_olds>13 then age4=4;
run;
proc freq data=dtc2;
        tables age4*surgery1/ nopercent norow;
run;

*Race;
proc freq data=dtc2;
        tables Race_recode_White_Black_Other RaceandoriginrecodeNHWNHBNHAIAN;
run;
data dtc3;
        set dtc2;
        if Race_recode_White_Black_Other=9 then Race_recode_White_Black_Other=.;
          /*set unknown age to missing*/
        if RaceandoriginrecodeNHWNHBNHAIAN=1 then race="NHW      ";
        if RaceandoriginrecodeNHWNHBNHAIAN=2 then race="NWB      ";
        if RaceandoriginrecodeNHWNHBNHAIAN=5 then race="Hisp     ";
        if RaceandoriginrecodeNHWNHBNHAIAN in (3,4,9) then race="NH other";
        if RaceandoriginrecodeNHWNHBNHAIAN=9 then race="";
run;
proc freq data=dtc3;
        tables race*surgery1/ nopercent norow missing;
run;

*Sex;
proc freq data=dtc3;
tables sex*surgery1/ nopercent norow missing;
run;

*Year of diagnosis;
proc freq data=dtc3;
        tables Year_of_diagnosis;
run;
data dtc3;
        set dtc3;
        if Year_of_diagnosis<208 then year_dx=1;
        else if Year_of_diagnosis<212 then year_dx=2;
        else year_dx=3;
run;
proc freq data=dtc3;
        tables year_dx*surgery1/ nopercent norow missing;
run;

*Histologic type ICD-O-3;
proc freq data=dtc3;
        tables Histologic_Type_ICD_O_3;
run;
data dtc4;
        set dtc3;
        if Histologic_Type_ICD_O_3 in (50,260,340,341,342,343,344,450) then histology=1;
```

```
        if Histologic_Type_ICD_O_3 in (330,331,332,335,37) then histology=2;
        if Histologic_Type_ICD_O_3 = 290 then histology=3;
run;
proc freq data=dtc4;
        tables histology*surgery1/ nopercent norow missing;
run;

*Stage1998+;
proc freq data=dtc4;
        tables Summary_stage_2000_1998;
run;
data dtc5;
        set dtc4;
        if Summary_stage_2000_1998 in (8,9,14) then Summary_stage_2000_1998=.;
          /*set unknown or unstaged or blanks to missing*/
        if Summary_stage_2000_1998=1 then stage="early";
        if Summary_stage_2000_1998 in (2,7) then stage="late";
        if Summary_stage_2000_1998=. then stage="";
run;
proc freq data=dtc5;
        tables stage*surgery1/ nopercent norow missing;
run;

*Stage2004+;
proc freq data=dtc5;
        tables SEERCombinedSummaryStage2000200;
run;
data dtc6;
        set dtc5;
        if SEERCombinedSummaryStage2000200 in (5,8,9,14) then
SEERCombinedSummaryStage2000200=.;
          /*set unknown or unstaged or unspecified to missing*/
run;

*Tumor size;
proc freq data=dtc6;
        tables CS_tumor_size_2004_2015*Tumor_Size_Summary_2016;
run;
data dtc7;
        set dtc6;
        if CS_tumor_size_2004_2015=1022 then size=Tumor_Size_Summary_2016;
        else size=CS_tumor_size_2004_2015;/*combine CS 2004-2015 and Tumor Size Summary 2016*/
        if size=990 then size=0;/*set foci to size 0 mm*/
        if size=991 then size=.;/*set <1cm but unknown size to missing*/
        if size<=6 then size_cat="<=6";/*categorical tumor size*/
        if size> 6 then size_cat=">6";
  if size=.  then size_cat="";
run;
proc means data=dtc7 n nmiss mean std median q1 q3 min max;
        var size;
        class size_cat;
run;
proc means data=dtc7 n nmiss mean std median q1 q3 min max;
```

```
        var size;
        class surgery1;
run;
proc freq data=dtc7;
        tables size_cat*surgery1/ nopercent norow missing;
run;

*Survival time;
proc means data=dtc7 n nmiss mean median min max;
        var Survival_months;
run;

*Survival time flag;
proc freq data=dtc7;
tables Survival_months_flag;
run;
/*set survival time to missing for those with insufficient follow-up or unclear follow-up time: flag=0, 2, 8*/
data dtc8;
        set dtc7;
        if Survival_months_flag in (0,2,8) then Survival_months=.;
        if Survival_months=0 then Survival_months=.;
run;

*All cause desth;
proc freq data=dtc8;
        tables Vitalstatusrecodestudycutoffuse;
run;

*Cause-specific death;
proc freq data=dtc8;
        tables SEERcausespecificdeathclassific Firstmalignantprimary_indicator;
run;
data dtc9;
set dtc8;
        /*unknown cause of death to missing*/
        if SEERcausespecificdeathclassific in (8,9) then SEERcausespecificdeathclassific=.;
run;

*Insurance status;
proc freq data=dtc9;
tables Insurance_Recode_2007;
run;
data dtc10;
        set dtc9;
        /*unknown or blanks to missing*/
        if Insurance_Recode_2007 in (5,14) then Insurance="       ";
        /*combine insured*/
        if Insurance_Recode_2007 in (2,3,4) then Insurance="insured";
        if Insurance_Recode_2007=1 then insurance="uninsured";
run;
proc freq data=dtc10;
        tables insurance*surgery1/ nopercent norow missing;
        where age4<4;
```

```
run;
proc freq data=dtc10;
        tables insurance*surgery1/ nopercent norow missing;
        where age4=4;
run;

data dtc10;
set dtc10;
if age4=4 then age65=1;
else age65=0;
run;
proc freq data=dtc10;
        tables Insurance_Recode_2007*age65/ nopercent norow missing;
run;

*Multifocality;
proc freq data=dtc10;
tables CSsitespecificfactor12004varyin;
run;
data dtc10;
        set dtc10;
        if CSsitespecificfactor12004varyin=10 then multifocal=0;
        if CSsitespecificfactor12004varyin=20 then multifocal=1;
        /*unknown to missing*/
        if CSsitespecificfactor12004varyin in (0,999) then multifocal=.;
run;
proc freq data=dtc10;
        tables multifocal*surgery1/ nopercent norow missing;
run;
```

SAS Code for Statistic Analysis

```
**********************
*Bivariate association*
**********************
*Polytomous logistic regression;
*Age group (ordinal);
proc freq data=dtc10;
tables surgery1;
run;
proc logistic data=dtc10;
class age4 (ref="1");
model surgery1(ref="TT")=age4/link=glogit;
run;

*Sex;
proc logistic data=dtc10;
class sex;
model surgery1(ref="TT")=sex/link=glogit;
run;

*Race;
```

```
proc logistic data=dtc10;
class race (ref="NHW");
model surgery1(ref="TT")=race/link=glogit;
run;

*Insurance;
proc logistic data=dtc10;
class insurance;
model surgery1(ref="TT")=insurance/link=glogit;
run;
proc logistic data=dtc10;
class insurance;
model surgery1(ref="TT")=insurance/link=glogit;
where age4<4;/*stratify on age<65*/
run;
proc logistic data=dtc10;
class insurance;
model surgery1(ref="TT")=insurance/link=glogit;
where age4=4;/*stratify on age>=65*/
run;

*Year of diagnosis (ordinal);
proc logistic data=dtc10;
class year_dx (ref="1");
model surgery1(ref="TT")=year_dx/link=glogit;
run;

*Histology;
proc logistic data=dtc10;
class histology (ref="1");
model surgery1(ref="TT")=histology/link=glogit;
run;

*Tumor size;
proc logistic data=dtc10;
class size_cat (ref="<=6");
model surgery1(ref="TT")=size_cat/link=glogit;
run;

*SEER stage;
proc logistic data=dtc10;
class stage (ref="early");
model surgery1(ref="TT")=stage/link=glogit;
run;

*Multifocal;
proc logistic data=dtc10;
class multifocal (ref="0");
model surgery1(ref="TT")=multifocal/link=glogit;
run;

************************
*Multivariable regression*
```

```
*************************
*Polytomous logistic regression;
proc logistic data=dtc10;
class age4 (ref="1") sex race (ref="NHW") year_dx (ref="1")
            histology (ref="1") size_cat (ref="<=6")
            stage (ref="early") multifocal (ref="0");
model surgery1(ref="TT")=age4 sex race year_dx histology
    size_cat stage multifocal/link=glogit;
run;
/*Restricted to localized patients*/
proc logistic data=dtc10;
class age4 (ref="1") sex race (ref="NHW") year_dx (ref="1")
            histology (ref="1") size_cat (ref="<=6")
            multifocal (ref="0");
model surgery1(ref="TT")=age4 sex race year_dx histology
    size_cat multifocal/link=glogit;
where stage="early" & surgery1>0;
run;

*******************
*Survival analysis*
*******************
;
*Kaplan Meier curves;
*Overall survival;
%include "C:\Users\fishball\Desktop\DW\thesis\macro\KMmacro.sas";
%ProvideSurvivalMacros
/*Localized, PTMC*/
%let tatters    = textattrs=(size=14pt weight=bold family='arial');
%let TitleText0 = "Localized, PTMC";
%let TitleText1 = &titletext0 " for " STRATUMID;
%let TitleText2 = &titletext0;
%let ntitles = 1;
/*Line thickness*/
%let StepOpts = lineattrs=(thickness=1.5);
/*Line color*/
%let GraphOpts = DataContrastColors=(green red blue)
            DataColors=(green red blue);
%let Censored   =markerattrs=(symbol=plus);
/*Legend*/
%let InsetOpts  = ;
%let LegendOpts = title="+ Censored" location=inside autoalign=(Bottom);
/*yaxis*/
%let yOptions   = label="Survival Probability"
            shortlabel="Survival"
            labelattrs=(size=10pt)
            tickvalueattrs=(size=10pt)
            linearopts=(viewmin=0.7 viewmax=1
                    tickvaluelist=(.7 .8 .9 1.0));
/*xaxis*/
%let xOptions = label="Cause-specific Survival Time (month)"
            labelattrs=(size=10pt)
            tickvalueattrs=(size=10pt)
```

```
        linearopts=(tickvaluelist=(0 30 60 90 120 150));
%CompileSurvivalTemplates          /* Compile the templates*/

proc lifetest data=dtc10 method=km plots=survival(test atrisk(outside maxlen=13)=(0 to 150 by 30))
notable;
time Survival_months*SEERcausespecificdeathclassific(0);
strata surgery1/adjust=tukey;
where Summary_stage_2000_1998=1 & histology=1;
run;

/*Localized, FTMC or HCTMC*/
%let TitleText0 = "Localized, FTMC/HCTMC";
%let TitleText1 = &titletext0 " for " STRATUMID;
%let TitleText2 = &titletext0;
%CompileSurvivalTemplates

proc lifetest data=dtc10 method=km plots=survival(test atrisk(outside maxlen=13)=(0 to 150 by 30))
notable;
time Survival_months*SEERcausespecificdeathclassific(0);
strata surgery1/adjust=tukey;
where Summary_stage_2000_1998=1 & histology>1;
run;

/*Regional and Distant, PTMC*/
%let TitleText0 = "Regional/Distant, PTMC";
%let TitleText1 = &titletext0 " for " STRATUMID;
%let TitleText2 = &titletext0;
%CompileSurvivalTemplates

proc lifetest data=dtc10 method=km plots=survival(test atrisk(outside maxlen=13)=(0 to 150 by 30));
time Survival_months*SEERcausespecificdeathclassific(0);
strata surgery1/adjust=tukey;
where Summary_stage_2000_1998>1 & histology=1;
run;

/*Regional and Distant, FTMC or HCTMC*/
%let TitleText0 = "Regional/Distant, FTMC/HCTMC";
%let TitleText1 = &titletext0 " for " STRATUMID;
%let TitleText2 = &titletext0;
%let GraphOpts = DataContrastColors=(red blue)
        DataColors=(red blue);
%CompileSurvivalTemplates

proc lifetest data=dtc10 method=lt plots=survival(test atrisk(outside maxlen=13)=(0 to 150 by 30)) ;
time Survival_months*SEERcausespecificdeathclassific(0);
strata surgery1/adjust=tukey;
where Summary_stage_2000_1998>1 & histology>1;
run;

*Surgery type;
*PH assumption;
*Graphical;
proc lifetest data=dtc10 method=lt plots=survival noleft;
```

```
time Survival_months*SEERcausespecificdeathclassific(0);
strata surgery1;
run;
proc lifetest data=dtc10 method=km plots=lls notable;
time Survival_months*SEERcausespecificdeathclassific(0);
strata surgery1;
run;
*GOF;
proc phreg data=dtc10;
class surgery1 age4 Race_recode_White_Black_Other year_dx
    sex histology Summary_stage_2000_1998;
model Survival_months*SEERcausespecificdeathclassific(0)=surgery1 age4
    Race_recode_White_Black_Other year_dx sex histology Summary_stage_2000_1998 size;
output out=resid ressch=sh_surgery10 sh_surgery11 sh_age41 sh_age42 sh_age43
    sh_Race1 sh_Race2
    sh_year_dx1 sh_year_dx2 sh_sex sh_histology1 sh_histology2
    sh_Summary_stage_2000_19981 sh_Summary_stage_2000_19982
    sh_size;
run;
data failures;
set resid;
where SEERcausespecificdeathclassific=1;
run;
proc rank data=failures out=ranked ties=mean;
var survival_months;
ranks timerank;
run;
proc corr data=ranked nosimple;
with timerank;
var sh_surgery10 sh_surgery11 sh_age41 sh_age42 sh_age43
    sh_Race1 sh_Race2
    sh_year_dx1 sh_year_dx2 sh_sex sh_histology1 sh_histology2
    sh_Summary_stage_2000_19981 sh_Summary_stage_2000_19982
    sh_size;
run;

*Cox analysis;
/*Stratified by stage*/
proc phreg data=dtc10;
class surgery1 age4 sex histology stage Race_recode_White_Black_Other year_dx multifocal;
model Survival_months*SEERcausespecificdeathclassific(0)=surgery1 age4
    Race_recode_White_Black_Other year_dx sex histology size multifocal
    surgery1*stage age4*stage Race_recode_White_Black_Other*stage
    year_dx*stage sex*stage histology*stage size*stage multifocal*stage/rl;
strata stage;
contrast "HR for early 1 vs 2" surgery1 0 1 surgery1*stage 0 1/estimate=exp;
contrast "HR for early 0 vs 2" surgery1 1 0 surgery1*stage 1 0/estimate=exp;
contrast "HR for late 1 vs 2" surgery1 0 1 surgery1*stage 0 0/estimate=exp;
contrast "HR for late 0 vs 2" surgery1 1 0 surgery1*stage 0 0/estimate=exp;
run;
```