2018

The Impact of Adjuvant Chemotherapy on Survival for Patients with Stage II Colon Cancer Portfolio

Li Ding  
*University of Kentucky, ldi234@uky.edu*

---

**Recommended Citation**  
Ding, Li, "The Impact of Adjuvant Chemotherapy on Survival for Patients with Stage II Colon Cancer Portfolio" (2018). *Theses and Dissertations--Public Health (M.P.H. & Dr.P.H.).* 218.  
https://uknowledge.uky.edu/cph_etds/218

---

This Graduate Capstone Project is brought to you for free and open access by the College of Public Health at UKnowledge. It has been accepted for inclusion in Theses and Dissertations--Public Health (M.P.H. & Dr.P.H.) by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
STUDENT AGREEMENT:

I represent that my thesis or dissertation and abstract are my original work. Proper attribution has been given to all outside sources. I understand that I am solely responsible for obtaining any needed copyright permissions. I have obtained and attached hereto needed written permission statements(s) from the owner(s) of each third-party copyrighted matter to be included in my work, allowing electronic distribution (if such use is not permitted by the fair use doctrine).

I hereby grant to The University of Kentucky and its agents the non-exclusive license to archive and make accessible my work in whole or in part in all forms of media, now or hereafter known. I agree that the document mentioned above may be made available immediately for worldwide access unless a preapproved embargo applies.

I retain all other ownership rights to the copyright of my work. I also retain the right to use in future works (such as articles or books) all or part of my work. I understand that I am free to register the copyright to my work.

REVIEW, APPROVAL AND ACCEPTANCE

The document mentioned above has been reviewed and accepted by the student’s advisor, on behalf of the advisory committee, and by the Director of Graduate Studies (DGS), on behalf of the program; we verify that this is the final, approved version of the student’s dissertation including all changes required by the advisory committee. The undersigned agree to abide by the statements above.

Li Ding, Student

Dr. Bin Huang, Major Professor

Dr. Sarah Wackerbarth, Director of Graduate Studies
The Impact of Adjuvant Chemotherapy on Survival for Patients with Stage II Colon Cancer Portfolio

Li Ding
MPH
Biostatistics
# Contents

1. PROJECT DESCRIPTION ........................................................................................................... 1

2. NOTEs ....................................................................................................................................... 2
   2.1 Patient selection algorithm in SEER dataset ........................................................................ 2
   2.2 High-risk features of recurrence in SEER dataset ............................................................... 2
   2.3 Risk group definition ........................................................................................................... 3
   2.4 Subgroups of high-risk group .............................................................................................. 3
   2.5 Characteristics of patients ................................................................................................. 4
   2.6 Chemotherapy Records ..................................................................................................... 4

3. SAS CODE ................................................................................................................................ 5
   3.1 Data Input and Format ........................................................................................................ 5
   3.2 Describe the Characteristics of Patients ............................................................................ 7
   3.3 Survival Analysis .............................................................................................................. 8

4. OUTPUTS ................................................................................................................................... 13
   4.1 Frequency Table Outputs .................................................................................................. 13
   4.2 Survival analysis Outputs ................................................................................................. 18

5. RESULTS ANALYSIS ................................................................................................................ 27
   5.1 Frequency Table Results .................................................................................................. 27
   5.2 Survival Analysis Results ................................................................................................. 27

6. REPORTS (Analysis outcomes summary) ................................................................................ 28

7. LESSONS .................................................................................................................................. 28

8. PAPER DRAFT ............................................................................................................................ 29
   ABSTRACT ............................................................................................................................... 29
     Purpose ..................................................................................................................................... 29
     Patients and Methods ........................................................................................................... 29
     Results ...................................................................................................................................... 29
     Conclusion .............................................................................................................................. 29
   INTRODUCTION ......................................................................................................................... 29
   PATIENTS AND METHODS ........................................................................................................ 30
     Data source ............................................................................................................................ 30
     Study Population .................................................................................................................. 30
Variables ........................................................................................................................................... 31
Statistical Analysis ............................................................................................................................ 31
RESULTS ........................................................................................................................................... 32
Characteristics of Patients .............................................................................................................. 32
Survival Benefit of Chemotherapy ................................................................................................. 32
DISCUSSION .................................................................................................................................... 33
CONCLUSIONS ................................................................................................................................. 33
REFERENCES ................................................................................................................................... 39
9. Defense Presentation ..................................................................................................................... 40
1. PROJECT DESCRIPTION

It is controversial for patients with stage II colon cancer to use adjuvant chemotherapy after surgery. Although in theory, adjuvant chemotherapy can reduce the risk of cancer recurrence after surgery, many studies indicated that the improvement of adjuvant chemotherapy on survival is minimal for patients with stage II colon cancer. Adjuvant chemotherapy are not recommended to routine use in patients with stage II colon cancer; However, if patients present any high risk feature of recurrence, including T4 tumor, poorly/differentiated histology, lymphovascular invasion, perineural invasion, less than 12 lymph nodes were removed or could be assessed, positive margin, obstruction, or perforation, adjuvant chemotherapy can be considered to use, but it does not mean that adjuvant chemotherapy should be used for all stage II patients who present high-risk features. Patients with these features still need to talk with their doctors to decide if they can use chemotherapy or not. Until now, there are not uniform standards about which part of patients with stage II colon cancer should receive adjuvant chemotherapy. The purpose of this study was to explore if adjuvant chemotherapy can really improve the survival of patients with stage II colon cancer, especially for patients who presented any high-risk feature of recurrence.

A total of 23,354 patients with stage II colon cancer and received colon cancer primary site surgery were selected from SEER dataset based on patient selection algorithm (Note 1). Considering that some patients had bad health condition and died before they could receive adjuvant chemotherapy, we excluded patients who survived at less than 6 months after diagnosis. We found the information of five high-risk features in SEER dataset, including T4 tumor, poorly/differentiated histology, perineural invasion, less than 12 lymph nodes were removed or could be assessed, and positive margin involvement (Note 2). Patients were divided into low-risk group and high-risk group according to if they presented any high-risk feature (Note 3). High-risk group patients were further divided into five subgroups (Note 4). We used $x^2$ test to describe the characteristics of patients in low and high risk groups. These characteristics included five sociodemographic variables (age, gender, race, year of diagnosis, and urban/rural status) and five cancer-related variables (T stage, nodes examined, histology grades, margin involvement, and perineural invasion) (Note 5). Then, we used survival analysis (Kaplan-Meier test and Cox regression analysis) to compare the survival difference by chemotherapy record (Yes vs. No/Unknown) in each risk group. For survival analysis, the primary explanatory variable was chemotherapy. The primary outcome variable was overall survival. The primary statistical tool is SAS 9.0.

The univariate survival analysis reflected that no matter patients presented any high-risk feature or not, the survival rate of patients who received chemotherapy was obviously better than those who did not or unknown. This difference disappeared after multivariate survival analysis (patients’ sociodemographic characteristics were controlled). The possible reason was that younger patients were more likely to receive chemotherapy and younger patients had better survival than older patients. Chemotherapy did not really improve patients’ survival. We guessed that maybe for some special high-risk features, chemotherapy could improve patients’ survival, but for some other features, chemotherapy could not improve patients’ survival. After analysis, we found when patients presented T4 tumor feature, no matter they presented any other high-risk features or not, chemotherapy could improve their survival. For other four high-risk features, we could not decide if chemotherapy really improved patients’ survival.
2. NOTES

2.1 Patient selection algorithm in SEER dataset

<table>
<thead>
<tr>
<th>Variables</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colon cancer Primary Site</td>
<td>Site recode ICD-O-3/WHO 2008= Colon excluding Rectum; Primary Site= C18.0-18.9 Histology = 8140-8147,8210-8211,8220-8221,8260-8263,8480-8481,8490 First malignant primary indicator=Yes</td>
</tr>
<tr>
<td>Colon cancer stage</td>
<td>AJCC.Derived AJCC Stage Group, 7th ed (2010+)= 'II','IIINOS','IIA','IIIB','IIIC','IIIE','IIIEA','IIIEB','IIIEC','IIIESA','IIIESB','IIIES'</td>
</tr>
<tr>
<td>Colon cancer primary surgery</td>
<td>Surg Prim Site (1998+)=30-80</td>
</tr>
<tr>
<td>Age</td>
<td>&gt;=20</td>
</tr>
<tr>
<td>Diagnosis year</td>
<td>2010-2015</td>
</tr>
<tr>
<td>Race</td>
<td>All race</td>
</tr>
<tr>
<td>gender</td>
<td>Male and Female</td>
</tr>
<tr>
<td>Survival Month</td>
<td>Complete data are available and there are more than 0 days of survival Survival Month &gt;=6</td>
</tr>
</tbody>
</table>

2.2 High-risk features of recurrence in SEER dataset

<table>
<thead>
<tr>
<th>High risk feature</th>
<th>Seer code field</th>
<th>Seer code</th>
<th>Other definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher T stage (T4 versus T3)</td>
<td>Derived AJCC T, 6th ed (2004+)</td>
<td>T3 and T4</td>
<td></td>
</tr>
<tr>
<td>Less than 12 lymph nodes were removed or could be assessed.</td>
<td>Regional nodes examined (1988+)</td>
<td>00-99, Records the total number of regional lymph nodes that were removed and examined by the pathologist.</td>
<td></td>
</tr>
<tr>
<td>Poorly differentiated histology</td>
<td>Grade</td>
<td>Well differentiated; Grade I Moderately differentiated; Grade II Poorly differentiated; Grade III Undifferentiated; anaplastic; Grade IV</td>
<td></td>
</tr>
<tr>
<td>Presence of lymphovascular invasion</td>
<td>Lymph-vascular Invasion (2004+ varying by schema)</td>
<td>Blank(s)</td>
<td></td>
</tr>
<tr>
<td>bowel obstruction or bowel perforation</td>
<td>No</td>
<td></td>
<td>Obstruction: Medicare, ICD-9-CM Diagnosis</td>
</tr>
</tbody>
</table>
2.3 Risk group definition

<table>
<thead>
<tr>
<th>Low-risk group</th>
<th>High-risk group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patients presented features: “T3 stage” AND “&gt;=12 lymph nodes were removed or could be assessed” AND “Well or Moderately differentiated” AND “Negative margins” AND “No presence of perineural invasion”</td>
<td>Patients present features: “T4 stage” OR “&lt;12 lymph nodes were removed or could be assessed” OR “Poorly or Undifferentiated differentiated” OR “Positive margins” OR “Presence of perineural invasion”</td>
</tr>
</tbody>
</table>

2.4 Subgroups of high-risk group

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subgroup 1</td>
<td>Patients presented high-risk feature: T4 tumor feature</td>
</tr>
<tr>
<td>Subgroup 2</td>
<td>Patients presented high-risk feature: &lt;12 lymph nodes were removed or could be assessed</td>
</tr>
<tr>
<td>Subgroup 3</td>
<td>Patients presented high-risk feature: Poorly or undifferentiated histology</td>
</tr>
<tr>
<td>Subgroup 4</td>
<td>Patients presented high-risk feature: Positive Margin</td>
</tr>
<tr>
<td>Subgroup 5</td>
<td>Patients presented high-risk feature: Presence of perineural invasion</td>
</tr>
</tbody>
</table>
2.5 Characteristics of patients

| Sociodemographic characteristics (Control variables) | Age, Gender, Race, Year of Diagnosis, and Urban/Rural Status |
| Cancer-related characteristics (Stratification variables) | T stage, nodes examined, histology grades, margin involvement, and perineural invasion |

2.6 Chemotherapy Records

Because SEER dataset did not record the time of chemotherapy, we could not distinguish adjuvant and non-adjuvant of chemotherapy. However, considering that for stage II colon cancer, there were few patients received chemotherapy before surgery. We can consider all chemotherapy record in SEER dataset as adjuvant chemotherapy. We will use KCR data to check the percentage of adjuvant chemotherapy among all chemotherapy records.
3. SAS CODE

3.1 Data Input and Format

```sas
filename in1 'D:\Chemotherapy for stage II colon cancer\exportnew.txt';
ods rtf file='D:\Chemotherapy for stage II colon cancer\export.rtf';
*Input data;
data casedat;
infile in1 LRECL = 32000 delimiter = '09'X TERMSTR = CRLF;
length Patient_ID $19
    Site_recode_ICD_0_3_WHO_2008 $53
    RX_Summ_SurgPrimSite1998 $20
    Age_recode_with_1_year_old $11
    Race_recode_White_Black_Other $59
    Sex $15
    Year_of_Diagnosis $9
    RuralUrban_Continuum_Code_2013 $60
    DerivedAJCC_T_7thed2010 $9
    Regional_nodes_examined1998 $9
    Grade $9
    CRM6 $60
    Perineural_Invasion8 $9
    Survival_months $11
    Vitalstatusrecodestudycutoffus $5
    Chemotherapy_Recode $5;
/*NOTE: skipping over field names*/
if _N_ = 1 then input;
input Patient_ID $
    Site_recode_ICD_0_3_WHO_2008 $
    RX_Summ_SurgPrimSite1998 $
    Age_recode_with_1_year_old $
    Race_recode_White_Black_Other $
    Sex $
    Year_of_Diagnosis $
    RuralUrban_Continuum_Code_2013 $
    DerivedAJCC_T_7thed2010 $
    Regional_nodes_examined1998 $
    Grade $
    CRM6 $
    Perineural_Invasion8 $
    Survival_months $
    Vitalstatusrecodestudycutoffus $
    Chemotherapy_Recode $
;
* Label data variables;
label Patient_ID = "Patient ID"
    Site_recode_ICD_0_3_WHO_2008 = "Site recode ICD-O-3/WHO 2008"
    RX_Summ_SurgPrimSite1998 = "Surgery Record"
    Age_recode_with_1_year_old = "Age recode with <1 year olds"
    Race_recode_White_Black_Other = "Race recode (White, Black, Other)"
    Sex = "Sex"
    Year_of_Diagnosis = "Year of Diagnosis"
    RuralUrban_Continuum_Code_2013 = "Rural-Urban Continuum Code 2013"
    DerivedAJCC_T_7thed2010 = "T stage"
    Regional_nodes_examined1998 = "Lymph nodes examined"
```
 Grade="Histology grade"
CRM6="Tumor Margin Involvement"
Perineural_Invasion8="Perineural Invasion"
Survival_months = "Survival months"
Vital_status_recode = "Vital status recode (study cutoff used)"
Chemotherapy_Recode="Chemotherapy Record"

run;
*Foramt data value;

proc format;
value agegf 1='20-49' 2='50-64' 3='65-74' 4='75+';
value racef 1='White' 2='Black' 3='Other' 9='unknown';
value sexf 1='Male' 2='Female';
value yearf 1="2010-2011" 2="2012-2013" 3="2014-2015";
value metorf 1="Metro" 2="Rural" 3="Unknown";
value Tstagef 1='T3 ' 2='T4' 3='Unknown';
value Nodesf 1='<12' 2='>=12' 3='Unknown';
value gradef 1="Well or Moderately differentiated" 2="Poorly or Undifferentiated differentiated" 3='Unknown';
value marginf 1="Positive resection margin" 2="Negative resection margin" 3='unknown';
value invasionf 1="No perineural invasion present" 2="Perineural invasion present" 3='unknown';
value vitalf 0='Dead' 1='Alive';
value chemof 0='No/Unknown' 1='Yes';
value riskgroupf 1='low-risk' 2='high-risk' 3='unknown';
*Check the original dataset;
Title Original dataset;

proc print data=casedat (obs=10) label;
run;
*Define new variables;
data test;
set casedat;
if 5<=Age_recode_with_1_year_olds <=10 then ageg=1;
else if 11<=Age_recode_with_1_year_olds <=13 then ageg=2;
else if 14<=Age_recode_with_1_year_olds <=15 then ageg=3;
else if Age_recode_with_1_year_olds >=16 then ageg=4;
race=Race_recode_White_Black_Other*1;
gender=sex*1;
if Year_of_Diagnosis in (210,211) then year_group=1;
else if Year_of_Diagnosis in (212,213) then year_group=2;
else if Year_of_Diagnosis in (214,215) then year_group=3;
if Rural_Urban_Continuum_Code_2013 in (1,2,3) then metro=1;
else if Rural_Urban_Continuum_Code_2013 in (4,5,6,7,8,9) then metro=2;
else if metro=3;
if DerivedAJCC_T_7thed2010=300 then T_stage=1;
else if DerivedAJCC_T_7thed2010 in (410,420) then T_stage=2;
else if T_stage=3;

if 0<=Regional_nodes_examined1998<=11 then Lymph_Nodes =1;
else if 12<=Regional_nodes_examined1998<=90 then Lymph_Nodes =2;
else if Regional_nodes_examined1998>90 then Lymph_Nodes =3;

if Grade in (1,2) then Histology_Grade=1;
else if Grade in (3,4) then Histology_Grade=2;
else if Grade =9 then Histology_Grade=3;

if 0<=CRM6<=10 then Margin_involvement=1;
else if 10<CRM6<=996 then Margin_involvement=2;
else Margin_involvement=3;

if Perineural_Invasion8="0" then Perineural_Invasion=1;
else if Perineural_Invasion8="10" then Perineural_Invasion=2;
else Perineural_Invasion=3;

if T_stage=1 and Lymph_Nodes =2 and Histology_Grade=1 and Margin_involvement=2 and Perineural_Invasion=1 then Risk_group=1;
else if T_stage=2 or Lymph_Nodes =1 or Histology_Grade=2 or Margin_involvement=1 or Perineural_Invasion=2 then Risk_group=2;
else Risk_group=3;

month_survival=Survival_months*1;

if Vitalstatusrecodestudycutoffus='4' then vitalstatus=0;
else if Vitalstatusrecodestudycutoffus='1' then vitalstatus=1;

Chemotherapy=Chemotherapy_Recode*1;
run;
*Check the new dataset with new variables;
Title Dataset with new defined variables;
proc print data=test (obs=10) label;
var Patient_ID ageg race gender  year_group  metro  T_stage  Lymph_Nodes  Histology_Grade  Margin_involvement  Perineural_Invasion  vitalstatus  month_survival  Chemotherapy  Risk_group  ;
run;

### 3.2 Describe the Characteristics of Patients

*Descriptive Statistics;
Title1 Characteristics of patients undergoing surgery for stage II colon cancer;
Title2 Patients in each risk feature group;
proc freq data=test;table Risk_group *Chemotherapy / chisq; run;
Title2 Age distribution in low-risk features group;
proc freq data=test;where Risk_Group=1; table ageg *Chemotherapy / chisq; run;
Title2 Age distribution in high-risk features group;
proc freq data=test;where Risk_Group=2; table ageg *Chemotherapy / chisq; run;
Title2 Gender distribution in low-risk features group;
proc freq data=test;where Risk_Group=1; table gender *Chemotherapy / chisq; run;
**3.3 Survival Analysis**
*For patients in low-risk feature group, testing if chemo will influence their survival;  
*Kaplan-Meier test;  
Title Patients without any high-risk of recurrence feature;  
proc lifetest data=test notable method=km;  
time month_survival*Vitalstatus(1);  
Where Risk_group=1;  
strata Chemotherapy;  
run;  
*Kaplan-Meier test;  
PROC PHREG DATA=test;  
    class Chemotherapy (PARAM=REF REF='No/Unknown');  
Where Risk_group=1;  
MODEL month_survival*Vitalstatus(1)= Chemotherapy /rl;  
RUN;  
*adjusted hared ratio;  
PROC PHREG DATA=test;  
    CLASS ageg (PARAM=REF REF='20-49') race(PARAM=REF REF='White')  
gender(PARAM=REF REF='Female') year_group(PARAM=REF REF='2010-2011')  
metro(PARAM=REF REF='Rural') Chemotherapy (PARAM=REF REF='No/Unknown');  
Where Risk_group=1;  
MODEL month_survival*Vitalstatus(1)= ageg race gender year_group metro  
Chemotherapy /rl;  
RUN;  

*For patients with any high risk factor, testing if chemo will influence survival;  
*Kaplan-Meier test;  
Title Patients with any high-risk of recurrence feature;  
proc lifetest data=test notable method=km;  
time month_survival*Vitalstatus(1);  
Where Risk_group=2;  
strata Chemotherapy;  
run;  
*unadjusted hared ratio;  
PROC PHREG DATA=test;  
    CLASS Chemotherapy (PARAM=REF REF='No/Unknown');  
Where Risk_group=2;  
MODEL month_survival*Vitalstatus(1)= Chemotherapy /rl;  
RUN;  
*adjusted hared ratio;  
PROC PHREG DATA=test;  
    CLASS ageg (PARAM=REF REF='20-49') race(PARAM=REF REF='White')  
gender(PARAM=REF REF='Female') year_group(PARAM=REF REF='2010-2011')  
metro(PARAM=REF REF='Rural') Chemotherapy (PARAM=REF REF='No/Unknown');  
Where Risk_group=2;  
MODEL month_survival*Vitalstatus(1)= ageg race gender year_group metro  
Chemotherapy /rl;  
RUN;  

*For patients at least with high-risk feature T4 tumor, testing if chemo will influence survival;  
*Kaplan-Meier test;  
Title Patients at least with high-risk feature T4 tumor;  
proc lifetest data=test notable method=km;  
time month_survival*Vitalstatus(1);
Where $T_{\text{stage}}=2$;

strata Chemotherapy;

run;

*unadjusted hazard ratio;

PROC PHREG DATA=test;
  class Chemotherapy (PARAM=REF REF='No/Unknown');
Where $T_{\text{stage}}=2$;
MODEL month_survival*Vitalstatus(1)= Chemotherapy / rl;
RUN;

*adjusted hazard ratio;

PROC PHREG DATA=test;
  class ageg (PARAM=REF REF='20-49') race(PARAM=REF REF='White')
gender(PARAM=REF REF='Female') year_group(PARAM=REF REF='2010-2011')
metro(PARAM=REF REF='Rural') Chemotherapy (PARAM=REF REF='No/Unknown');
Where $T_{\text{stage}}=2$;
MODEL month_survival*Vitalstatus(1)= ageg race gender year_group metro
Chemotherapy / rl;
RUN;

*For patients at least with high-risk feature $< 12$ lymph nodes removed or assessed, testing if chemo will influence survival;
*Kaplan-Meier test;
Title Patients at least with high-risk feature $< 12$ lymph nodes removed or assessed;

proc lifetest data=test notable method=km ;
time month_survival*Vitalstatus(1);
Where Lymph_Nodes =1 ;
strata Chemotherapy;
run;

*unadjusted hazard ratio;

PROC PHREG DATA=test;
  class Chemotherapy (PARAM=REF REF='No/Unknown');
Where Lymph_Nodes =1 ;
MODEL month_survival*Vitalstatus(1)= Chemotherapy / rl;
RUN;

*adjusted hazard ratio;

PROC PHREG DATA=test;
  class ageg (PARAM=REF REF='20-49') race(PARAM=REF REF='White')
gender(PARAM=REF REF='Female') year_group(PARAM=REF REF='2010-2011')
metro(PARAM=REF REF='Rural') Chemotherapy (PARAM=REF REF='No/Unknown');
Where Lymph_Nodes =1 ;
MODEL month_survival*Vitalstatus(1)= ageg race gender year_group metro
Chemotherapy / rl;
RUN;

*For patients at least with high-risk feature poorly/differentiated histology, testing if chemo will influence survival;
*Kaplan-Meier test;
Title Patients at least with high-risk feature poorly/differentiated histology;

proc lifetest data=test notable method=km ;
time month_survival*Vitalstatus(1);
Where Histology_Grade=2 ;
strata Chemotherapy;
run;

*unadjusted hazard ratio;

PROC PHREG DATA=test;
PROC PHREG DATA=test;
  class ageg (PARAM=REF REF='20-49') race(PARAM=REF REF='White')
gender(PARAM=REF REF='Female') year_group(PARAM=REF REF='2010-2011')
metro(PARAM=REF REF='Rural') Chemotherapy (PARAM=REF REF='No/Unknown');
  Where Histology_Grade=2;
MODEL month_survival*Vitalstatus(1)= ageg race gender year_group metro
Chemotherapy /rl;
RUN;
*For patients at least with high-risk feature positive margin, testing if chemo will influence survival;
*Kaplan-Meier test;
Title Patients at least with high-risk feature positive margin;
proc lifetest data=test notable method=km;
time month_survival*Vitalstatus(1);
  Where Margin_involvement=1;
strata Chemotherapy;
run;
*unadjusted hazed ratio;
PROC PHREG DATA=test;
  class ageg (PARAM=REF REF='20-49') race(PARAM=REF REF='White')
gender(PARAM=REF REF='Female') year_group(PARAM=REF REF='2010-2011')
metro(PARAM=REF REF='Rural') Chemotherapy (PARAM=REF REF='No/Unknown');
  Where Histology_Grade=2;
MODEL month_survival*Vitalstatus(1)= ageg race gender year_group metro
Chemotherapy /rl;
RUN;
*For patients at least with high-risk feature perineural invasion, testing if chemo will influence survival;
*Kaplan-Meier test;
Title Patients at least with high-risk feature perineural invasion;
proc lifetest data=test notable method=km;
time month_survival*Vitalstatus(1);
  Where Perineural_Invasion=2;
strata Chemotherapy;
run;
*unadjusted hazed ratio;
PROC PHREG DATA=test;
  class Chemotherapy (PARAM=REF REF='No/Unknown');
  Where Perineural_Invasion=2;
MODEL month_survival*Vitalstatus(1)= Chemotherapy /rl;
RUN;
*adjusted hazed ratio;
PROC PHREG DATA=test;
  class Chemotherapy (PARAM=REF REF='No/Unknown');
  Where Perineural_Invasion=2;
MODEL month_survival*Vitalstatus(1)= Chemotherapy /rl;
RUN;
*adjusted hazed ratio;
PROC PHREG DATA=test;
  class Chemotherapy (PARAM=REF REF='No/Unknown');
  Where Perineural_Invasion=2;
MODEL month_survival*Vitalstatus(1)= Chemotherapy /rl;
RUN;
class ageg (PARAM=REF REF='20-49') race (PARAM=REF REF='White')
gender (PARAM=REF REF='Female') year_group (PARAM=REF REF='2010-2011')
metro (PARAM=REF REF='Rural') Chemotherapy (PARAM=REF REF='No/Unknown');
Where Perineural_Invasion=2;
MODEL month_survival*Vitalstatus(1)= ageg race gender year_group metro
Chemotherapy / rl;
RUN;
doors rtf close;
4. OUTPUTS

4.1 Frequency Table Outputs

Patients in each risk group

<table>
<thead>
<tr>
<th>Group</th>
<th>No chemo or Unknown</th>
<th>Chemo</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low-risk group</td>
<td>8509</td>
<td>1008</td>
<td>9517</td>
</tr>
<tr>
<td>High-risk group</td>
<td>7668</td>
<td>2692</td>
<td>10360</td>
</tr>
<tr>
<td>Unknown</td>
<td>3058</td>
<td>419</td>
<td>3477</td>
</tr>
<tr>
<td>Total</td>
<td>19235</td>
<td>4119</td>
<td>23354</td>
</tr>
</tbody>
</table>

Age Distribution

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Chemo or Unknown (n=8509)</td>
<td>Chemo (n=1008)</td>
</tr>
<tr>
<td>Age, years, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-49</td>
<td>7.05</td>
<td>26.79</td>
</tr>
<tr>
<td>50-64</td>
<td>27.01</td>
<td>44.25</td>
</tr>
<tr>
<td>65-74</td>
<td>27.12</td>
<td>20.14</td>
</tr>
<tr>
<td>75+</td>
<td>38.82</td>
<td>8.83</td>
</tr>
</tbody>
</table>
Gender Distribution

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Chemo or Unkonw (n=8509)</td>
<td>Chemo (n=1008)</td>
</tr>
<tr>
<td>Gender, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>49.75</td>
<td>49.7</td>
</tr>
<tr>
<td>Female</td>
<td>50.25</td>
<td>50.3</td>
</tr>
</tbody>
</table>

Race Distribution

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No Chemo or Unkonw (n=8509)</td>
<td>Chemo (n=1008)</td>
</tr>
<tr>
<td>Race, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>79.84</td>
<td>73.81</td>
</tr>
<tr>
<td>Black</td>
<td>11.04</td>
<td>15.77</td>
</tr>
<tr>
<td>Other</td>
<td>8.56</td>
<td>10.12</td>
</tr>
<tr>
<td>Unknown</td>
<td>0.56</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Diagnosis year distribution

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year of Diagnosis, %</td>
<td>No Chemo or Unkonw (n=8509)</td>
<td>Chemo (n=1008)</td>
<td>No Chemo or Unkonw (n=7668)</td>
</tr>
<tr>
<td>2010-2011</td>
<td>33.66</td>
<td>35.42</td>
<td>38.81</td>
</tr>
<tr>
<td>2012-2013</td>
<td>36.57</td>
<td>35.12</td>
<td>36.54</td>
</tr>
<tr>
<td>2014-2015</td>
<td>29.77</td>
<td>29.46</td>
<td>24.65</td>
</tr>
</tbody>
</table>

Metra/rural status distribution

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metro/Rural, %</td>
<td>No Chemo or Unkonw (n=8509)</td>
<td>Chemo (n=1008)</td>
<td>No Chemo or Unkonw (n=7668)</td>
</tr>
<tr>
<td>Metro</td>
<td>87.3</td>
<td>88.1</td>
<td>86.92</td>
</tr>
<tr>
<td>Rural</td>
<td>12.7</td>
<td>11.9</td>
<td>13.08</td>
</tr>
</tbody>
</table>
**T stage distribution**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T Stage, %</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>No Chemo or Unkonw (n=8509) Chemo (n=1008)</td>
<td>P No Chemo or Unkonw (n=7668) Chemo (n=2692) P</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>72.08 41.05</td>
</tr>
<tr>
<td>T4</td>
<td>0</td>
<td>27.92 58.95</td>
</tr>
</tbody>
</table>

**Nodes examined distribution**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes Examined,%</td>
<td>No Chemo or Unkonw (n=8509) Chemo (n=1008)</td>
<td>P No Chemo or Unkonw (n=7668) Chemo (n=2692) P</td>
</tr>
<tr>
<td>Nodes examined &lt;12</td>
<td>0</td>
<td>27.03 21.51</td>
</tr>
<tr>
<td>Nodes examined &gt;12</td>
<td>100</td>
<td>72.85 78.27</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0.12 0.22</td>
</tr>
</tbody>
</table>
Histology grades distribution

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Histology Grade, %</td>
<td>No Chemo or Unknown (n=8509)</td>
<td>Chemo (n=1008)</td>
<td>No Chemo or Unknown (n=7668)</td>
</tr>
<tr>
<td>Well or Moderately differentiated</td>
<td>100</td>
<td>100</td>
<td>60.42</td>
</tr>
<tr>
<td>Poorly or Undifferentiated</td>
<td>0</td>
<td>0</td>
<td>38.12</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
<td>1.46</td>
</tr>
</tbody>
</table>

Margin involvement distribution

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Stage II without any high risk factor (n=9517)</th>
<th>Stage II with at least one high risk factor (n=10360)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Margin involvement, %</td>
<td>No Chemo or Unknown (n=8509)</td>
<td>Chemo (n=1008)</td>
<td>No Chemo or Unknown (n=7668)</td>
</tr>
<tr>
<td>Positive margin</td>
<td>0</td>
<td>0</td>
<td>20.51</td>
</tr>
<tr>
<td>Negative margin</td>
<td>100</td>
<td>100</td>
<td>58.06</td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
<td>21.43</td>
</tr>
</tbody>
</table>
4.2 Survival analysis Outputs

Patients without any high-risk factor:

(a) KM plot (Log-Rank, P<0.0001)
(b) Cox regression

Unadjusted HR (Chemo Yes VS Chemo No or Unknown): 0.534, 95% CI (0.422, 0.676), P<0.0001

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
<th>95% Hazard Ratio Confidence Limits</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemotherapy Yes</td>
<td>1</td>
<td>-0.62682</td>
<td>0.12010</td>
<td>27.2393</td>
<td>&lt;.0001</td>
<td>0.534</td>
<td>0.422 0.676</td>
<td>Chemotherapy Yes</td>
</tr>
</tbody>
</table>

Adjusted HR (Chemo Yes VS Chemo No or Unknown): 0.926, 95% CI (0.726, 1.180), P=0.5323 (adjusting age, race, gender, year, metro/rural status)
Patients with at least one high risk factor:

(a) KM plot (Log-Rank, P<0.0001)

(b) Cox regression

Unadjusted HR (Chemo Yes VS Chemo No or Unknown): 0.645, 95% CI (0.576, 0.721), P<0.0001
Adjusted HR (Chemo Yes VS Chemo No or Unknown): 0.965, 95% CI (0.856, 1.088), P=0.5592

Patients at least with high risk factor T4 stage

(a) KM plot (Log-Rank, P<0.0001)

![KM plot](image)

(b) Cox regression

Unadjusted HR (Chemo Yes VS Chemo No or Unknown): 0.553, 95% CI (0.478, 0.640), P<0.0001

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
<th>95% Hazard Ratio Confidence Limits</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>ageg 50-64</td>
<td>1</td>
<td>0.24395</td>
<td>0.11947</td>
<td>4.1691</td>
<td>0.0412</td>
<td>1.276</td>
<td>1.010 1.613</td>
<td>ageg 50-64</td>
</tr>
<tr>
<td>ageg 65-74</td>
<td>1</td>
<td>0.55168</td>
<td>0.11821</td>
<td>21.7794</td>
<td>&lt;.0001</td>
<td>1.736</td>
<td>1.377 2.189</td>
<td>ageg 65-74</td>
</tr>
<tr>
<td>ageg 75+</td>
<td>1</td>
<td>1.30949</td>
<td>0.11378</td>
<td>132.4493</td>
<td>&lt;.0001</td>
<td>3.704</td>
<td>2.964 4.630</td>
<td>ageg 75+</td>
</tr>
<tr>
<td>race Black</td>
<td>1</td>
<td>0.22407</td>
<td>0.07057</td>
<td>10.0818</td>
<td>0.0015</td>
<td>1.251</td>
<td>1.090 1.437</td>
<td>race Black</td>
</tr>
<tr>
<td>race Other</td>
<td>1</td>
<td>-0.19608</td>
<td>0.09595</td>
<td>4.1762</td>
<td>0.0410</td>
<td>0.822</td>
<td>0.681 0.992</td>
<td>race Other</td>
</tr>
<tr>
<td>race unknown</td>
<td>1</td>
<td>-0.48799</td>
<td>0.70784</td>
<td>0.4753</td>
<td>0.4906</td>
<td>0.614</td>
<td>0.153 2.458</td>
<td>race unknown</td>
</tr>
<tr>
<td>gender Male</td>
<td>1</td>
<td>0.09963</td>
<td>0.04531</td>
<td>4.8344</td>
<td>0.0279</td>
<td>1.105</td>
<td>1.011 1.207</td>
<td>gender Male</td>
</tr>
<tr>
<td>year_group 2012-2013</td>
<td>1</td>
<td>0.00855</td>
<td>0.05168</td>
<td>0.0274</td>
<td>0.8685</td>
<td>1.009</td>
<td>0.911 1.116</td>
<td>year_group 2012-2013</td>
</tr>
<tr>
<td>year_group 2014-2015</td>
<td>1</td>
<td>-0.02895</td>
<td>0.09494</td>
<td>0.0930</td>
<td>0.7604</td>
<td>0.971</td>
<td>0.807 1.170</td>
<td>year_group 2014-2015</td>
</tr>
<tr>
<td>metro Metro</td>
<td>1</td>
<td>0.04006</td>
<td>0.06611</td>
<td>0.3673</td>
<td>0.5445</td>
<td>1.041</td>
<td>0.914 1.185</td>
<td>metro Metro</td>
</tr>
<tr>
<td>Chemotherapy Yes</td>
<td>1</td>
<td>-0.03569</td>
<td>0.06112</td>
<td>0.3411</td>
<td>0.5592</td>
<td>0.965</td>
<td>0.856 1.088</td>
<td>Chemotherapy Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
<th>95% Hazard Ratio Confidence Limits</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemotherapy Yes</td>
<td>1</td>
<td>-0.59177</td>
<td>0.07449</td>
<td>63.1158</td>
<td>&lt;.0001</td>
<td>0.553</td>
<td>0.478 0.640</td>
<td>Chemotherapy Yes</td>
</tr>
</tbody>
</table>
Adjusted HR (Chemo Yes VS Chemo No or Unknown): 0.770, 95% CI (0.656, 0.904), P=0.0014

Patients at least with high risk factor <12 nodes examined:

(a) KM plot (Log-Rank, P=0.0001)

(b) Cox regression

Unadjusted HR (Chemo Yes VS Chemo No or Unknown): 0.667, 95% CI (0.540, 0.823), P=0.0002
Adjusted HR (Chemo Yes VS Chemo No or Unknown): 0.965, 95% CI (0.772, 1.207), P=0.7543

Patients at least with high risk factor poorly or undifferentiated grade:

(a) KM plot (Log-Rank, P=0.0001)

(b) Cox regression

Unadjusted HR (Chemo Yes VS Chemo No or Unknown): 0.619, 95% CI (0.505, 0.758), P<0.0001
Adjusted HR (Chemo Yes VS Chemo No or Unknown): 0.960, 95% CI (0.773, 1.192), P=0.7127

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
<th>95% Hazard Confidence Limits</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemotherapy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1</td>
<td>-0.48004</td>
<td>0.10346</td>
<td>21.5284</td>
<td>&lt;.0001</td>
<td>0.619</td>
<td>0.505 0.758</td>
<td>Chemotherapy Yes</td>
</tr>
</tbody>
</table>

Patients at least with high risk factor positive margin:

(a) KM plot (Log-Rank, P=0.1501)

(b) Cox regression

Unadjusted HR (Chemo Yes VS Chemo No or Unknown): 0.852, 95% CI (0.684, 1.061), P=0.1520
Adjusted HR (Chemo Yes VS Chemo No or Unknown): 1.196, 95% CI (0.948, 1.509), P=0.1319

Patients at least with high risk factor perineural invasion present:

(a) KM plot (Log-Rank, P=0.0101)

(b) Cox regression

Unadjusted HR (Chemo Yes VS Chemo No or Unknown): 0.0109, 95% CI (0.501, 0.914), P=0.1520
Adjusted HR (Chemo Yes VS Chemo No or Unknown): 0.927, 95% CI (0.672, 1.280), P=0.6473

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
<th>95% Hazard Ratio Confidence Limits</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>ageg 50-64</td>
<td>1</td>
<td>0.27596</td>
<td>0.30227</td>
<td>0.8335</td>
<td>0.3613</td>
<td>1.318</td>
<td>0.729</td>
<td>2.383</td>
</tr>
<tr>
<td>ageg 65-74</td>
<td>1</td>
<td>0.41672</td>
<td>0.30692</td>
<td>1.8436</td>
<td>0.1745</td>
<td>1.517</td>
<td>0.831</td>
<td>2.768</td>
</tr>
<tr>
<td>ageg 75+</td>
<td>1</td>
<td>1.06357</td>
<td>0.29077</td>
<td>13.3790</td>
<td>0.0003</td>
<td>2.897</td>
<td>1.638</td>
<td>5.122</td>
</tr>
<tr>
<td>race Black</td>
<td>1</td>
<td>0.10925</td>
<td>0.17877</td>
<td>0.3734</td>
<td>0.5411</td>
<td>1.115</td>
<td>0.786</td>
<td>1.584</td>
</tr>
<tr>
<td>race Other</td>
<td>1</td>
<td>-0.12820</td>
<td>0.25350</td>
<td>0.2557</td>
<td>0.6131</td>
<td>0.880</td>
<td>0.535</td>
<td>1.446</td>
</tr>
<tr>
<td>race unknown</td>
<td>1</td>
<td>-10.02699</td>
<td>258.67270</td>
<td>0.0015</td>
<td>0.9691</td>
<td>0.000</td>
<td>0.000</td>
<td>6.73E215</td>
</tr>
<tr>
<td>gender Male</td>
<td>1</td>
<td>0.07629</td>
<td>0.12708</td>
<td>0.3604</td>
<td>0.5483</td>
<td>1.079</td>
<td>0.841</td>
<td>1.385</td>
</tr>
<tr>
<td>year_group 2012-2013</td>
<td>1</td>
<td>0.19132</td>
<td>0.14238</td>
<td>1.8056</td>
<td>0.1790</td>
<td>1.211</td>
<td>0.916</td>
<td>1.601</td>
</tr>
<tr>
<td>year_group 2014-2015</td>
<td>1</td>
<td>-0.47937</td>
<td>0.29005</td>
<td>2.7314</td>
<td>0.0984</td>
<td>0.619</td>
<td>0.351</td>
<td>1.093</td>
</tr>
<tr>
<td>metro Metro</td>
<td>1</td>
<td>0.24582</td>
<td>0.21192</td>
<td>1.3455</td>
<td>0.2461</td>
<td>1.279</td>
<td>0.844</td>
<td>1.937</td>
</tr>
<tr>
<td>Chemotherapy Yes</td>
<td>1</td>
<td>-0.07528</td>
<td>0.16454</td>
<td>0.2094</td>
<td>0.6473</td>
<td>0.927</td>
<td>0.672</td>
<td>1.280</td>
</tr>
</tbody>
</table>

---

### Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
<th>95% Hazard Ratio Confidence Limits</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chemotherapy Yes</td>
<td>1</td>
<td>-0.39056</td>
<td>0.15342</td>
<td>6.4804</td>
<td>0.0109</td>
<td>0.677</td>
<td>0.501</td>
<td>0.914 Chemotherapy Yes</td>
</tr>
</tbody>
</table>

---

Adjusted HR (Chemo Yes VS Chemo No or Unknown): 0.927, 95% CI (0.672, 1.280), P=0.6473
5. RESULTS ANALYSIS

5.1 Frequency Table Results

23,354 patients were included in this study based on the patient selection algorithm. These patients were divided into low-risk feature group (n=9,517) and high-risk feature group (n=10,360). Additionally, 3,477 patients could not be decided into any group because of the missing values for some cancer-related variables. In this study, we were only interested in patients in low-risk feature group and high-risk feature group.

The above bar charts compare the sociodemographic and cancer-related characteristics of patients by chemotherapy records within each risk group. Compared to low-risk group, the percentage of patients who received chemotherapy in high-risk group was obviously higher. No matter in low-risk feature group or high-risk feature group, compared to those who did not receive chemotherapy or unknown, patients who received chemotherapy tended to be younger. There was not obvious chemotherapy record difference for different gender, race, year of diagnosis, or rural/urban status. For patients in high-risk group, compared to those who did not receive chemotherapy or unknown, patients who received chemotherapy were more likely to have T4 tumor. There was not obvious chemotherapy record difference for other four cancer-related characteristics.

5.2 Survival Analysis Results

For patients in low-risk group, Kaplan-Meier Curve (Log-Rank, P<0.0001) and unadjusted Hazard Ratio [0.534, 95% CI (0.422, 0.676), P<0.0001] showed that the OS of patients who received chemotherapy was significantly higher than those who did not or unknown. However, when controlling for age of diagnosis, gender, race, year of diagnosis, and rural/urban status, the adjusted HR [0.926, 95% CI (0.726, 1.180), P=0.5323] is not significant, indicating that there was not significant survival difference for patients who received chemotherapy and those who did not or unknow.

Likewise, for patients in high risk group, patients who received chemotherapy had better survival than those not or unknown as the Kaplan-Meier Curve (Log-Rank, P<0.0001) and unadjusted HR [0.645, 95% CI (0.576, 0.721), P<0.0001] showed, but the adjusted HR [0.965, 95% CI (0.856, 1.088), P=0.5592] showed that this survival difference was insignificant.

Further, for each subgroup of high-risk feature group, we found that when patients at least presented T4 tumor, Kaplan-Meier test (Log-Rank, P<0.0001), unadjusted HR [0.553, 95% CI (0.478, 0.640), P<0.0001] and adjusted HR [0.770, 95% CI (0.656, 0.904), P=0.0014] consistently reflected that patients could benefit from Chemotherapy. Differently, for patients with positive margin, Kaplan-Meier test (Log-Rank, P=0.1501), unadjusted HR [0.852, 95% CI (0.684, 1.061), P=0.1520] and adjusted HR [1.196, 95% CI (0.948, 1.509), P=0.1319] consistently reflected that there was not significant OS difference for patients who received chemotherapy and those who did not or unknow. The analysis for other 3 high-risk features were similar: Although Kaplan-Meier test and unadjusted HR indicated that chemotherapy could improve OS, adjusted HR indicated this improvement was insignificant.
6. REPORTS (Analysis outcomes summary)

After analysis, the following items were concluded.

1. Younger patients were more likely to receive adjuvant chemotherapy.
2. Patients with any high-risk features of recurrence were more likely to receive adjuvant chemotherapy than those without any high-risk features.
3. When patients did not present any high-risk features of recurrence, adjuvant chemotherapy did not substantially improve patients’ survival.
4. When patients presented any high-risk features of recurrence, adjuvant chemotherapy did not substantially improve patients’ survival.
5. When patients presented T4 tumor, adjuvant chemotherapy substantially improved patients’ survival.
6. When patients presented less than 12 lymph nodes examined, adjuvant chemotherapy did not substantially improve patients’ survival.
7. When patients presented poorly/differentiated histology, adjuvant chemotherapy did not substantially improve patients’ survival.
8. When patients presented less than positive margin, adjuvant chemotherapy did not substantially improve patients’ survival.
9. When patients presented less than perineural invasion, adjuvant chemotherapy did not substantially improve patients’ survival.
10. It is inappropriate to recommend adjuvant chemotherapy for patients with any high-risk features of recurrence.
11. The presence of T4 tumor could be one indicator of using adjuvant chemotherapy.

7. LESSONS

This capstone project was a wonderful research experience for me. By reviewing the literatures, I learned the standard treatment of colon cancer and the high-risk features of recurrence for colon cancer. By using SEER dataset, I learned the ICD code of colon cancer site. By conducting data analysis, I reviewed the knowledge on SAS programming and survival analysis. Because of this project, I develop great interest in data analysis in cancer epidemiology.

In this research experience, I realized the importance of literature review. I defined many variables in this project. The rationality of these new defined variables was based on other published literatures. In addition, I also realized that the result of data analysis is not unchanged. Even though for same research question, different statistical methods can lead different results. Thinking reasonable explanation for these different results is the key point.

From this project, I felt that I need to learn more knowledge about biostatistics, especially some different statistical methods. In this research, I met the trouble of missing values. There were 3477 patients who could not be divided into any risk group because of the missing values of some cancer related variables. I could not handle these patients but just simply ignored them. I desired to learn more statistical methods about how to handle some missing values. I also felt that I need to lean more SAS advanced programming knowledge, such as Macro programming.
Adjuvant Chemotherapy for Patients with Stage II Colon Cancer

ABSTRACT

Purpose

It is controversial for patients with stage II colon cancer to use adjuvant chemotherapy after surgery. Although in theory, adjuvant chemotherapy can reduce the risk of cancer recurrence after surgery, some studies showed that adjuvant chemotherapy had limited influence on patients’ survival improvement. The purpose of this study is to explore if adjuvant chemotherapy can improve patients’ survival, especially for patients who presented any high-risk feature of recurrence.

Patients and Methods

A total of 23,354 patients with stage II colon cancer from SEER dataset were included in this study. Patients were divided into low-risk feature group and high-risk feature group. High-risk group patients were further divided into five subgroups according to the presence of different high-risk feature. We used \( \chi^2 \) tests to describe the characteristics of patients. Then we used survival analysis (Kaplan-Meier test and Cox regression analysis) to compare the survival difference by chemotherapy record (Yes vs. No/Unknown) in each risk group.

Results

No matter in low-risk or high-risk feature group, there was not enough evidence to prove that chemotherapy could improve patients’ survival. However, when patients presented T4 tumor, the Kaplan-Meier test (Log-Rank, \( P<0.0001 \)), unadjusted HR \([0.553, 95\%\ CI (0.478, 0.640), P<0.0001]\) and adjusted HR \([0.770, 95\%\ CI (0.656, 0.904), P=0.0014]\) consistently reflected that chemotherapy improved patients’ survival. For other four high-risk features, we could not prove that chemotherapy really improved patients’ survival.

Conclusion

The presence of T4 tumor is an important indicator of adjuvant chemotherapy for stage II colon cancer.

INTRODUCTION

Colorectal cancer is the third most common cancer in the United States and it is also the third leading cause of cancer-related deaths in the US. Colorectal cancer includes colon cancer and rectal cancer. Compared to rectal cancer, colon cancer is more common. It is estimated that there are 97,220 new cases of colon cancer in the US for 2018 (American Cancer Society, 2018a).

Treatment of colon cancer is largely based on its stage. Colon cancer includes five stages: Stage 0 is the earliest stage, and then range from Stage I to Stage IV. The treatments for stage 0 - III are relatively clear (stage IV means cancer has spread to distant sites, so the treatment is complex). For
patients in stage 0 or stage I. Surgery only is the standard treatment. For patients in stage III, Surgery with adjuvant chemotherapy (AC) is the standard treatment. For patients in stage II, there is no doubt that surgery is necessary, but the using of AC is controversial (American Cancer Society, 2018b; American Society of Clinical Oncology, 2004, 2017; Varghese, 2015).

In practice, the 5-year overall survival (OS) rate for stage II patients who received surgery can research 80%. The main cause of cancer recurrence is that some patients may have micrometastatic diseases at the time of surgery, but surgery cannot recognize these diseases. In theory, AC can eradicate this micrometastatic disease, prevent cancer recurrence, and then improve survival, but at the same time, AC brings potentially serious side effects, such as tiredness, nausea, vomiting, nerve damage, and so on (American Society of Clinical Oncology, 2004; Varghese, 2015). Compared to its serious side effects, the improvement of AC on survival is very small for patients with stage II colon cancer. Some studies even indicated that there is not significant survival difference between stage II patients who received AC and those who did not (American Society of Clinical Oncology, 2004; Böckelman, Engelmann, Kaprio, Hansen, & Glimelius, 2015; Fang, Efron, Berho, & Wexner, 2014; O'Connor et al., 2011; Varghese, 2015).

Based on national and international guidelines, AC are not recommended to routine use in patients with stage II colon cancer, however, if patients present high risk features of recurrence, including T4 tumor, poorly/differentiated histology, lymphovascular invasion, perineural invasion, less than 12 lymph nodes were removed or could be assessed, obstruction, or perforation, AC can be considered, but it does not mean that AC should be used for all stage II patients who present any high-risk features. Patients need to talk with their doctors (American Cancer Society, 2018b; American Society of Clinical Oncology, 2004; Varghese, 2015). Two studies indicated that even for patients with high-risk features, AC did not contribute to higher survival (Kucukzeybek et al., 2015; O'Connor et al., 2011). Two studies showed that only for patients with T4 tumor, the AC was associated with higher survival, but for patients with other high-risk features, there was no significant association between AC and survival (Kumar et al., 2014; Verhoeff, van Erning, Lemmens, de Wilt, & Pruijt, 2016). Until now, there are not uniform standards about which part of patients with stage II colon cancer should be given AC. More statistical data are needed to evaluate the benefit of AC for patients with stage II colon cancer, especially for those presenting high-risk features.

In this study, the association between AC and survival are further explored among patients who are diagnosed as stage II colon cancer, especially for those presenting any high-risk feature.

**PATIENTS AND METHODS**

**Data source**

The data came from the Surveillance, Epidemiology, and End Results (SEER) cancer database “Incidence -SEER 18 Regs Custom Data (with additional treatment fields), No2017 Sub (1973-215 varying) Linked to County Attributes - Total U.S., 1969-2016 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, released April 2018, based on the November 2017 submission.”

**Study Population**
All patients aged 20 years and older as well as diagnosed with primary American Joint Committee on Cancer (AJCC) stage II colon adenocarcinoma in a SEER area from 2010 to 2015 were eligible for this study. Colon Cancer were identified by site and histology codes (Primary Site=C18.0-18.9; Histology=8140-8147,8210-8221,8260-8263,8480-8481,8490) (O'Connor et al., 2011; Weiss et al., 2014). Further, Patients who received surgical resection of primary colon cancer were selected. Moreover, considering that some patients had bad health condition and died before they could receive AC, we excluded patients who survived at less than 6 months after diagnosis. Eventually, a total of 23,354 patients were included in this study.

**Variables**

Outcome variable: The primary outcome variable is overall survival rate (OS), which is the percentage of patients who are alive for a period of time after their diagnosis. SEER Database recorded patients’ survival months and death/alive status.

Explanatory variable: The primary explanatory variable is AC. AC is chemotherapy offered after primary site surgery with an intent of reducing the risk of cancer recurrence (American Society of Clinical Oncology, 2004; Kopetz, Freitas, Calabrich, & Hoff, 2008). SEER database had chemotherapy records but did not indicate this chemotherapy was AC or not. We used Kentucky Cancer Registry (KCR) database to check the percentage of patients who received chemotherapy before primary surgery and found this percentage is minimal. Considering the similarity of KCR data and SEER data, we believed that most chemotherapy records in SEER data could be regarded as AC records.

Stratification variables: The primary stratification variables were high risk features of recurrence. SEER database included the information of five high risk features (T4 tumor, poorly/differentiated histology, less than 12 lymph nodes removed or assessed, positive margin, perineural invasion,). The information of other three high risk features (lymphovascular invasion, obstruction, or perforation) was not recorded in SEER database.

Control variables: The primary control variables included age of diagnosis (>=20 years old), gender (male and female), race (white, black, other or unknown), year of diagnosis (2010-2015), and urban/rural status.

**Statistical Analysis**

Patients were divided into two risk-feature groups. If patients presented any one of the five SEER-listed high-risk features of recurrence, they were in the high-risk features group. If they did not present any of the five SEER-listed high-risk features, they were in the low-risk features group.

Then we used \(x^2\) tests to compare the frequency of patients’ sociodemographic and cancer-related variables by chemotherapy records (Yes vs. No/Unknown) within each risk group. The sociodemographic variables were five control variables: age, gender, race, year of diagnosis, and urban/rural status. The cancer-related variables included T stage, nodes examined, histology grades, margin involvement, and perineural invasion.

Kaplan-Meier test were used to compare the OS difference by chemotherapy records (Yes vs. No/Unknown) within each risk-feature group. Then Cox regression analysis was used to compare the hazard ratio (HR) for different chemotherapy records (Yes vs. No/Unknown) within each risk-feature
group. Compared to patients in low-risk group, we were more interested in the effect of chemotherapy on patients in high-risk group because patients in high-risk group were more likely to be recommended to receive chemotherapy. Based on the presence of different high-risk features, the high-risk group patients were further divided into five subgroups: patients at least with T4 tumor, patients at least with poorly/differentiated histology, patients at least with less than 12 lymph nodes removed or assessed, patients at least with positive margin, and patients at least with perineural invasion. For each of these five subgroups, we repeated to conduct Kaplan-Meier test and Cox regression analysis.

RESULTS

Characteristics of Patients

As described in the patients and methods section, 23,354 patients were included in this study based on the patient selection algorithm. These patients were divided into low-risk feature group (n=9,517) and high-risk feature group (n=10,360). Additionally, 3,477 patients could not be decided into any group because of the missing values for some cancer-related variables. In this study, we were only interested in patients in low-risk feature group and high-risk feature group.

Table 1 compare the sociodemographic and cancer-related characteristics of patients by chemotherapy records within each risk group. Compared to low-risk group, the percentage of patients who received chemotherapy in high-risk group was obviously higher. No matter in low-risk feature group or high-risk feature group, compared to those who did not received chemotherapy or unknown, patients who received chemotherapy tended to be younger. There was not obvious chemotherapy record difference for different gender, race, year of diagnosis, or rural/urban status. For patients in high-risk group, compared to those who did not receive chemotherapy or unknown, patients who received chemotherapy were more likely to have T4 tumor. There was not obvious chemotherapy record difference for other four cancer-related characteristics.

Survival Benefit of Chemotherapy

For patients in low-risk group (Figure 1), Kaplan-Meier Curve (Log-Rank, P<0.0001) and unadjusted Hazard Ratio [0.534, 95% CI (0.422, 0.676), P<0.0001] showed that the OS of patients who received chemotherapy was significantly higher than those who did not or unknown. However, when controlling for age of diagnosis, gender, race, year of diagnosis, and rural/urban status, the adjusted HR [0.926, 95% CI (0.726, 1.180), P=0.5323] is not significant, indicating that there was not significant survival difference for patients who received chemotherapy and those who did not or unknow.

Likewise, for patients in high risk group (Figure 2), patients who received chemotherapy had better survival than those not or unknown as the Kaplan-Meier Curve (Log-Rank, P<0.0001) and unadjusted HR [0.645, 95% CI (0.576, 0.721), P<0.0001] showed, but the adjusted HR [0.965, 95% CI (0.856, 1.088), P=0.5592] showed that this survival difference was insignificant.

Further, for each subgroup of high-risk feature group, we found that when patients at least presented T4 tumor, Kaplan-Meier test (Log-Rank, P<0.0001), unadjusted HR [0.553, 95% CI (0.478, 0.640), P<0.0001] and adjusted HR [0.770, 95% CI (0.656, 0.904), P=0.0014] consistently reflected that patients could benefit from Chemotherapy (Figure 3). Differently, for patients with positive margin, Kaplan-Meier test (Log-Rank, P=0.1501), unadjusted HR [0.852, 95% CI (0.684, 1.061), P=0.1520] and
adjusted HR [1.196, 95% CI (0.948, 1.509), P=0.1319] consistently reflected that there was not significant OS difference for patients who received chemotherapy and those who did not or unknown (Figure 6). The analysis for other 3 high-risk features were similar (Figure 4, Figure 5, Figure 7): Although Kaplan-Meier test and unadjusted HR indicated that chemotherapy could improve OS, adjusted HR indicated this improvement was insignificant.

DISCUSSION

The univariate analysis reflected that no matter patients presented any high-risk feature or not, the survival rate for patients who received chemotherapy was obviously better than those who did not or unknown. This difference disappeared after patients’ sociodemographic characteristics were controlled. The possible reason was that younger patients were more likely to receive chemotherapy and younger patients had better survival than older patients. Chemotherapy did not really improve patients’ survival.

Patients who presented any high-risk features of recurrence were more likely to be recommended to receive chemotherapy. However, in this study, even when patients presented any high-risk features, chemotherapy did not really improve patients’ survival. We guessed that maybe for some special high-risk features, chemotherapy could improve patients’ survival, but for some other features, chemotherapy could not improve patients’ survival. After analysis, we found when patients presented T4 tumor feature, no matter they presented any other high-risk features or not, chemotherapy could improve their survival. For other four high-risk features, we could not decide if chemotherapy really improved patients’ survival.

In this analysis, the data was the information of patients diagnosed as colon adenocarcinoma from 2010 to 2015. This data is the newest data for colon cancer in SEER database. From 2010, the information of margin involvement and perineural invasion became available. Few previous studies involved these two factors. The other advantage of this study was that the age range. Some previous studies only involved patients aged 65 or older, but this study included younger patients (O’Connor et al., 2011; Weiss et al., 2014).

Because SEER data did not include the information of perineural invasion, obstruction, and perforation. The definition of risk group was not precise, and we could not test the effects of these three features. SEER data recorded chemotherapy as “Yes” and “No/Unknow”, which also influenced the accuracy of analysis. In addition, SEER data did not record the time of chemotherapy, thus we could not precisely distinguish the adjuvant chemotherapy and non-adjuvant chemotherapy. Two solutions for these limitations were to use SEER-Medicare data or KCR data. However, SEER-Medicare data lacked the information of patients aged <65. KCR data for stage II colon cancer was too small to impactable. Further analysis needs more advanced database.

CONCLUSIONS

For patients with stage II colon cancer, chemotherapy improved the overall survival for those presented T4 tumor feature, no matter they presented any other high-risk features or not. The presence of T4 tumor was an important indicator of chemotherapy for stage II colon cancer.
<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Stage II without any high-risk features (n=9,517)</th>
<th>Chemo (n=1,008)</th>
<th>P</th>
<th>Stage II with any high-risk feature (n=10,360)</th>
<th>Chemo (n=2,692)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, years, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-49</td>
<td>7.05</td>
<td>26.79</td>
<td>6.17</td>
<td>19.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-64</td>
<td>27.01</td>
<td>44.25</td>
<td>22.4</td>
<td>43.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>27.12</td>
<td>20.14</td>
<td>25.12</td>
<td>26.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75+</td>
<td>38.82</td>
<td>8.83</td>
<td>46.31</td>
<td>11.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>49.75</td>
<td>49.7</td>
<td>46.35</td>
<td>51.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>50.25</td>
<td>50.3</td>
<td>53.65</td>
<td>48.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Race, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>79.84</td>
<td>73.81</td>
<td>81.7</td>
<td>79.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>11.04</td>
<td>15.77</td>
<td>10.54</td>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>other</td>
<td>8.56</td>
<td>10.12</td>
<td>7.46</td>
<td>7.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>0.56</td>
<td>0.3</td>
<td>0.3</td>
<td>0.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year of Diagnosis, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2011</td>
<td>33.66</td>
<td>35.42</td>
<td>38.81</td>
<td>37.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012-2013</td>
<td>36.57</td>
<td>35.12</td>
<td>36.54</td>
<td>35.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2014-2015</td>
<td>29.77</td>
<td>29.46</td>
<td>24.65</td>
<td>27.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metro/Rural, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metro</td>
<td>87.3</td>
<td>88.1</td>
<td>86.92</td>
<td>86.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural</td>
<td>12.7</td>
<td>11.9</td>
<td>13.08</td>
<td>13.52</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T Stage, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td>100</td>
<td>100</td>
<td>72.08</td>
<td>41.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td>0</td>
<td>0</td>
<td>27.92</td>
<td>58.95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nodes Examined, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;12</td>
<td>0</td>
<td>0</td>
<td>27.03</td>
<td>21.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;=12</td>
<td>100</td>
<td>100</td>
<td>72.85</td>
<td>78.27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
<td>0.12</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Histology Grade, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well or Moderately differentiated</td>
<td>100</td>
<td>100</td>
<td>60.42</td>
<td>65.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poorly or Undifferentiated differentiated</td>
<td>0</td>
<td>0</td>
<td>38.12</td>
<td>31.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>unknown</td>
<td>0</td>
<td>0</td>
<td>1.46</td>
<td>2.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Margin involvement, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>0</td>
<td>0</td>
<td>20.51</td>
<td>21.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>100</td>
<td>100</td>
<td>58.06</td>
<td>53.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
<td>21.43</td>
<td>25.48</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perineural Invasion, %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No invasion present</td>
<td>100</td>
<td>100</td>
<td>76.02</td>
<td>73.22</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Invasion present</td>
<td>0</td>
<td>0</td>
<td>13.02</td>
<td>13.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>0</td>
<td>0</td>
<td>10.97</td>
<td>12.82</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. The effects of adjuvant chemotherapy on overall survival for patients without any high-risk feature.

Figure 2. The effects of adjuvant chemotherapy on overall survival for patients with any high-risk feature.
Figure 3. The effects of adjuvant chemotherapy on overall survival for patients at least with T4 tumor

Log-Rank, P<0.0001
Unadjusted HR: 0.553, 95% CI (0.478, 0.640), P<0.0001
Adjusted HR: 0.770, 95% CI (0.656, 0.904), P=0.0014

Figure 4. The effects of adjuvant chemotherapy on overall survival for patients at least with < 12 lymph nodes removed or assessed

Log-Rank, P=0.0001
Unadjusted HR: 0.667, 95% CI (0.540, 0.823), P=0.0002
Adjusted HR: 0.965, 95% CI (0.772, 1.207), P=0.7543
Figure 5. The effects of adjuvant chemotherapy on overall survival for patients at least with poorly/differentiated histology

Figure 6. The effects of adjuvant chemotherapy on overall survival for patients at least with positive margin
Figure 7. The effects of adjuvant chemotherapy on overall survival for patients at least with perineural invasion
REFERENCES


9. Defense Presentation

The Impact of Adjuvant Chemotherapy on Survival for Patients with Stage II Colon Cancer

Li Ding
MPH, biostatistics

Background

It is controversial to use adjuvant chemotherapy for patients with stage II colon cancer after surgery.

In theory, adjuvant chemotherapy can reduce the risk of cancer recurrence and improve survival, but it has serious side effects.

In practice, some studies indicated the impact of adjuvant chemotherapy on survival is minimal, even for patients with high-risk features of cancer recurrence. For some features, chemo can improve survival.
Study Purpose

1. For patients without any high risk features of cancer recurrence, can they benefit from chemotherapy?

2. For patients with any high risk features of cancer recurrence, can they benefit from chemotherapy?

3. For patients with one particular feature of cancer recurrence, can they benefit from chemotherapy?

SEER Database

1. The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) is an authoritative source of information on cancer incidence and survival in the United States.

2. The SEER Program registries routinely collect data on patient demographics, primary tumor site, tumor morphology and stage at diagnosis, first course of treatment, and follow-up for vital status.
Study Population

<table>
<thead>
<tr>
<th>Variable</th>
<th>Selection algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary Site</td>
<td>Site recode ICD-O-3/WHO 2008= Colon excluding Rectum;</td>
</tr>
<tr>
<td></td>
<td>Primary Site= C18.0-18.9</td>
</tr>
<tr>
<td></td>
<td>Histology = 8140-8147,8210-8211,8220-8221,8260-8263,8480-8481,8490</td>
</tr>
<tr>
<td></td>
<td>First malignant primary indicator=Yes</td>
</tr>
<tr>
<td>Surgery</td>
<td>Surg Prim Site (1998+)=30-80</td>
</tr>
<tr>
<td>Age</td>
<td>&gt;=20</td>
</tr>
<tr>
<td>Diagnosis year</td>
<td>2010-2015</td>
</tr>
<tr>
<td>Race</td>
<td>All race</td>
</tr>
<tr>
<td>gender</td>
<td>Male and Female</td>
</tr>
<tr>
<td>Survival Month</td>
<td>Complete data are available and Survival Month &gt;=6</td>
</tr>
</tbody>
</table>

Statistical Methods

1. Descriptive Statistics: Frequency tables and chi-square test
2. Survival Analysis:
   - Univariate Analysis (Kaplan-Meier Analysis and univariate Cox Regression)
   - Multivariate Analysis (Multivariate Cox Regression)
Patients in each risk group

Patients Distribution

- Low-risk Group (9,517)
- High-risk Group (10,360)
- T4 Tumor (3,728)
- <12 Lymph Nodes examined (2,652)
- Poor Histology (3,771)
- Positive Margin (2,148)
- Perineural Invasion (1,374)

- No chemo or Unknown
- Chemo

Patients Characteristics in each risk group

Low-risk group

- Age
- Race

High-risk group

- Age
- Gender

High risk subgroup (T4 Tumor)

- Age
- Gender

High risk subgroup (T4 Tumor): Race
For patients **without** any high risk features of cancer recurrence, chemotherapy **did not** substantially impact survival.

For patients **with** any high risk features of cancer recurrence, chemotherapy **did not** substantially impact survival.
Survival Analysis

High risk subgroup (T4 Tumor)

Analysis of Maximum Likelihood Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Hazard Ratio</th>
<th>95% Hazard Ratio CI</th>
<th>P &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-49 (reference)</td>
<td>1.00</td>
<td>(0.64, 1.57)</td>
<td>0.96</td>
</tr>
<tr>
<td>50-64</td>
<td>1.51</td>
<td>(1.03, 2.24)</td>
<td>0.03</td>
</tr>
<tr>
<td>65-74</td>
<td>1.71</td>
<td>(1.14, 2.57)</td>
<td>0.0026</td>
</tr>
<tr>
<td>75+</td>
<td>2.19</td>
<td>(1.37, 3.47)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>race</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White (reference)</td>
<td>1.00</td>
<td>(0.89, 1.15)</td>
<td>0.92</td>
</tr>
<tr>
<td>Black</td>
<td>1.10</td>
<td>(0.77, 1.57)</td>
<td>0.77</td>
</tr>
<tr>
<td>Other</td>
<td>0.83</td>
<td>(0.52, 1.33)</td>
<td>0.43</td>
</tr>
<tr>
<td>gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (reference)</td>
<td>1.00</td>
<td>(0.89, 1.15)</td>
<td>0.92</td>
</tr>
<tr>
<td>Male</td>
<td>1.00</td>
<td>(0.89, 1.15)</td>
<td>0.92</td>
</tr>
<tr>
<td>Diagnosis Year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2012-2013</td>
<td>1.00</td>
<td>(0.89, 1.15)</td>
<td>0.92</td>
</tr>
<tr>
<td>2014-2015</td>
<td>1.00</td>
<td>(0.89, 1.15)</td>
<td>0.92</td>
</tr>
<tr>
<td>Chemotherapy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>1.00</td>
<td>(0.89, 1.15)</td>
<td>0.92</td>
</tr>
<tr>
<td>No</td>
<td>1.00</td>
<td>(0.89, 1.15)</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Log-Rank, P<0.0001
Unadjusted HR: 0.553, 95% CI (0.478, 0.640), P<0.0001

For patients with T4 tumor, chemotherapy substantially impact survival.
For other four high-risk features, chemotherapy did not substantially impact survival.

Discussion

Overall, adjuvant chemotherapy did not substantially impact survival because younger patients are more likely to receive chemo than older. However, whey patients presented T4 tumor, they were more likely to benefit from adjuvant chemotherapy.

Study Strengths:
SEER database is a reliable cancer database.
Sample size is large enough.

Study Limitations:
Chemotherapy record: “No chemo or Unknown” Vs. “chemo”
Use KCR data to check the percentage of unknown chemotherapy
THANK YOU!!!