Paper PH10

Using the 7th edition American Joint Commission on Cancer (AJCC) Cancer Staging Manual to Determine Esophageal Cancer Staging in SEER-Medicare Data

Johnita L. Byrd, Emory University School of Medicine, Atlanta, GA;
Dr. Felix G. Fernandez, Emory University School of Medicine, Atlanta, GA

ABSTRACT

The purpose of this paper is to discuss the methodology in creating a consistent esophageal cancer staging system through the use of the 7th edition AJCC Cancer staging for patients with data across a time period of 10 years with different staging criteria.

In a study using Surveillance, Epidemiology, and End Results (SEER)-Medicare data, we examine health care utilization in surgically treated beneficiaries with esophageal cancer. Cancer staging is required in determining primary outcomes for this study. Across the time span of our study population, two different AJCC staging systems were needed. In each system, there are slight differences to the way the staging is developed. As it is important to use a uniform cancer staging system for consistency, we choose to use the 7th edition (most recent) to create a cancer staging for the entire population so that we could make assumptions about the population in terms of the present.

In order to use the 7th edition AJCC Cancer staging system for all patients, we create algorithms in SAS ®. These algorithms include a number of variables across several years. First, we determine different locations of a tumor based on a primary site variable. The T, N, and M staging variables are categorized by AJCC differently at different time periods. We researched each variable to standardize the different T, N, and M stages. We, then, classify the list of histology types into three groups: Adenocarcinoma, Squamous-Cell Carcinoma, or Other/Unknown stage groupings. The grade variable remains consistent across all years. Once all variables are uniform, algorithms are created to calculate the most recent AJCC Cancer staging, based on literature explaining the development of staging for esophageal cancer in the 7th Edition of the AJCC Cancer Staging Manual.

Methods used to derive the 7th edition AJCC cancer stages can be expanded and utilized in developing updated staging systems for other cancers. Such methods allow the use of the most recent staging system regardless of the dates of the SEER-Medicare data, especially since SEER has not updated its data to the most recent version.

BACKGROUND

The retrospective cohort study population included 1,734 Medicare beneficiaries undergoing an esophagectomy for esophageal cancer from 2001 through 2009 using the SEER-Medicare database to evaluate outcomes. Data files from the SEER-Medicare dataset that were utilized for the cancer staging was the Patient Entitlement and Diagnosis Summary File (PEDSF) which contains SEER cancer registry data as well as information on Medicare enrollment and mortality. Patient, disease, and first course treatment are available through the SEER Program and captured in the PEDSF data. This file includes variables such as birth date (month and year), race, gender, date of diagnosis (month and year), primary site, tumor grade, histology, tumor size, lymph node status, first course of therapy, tumor stage and vital status (date and cause of death). In other words, this file contains all the variables necessary to compute the cancer staging for the cohort based on the AJCC 7th edition cancer staging.

The analysis sample included patients 66 years of age or older who were not diagnosed at death or autopsy. All patients had to have both Part A and Part B Medicare coverage with concurrent health maintenance organization (HMO) enrollment for at least one year before their esophageal cancer diagnosis. Only full fee-for-service beneficiaries not enrolled in other insurance programs will have complete claims records. In addition, patients who were 65 at the time of diagnosis do not have claims available for the year before diagnosis of esophageal cancer.

INTRODUCTION

Different variables will be needed from the SEER-Medicare data set to obtain the final cancer staging. There will be instances where the variables change slightly over time. Our study time period happens to occur within the time period of both variables. In the interest of consistency, it is important to use the same variable throughout the entire

course of the study. This paper looks at tumor staging in SEER-Medicare data over a 10 year period to find the most efficient way to create a uniform variable for tumor staging. We will discuss the following methods for obtaining uniformity in cancer staging. They are

1. PROC FREQ
2. The DATA Step

All variables needed to perform the most recent cancer staging can be found in the PEDSF file. For all variable names shown below, the number '1' after the variable name indicates that there was a possibility of more than one field with data for this variable. However, for the purposes of our study, we decided to use the first field only. We examined the additional columns but there were an extreme number of missing values. The most comprehensive field was always found to be the first column. So, we used the first column only (represented by the '1' after the variable name) for all variables that had more than column of data available. The list of variables is shown below:

- Primary Site (SITE1 – name in the PEDSF file): this variable is used to determine the location of the tumor. For this study, the locations were Upper, Middle, Lower, or Other. This variable remained the same across the study period.
- Histology ICD-O-3 (HIST1 – name in the PEDSF file): this variable is used to determine the tumor type or histology. For this study, the histology type was Adenocarcinoma, Squamous-Cell Carcinoma, or Other/Unknown. This variable remained the same across the study period.
- Grade (GRADE1 – name in the PEDSF file): this variable is used to determine the grade of the tumor. The variable remained the same across the study period.
- T-stage (E10EX1 & CSEX1 – names in the PEDSF file): these variables are used to determine the T classification. The variable E10EX1 was used from 1988 – 2003 and CSEX1 was used from 2004 – present.
- N-stage (E10PN1 – name in the PEDSF file): this variable records the number of positive regional lymph nodes. This variable remained the same across the study period.
- M-stage (E10EX1 & CSMET1 – names in the PEDSF file): these variables are used to determine the M classification. The variable E10EX1 was used from 1988 – 2003 (this variable also included metastasis in addition to the tumor extent) and CSMET1 was used from 2004 – present.

The 7th edition of the AJCC cancer staging manual for the esophagus contained the appropriate stage groupings. These stage groupings are presented in Tables 1 and 2. Since we are interested in the most recent cancer staging for esophageal cancer, we will base all of our data steps below on these two tables.

Table 1. Adenocarcinoma (Other/Unknown) Stage Groupings

Stage	T	N	M	Grade
0	is(HGD)	0	0	1
IA	1	0	0	1,2,6,9
IB	1	0	0	3,4
	2	0	0	1,2,6,9
IIA	2	0	0	3,4
IIB	3	0	0	Any
	1,2	1	0	Any
IIIA	1,2	2	0	Any
	3	1	0	Any
	4	0	0	Any
IIIB	3	2	0	Any
IIIC	4	1,2	0	Any
	4	3	0	Any
	1,2,3	3	0	Any
IV	Any	Any	1	Any

Table 2. Squamous-cell Carcinoma Stage Groupings

Stage	T	N	M	Grade	Location
0	is(HGD)	0	0	Any	Any
IA	1	0	0	1,6,9	Any
IB	1	0	0	2,3,4	Any
	2,3	0	0	1,6,9	Lower
IIA	2,3	0	0	1,6,9	Upper, Middle
	2,3	0	0	2,3,4	Lower
IIB	2,3	0	0	2,3,4	Upper, Middle
	1,2	1	0	Any	Any
IIIA	1,2	2	0	Any	Any
	3	1	0	Any	Any
	4	0	0	Any	Any
IIIB	3	2	0	Any	Any
IIIC	4	1,2	0	Any	Any
	4	3	0	Any	Any
	1,2,3	3	0	Any	Any
IV	Any	Any	1	Any	Any

All examples were run using SAS ® V9.3.

There were 1,734 observations and 16 variables.

CLASSIFYING VARIABLES USING PROC FREQ AND THE DATA STEP

PROC FREQ performs a frequency on the data to summarize all categories within the data set. It was important to know what values were represented in our dataset to make sure that we knew the best way to categorize the data. To make sure that all values were considered, we used PROC FREQ to compile a complete list. Once we had a complete list of all of the variables within the data set, we were able to use the DATA step to place the values in the correct categories for the different variables. Each variable required a different method of categorization. The methods for categorization of the variables depended upon the documentation in the PEDSF file and the interpretation of the documentation, based on the literature and the experience of Surgeons who perform esophagectomies. Each of the different methods is described below:

PRIMARY SITE

In order to determine the location of the tumor, the primary site variable was used. Frequencies were run on the data to determine the different values for the primary site variable. Here is an example:

```
PROC FREQ DATA = final_cohort;
    TABLE site1;
RUN;
```

Once the different values were determined, we were able to group them into different locations for the tumor. Table 3 describes the different values of interest within the primary site variable and the correct grouping of the values.

Table 3. The different tumor locations of interest

SITE1	Description of Variable	Location
C15.0	Cervical esophagus	Upper
C15.1	Thoracic esophagus	Middle
C15.2	Abdominal esophagus	Lower
C15.3	Upper third of esophagus	Upper
C15.4	Middle third of esophagus	Middle
C15.5	Lower third of esophagus	Lower
C15.8	Overlapping lesion of esophagus	Other
C15.9	Esophagus, NOS	Other

Abbreviations: NOS, not otherwise specified

A new variable was created, named 'Location', to describe the location of the tumor. The location of the tumor could be Upper, Middle, Lower or Other.

```
DATA final_cohort;
SET final_cohort;
       if site1 = '150' or site1 = '153' then location = 'Upper';
       if site1 = '151' or site1 = '154' then location = 'Middle';
       if site1 = '152' or site1 = '155' then location = 'Lower';
       if site1 = '009' or site1 = '023' or site1 = '029' or site1 = '030' or site1 =
       '031' or site1 = '019' or site1 = '051' or site1 = '129' or site1 = '441' or
       site1 = '443' or site1 = '447' or site1 = '601' or site1 = '158' or site1 =
       '159' or site1 = '160' or site1 = '163' or site1 = '180' or site1 = '182' or
       site1 = '184' or site1 = '186' or site1 = '187' or site1 = '188' or site1 =
       '209' or site1 = '211' Or site1 = '239' or site1 = '320' or site1 = '341' or
       site1 = '343' or site1 = '421' or site1 = '444' or site1 = '446' or site1 =
       '501' or site1 = '503' or site1 = '505' or site1 = '508' or site1 = '509'
       or site1 = '619' or site1 = '649' or site1 = '659' or site1 = '674' or site1 =
       '678' or site1 = '679' or site1 = '770' or site1 = '774' or site1 = '778' or
       site1 = '671' or site1 = '672' or site1 = '676' or site1 = '739' then location
       = 'Other';
RUN;
```

It is important to note that there were several other values obtained when frequencies were performed on the primary site variable. It was decided that all other values besides those in the above list would be coded as 'Other'.

HISTOLOGY TYPE

Next, we used the histology type variable to determine the tumor type. In order to categorize the tumor types, frequencies were calculated to determine all possible values.

```
       PROC FREQ DATA = final_cohort;
           TABLE hist1;
       RUN;
```

Once all values for the cohort were known, we used the appendix in the PEDSF documentation to determine which staging group the values would be placed. An Excel spreadsheet was created to include a comprehensive list of all the values found for our cohort as well as the corresponding translation for the values, according to the appendix. A Thoracic Surgeon was consulted to determine how best to categorize the codes. Each value or code was placed in one of three categories: Adenocarcinoma, Squamous-Cell Carcinoma, or Other/Unknown. Table 4 shows all the different histology codes found for our cohort, description of the code, and its corresponding tumor type grouping.

Table 3. The different histology type codes of interest

HIST1	Description of Variable	Group	HIST1	Description of Variable	Group
8000	Neoplasm Malignant	Unknown/Other	8263	Adenocarcinoma in Tubulovill	Adenocarcinoma
8010	Carcinoma NOS	Unknown/Other	8310	Clear Cell Adenocarcinoma	Adenocarcinoma
8012	Large Cell Carcinoma NOS	Unknown/Other	8323	Mixed Cell Adenocarcinoma	Adenocarcinoma
8020	Undifferentiated Carcinoma	Unknown/Other	8331	Foll Adenocarcinoma	Adenocarcinoma
8021	Anaplastic Carcinoma	Unknown/Other	8480	Mucinous Adenocarcinoma	Adenocarcinoma
8041	Small Cell Carcinoma NOS	Unknown/Other	8481	Mucin Prod Adenocarcinoma	Adenocarcinoma
8046	Non-Small Cell Carcinoma	Unknown/Other	8490	Signet Ring Cell Adenocarcinoma	Adenocarcinoma
8051	Verrucous Carcinoma	Squamous Cell	8500	Duct Adenocarcinoma	Adenocarcinoma
8070	Squamous Cell Carcinoma	Squamous Cell	8504	Intracyst Pap Adenocarcinoma	Adenocarcinoma
8071	Squamous Cell Carcinoma Keratiniz	Squamous Cell	8520	Lobular Carcinoma NOS	Adenocarcinoma
8072	Squamous Cell Carcinoma Non-Keratiniz	Squamous Cell	8523	Infil Duct mixed w/other Carcinoma	Unknown/Other
8073	Squamous Cell Carcinoma SmallCell Non-Keratiniz	Squamous Cell	8550	Acinic Cell Adenocarcinoma	Adenocarcinoma
8074	Squamous Cell Carcinoma Spindle Cell	Squamous Cell	8560	Adenosquamous Carcinoma	Unknown/Other
8081	Bowen's Disease	Unknown/Other	8720	Malignant Melanoma NOS	Unknown/Other
8083	Basaloid Squamous Cell Carcinoma	Squamous Cell	8723	Regressing Melanoma	Unknown/Other
8120	Papillary Transitional Cell Carcinoma	Unknown/Other	8742	Lentigo Maligna Melanoma	Unknown/Other
8130	AdenoCarcinoma NOS	Unknown/Other	8936	Gastrointestinal Stromal Sarcoma	Unknown/Other
8140	Scirrhous Adenocarcinoma	Adenocarcinoma	8980	Carcinosarcoma NOS	Unknown/Other
8144	Intestinal Adenocarcinoma	Adenocarcinoma	9591	Malignant Lymphoma nonHodgkin	Unknown/Other
8145	Diffuse Adenocarcinoma	Adenocarcinoma	9663	Hodgkin Lympg nod Sclerosis	Unknown/Other
8210	AdenoCarcinoma in Adeno Polyp	Adenocarcinoma	9680	Malignant Lymph Large B Cell	Unknown/Other
8240	Carcinoid Tumor	Unknown/Other	9699	Marginal zone B cell Lymph	Unknown/Other
8246	Neuroendocrine Carcinoma	Unknown/Other	9732	Multiple Myeloma	Unknown/Other
8247	Merkel Cell Carcinoma	Unknown/Other	9823	Bcell chronic lymph Leukemia	Unknown/Other
8255	Adenocarcinoma w/ mixed subtypes	Adenocarcinoma	9950	Polycythemia Vera	Unknown/Other
8260	Papillary Adenocarcinoma NOS	Adenocarcinoma	9960	Chronic Myeloproliferative	Unknown/Other

A new variable was created, named 'Group', to describe the tumor type. The data steps taken to obtain this new variable are listed below:

```
DATA final_cohort;
SET final_cohort;
      if hist1 = '8000' or hist1 = '8010' or hist1 = '8046' then group = 'Unknown';
      if hist1 = '8012' or hist1 = '8020' or hist1 = '8021' or hist1 = '8041' or
      hist1 = '8081' or hist1 = '8120' or hist1 = '8130' or hist1 = '8240' or hist1 =
      '8246' or hist1 = '8247' or hist1 = '8523' or hist1 = '8560' or hist1 = '8720'
      or hist1 = '8723' or hist1 = '8742' or hist1 = '8936' or hist1 = '8980' or
      hist1 = '9591' or hist1 = '9663' or hist1 = '9680' or hist1 = '9699' or hist1 =
      '9732' or hist1 = '9823' or hist1 = '9950' or hist1 = '9960' then group =
      'Other';
      if hist1 = '8051' or hist1 = '8070' or hist1 = '8071' or hist1 = '8072' or
      hist1 = '8073' or hist1 = '8074' or hist1 = '8083' then group = 'Squamous';
      if hist1 = '8140' or hist1 = '8141' or hist1 = '8142' or hist1 = '8143' or
      hist1 = '8144' or hist1 = '8145' or hist1 = '8210' or hist1 = '8255' or hist1 =
      '8260' or hist1 = '8263' or hist1 = '8310' or hist1 = '8323' or hist1 = '8331'
      or hist1 = '8480' or hist1 = '8481' or hist1 = '8490' or hist1 = '8500' or
      hist1 = '8504' or hist1 = '8520' or hist1 = '8550'
      then group = 'Adenocarcinoma';
RUN;
```

GRADE

The grade of the tumor didn't have to be recoded. Since the grade variable was calculated with differentiation codes, we were able to use the values provided for purposes of calculating the cancer staging. If you notice in Tables 1 and 2, we only look at the grading and differentiation codes so no reclassification was necessary. A frequency was run on the data to make sure that all values were represented.

```
PROC FREQ DATA = final_cohort;
    TABLE grade1;
RUN;
```

All values were used within this cohort. The values were 1 – 4 and 9.

T-STAGE CLASSIFICATION

In order to perform cancer staging, it is necessary to obtain each person's T-stage classification. In order to obtain the T-stage for our cohort, it was important to look at two different variables within the SERR-Medicare data since the variables changed over time. From 1988 – 2003, the E10EX1 variable was used and from 2004 – present, the CSEX1 variable was used. In order to create the correct staging based on the AJCC 7th edition cancer staging, the same classification was needed for all years in the study. As a result, we decided to run some frequencies on both variables to determine the best way to classify them according to the AJCC 7th edition cancer staging guidelines and the PEDSF file's documentation for the variables.

```
PROC FREQ DATA = final_cohort;
    TABLES e10ex1 csex1;
RUN;
```

Once all of the possible values were determined, a uniform classification was developed. Table 5 shows how the variable 'E10EX1' was classified.

Table 5. T-stage classification for E10EX1 variable (2001 - 2003)

E10EX1	Description of the variable	T Classification
00	Noninvasive; intrapithelial	Tis
01, 03, 06, 10, 11, 12, 13, 14, 15, 16, 30, 34	Mucosa, NOS; Lamina propria; Muscularis mucosae; Submucosa	T1
20, 23	Muscularis propria invaded	T2
40, 42, 45, 55	Adventitia and/or soft tissue invaded	T3
60, 65, 70, 80	Tumor invades adjacent structures -- upper, middle or lower esophagus	T4
85, 99	Unknown	Unknown

Abbreviations: NOS, not otherwise specified

Table 6 shows how the variable 'CSEX1' was classified. The classification will be the same for both variables which will allow for a uniform classification that covers all years of the study cohort.

Table 6. T-stage classification for CSEX1 variable (2004+)

CSEX1	Description of the variable	T Classification
00, 000	Noninvasive; intrapithelial	Tis
05, 10, 100, 11, 110, 12, 120, 13, 15, 150, 16, 160, 30, 300, 010, 060, 030	Mucosa, NOS; Lamina propria; Muscularis mucosae; Submucosa	T1
20, 200, 23, 230, 24, 210, 220	Muscularis propria invaded	T2
40, 400, 45, 450, 50, 420, 421	Adventitia and/or soft tissue invaded	T3
60, 600, 65, 650, 70, 78, 80, 800, 615	Tumor invades adjacent structures -- upper, middle or lower esophagus	T4
95, 99, 999	Unknown	Unknown

Next, a new variable, named 'T', was created to describe the T-stage classification that is needed for the final cancer staging.

```
DATA final_cohort;
SET final_cohort;
      if e10ex1 = '00' then T = 'Tis';
      if e10ex1 = '01' or e10ex1 ='10' or e10ex1 = '11' or e10ex1 = '12' or e10ex1 =
      '13' or e10ex1 = '14' or e10ex1 = '15' or e10ex1 = '16' or e10ex1 = '30' or
      e10ex1 = '34' or e10ex1 = '03' or e10ex1 = '06' then T = 'T1';
      if e10ex1 = '20' or e10ex1 = '23' then T = 'T2';
      if e10ex1 = '40' or e10ex1 = '42' or e10ex1 = '45' or e10ex1 = '55' then T =
      'T3';
      if e10ex1 = '60' or e10ex1 = '65' or e10ex1 = '70' or e10ex1 = '80' then T =
      'T4';
      if e10ex1 = '85' or e10ex1 = '99' then T = 'Unknown';

      if csex1 = '00' or csex1 = '000' then T = 'Tis';
      if csex1 = '05' or csex1 = '10' or csex1 = '100' or csex1 = '11' or csex1 =
      '110' or csex1 = '12' or csex1 = '120' or csex1 = '13' or csex1 = '15' or csex1
      = '150' or csex1 = '16' or csex1 = '160' or csex1 = '30' or csex1 = '300' or
      csex1 = '010' or csex1 = '060' or csex1 = '030' then T = 'T1';
      if csex1 = '20' or csex1 = '200' or csex1 = '23' or csex1 = '230' or csex1 =
      '24' or csex1 = '210' or csex1 = '220' then T = 'T2';
      if csex1 = '40' or csex1 = '400' or csex1 = '45' or csex1 = '450' or csex1 =
      '50' or csex1 = '420' or csex1 = '421' then T = 'T3';
      if csex1 = '60' or csex1 = '600' or csex1 = '65' or csex1 = '650' or csex1 =
      '70' or csex1 = '78' or csex1 = '80' or csex1 = '800' or csex1 = '615' then T =
      'T4';
      if csex1 = '95' or csex1 = '99' or csex1 = '999' then T = 'Unknown';
RUN;
```

N-STAGE CLASSIFICATION

It was important to calculate the N-stage classification in order to calculate the cancer staging for each patient. In classifying N, it is necessary to use the number of positive nodes variable, 'E10PN1'. The values for this variable

determine the exact N classification. Frequencies were run on the variable to make sure all the values were captured.

```
PROC FREQ DATA = final_cohort;
    TABLE e10pn1;
RUN;
```

The new classification of the N-stage, according to the 7th edition AJCC cancer staging, is shown below in Table 7:

Table 7. N-stage classification

N Classification	Description
N0	No regional lymph node metastases
N1	1 to 2 positive regional lymph nodes
N2	3 to 6 positive regional lymph nodes
N3	≥ 7 positive regional lymph nodes
NX	Unknown whether nodes are positive

A new variable was created, named 'N', to describe the N-stage classification that would be used to determine the final cancer staging for each patient.

```
DATA final_cohort;
SET final_cohort;
      if e10pn1 = '00' then N = 'N0';
      if e10pn1 = '01' or e10pn1 = '02' or e10pn1 = '97' or e10pn1 = '95' then N =
      'N1';
      if e10pn1 = '03' or e10pn1 = '04' or e10pn1 = '05' or e10pn1 = '06' then N =
      'N2';
      if e10pn1 = '07' or e10pn1 = '08' or e10pn1 = '09' or e10pn1 = '10' or e10pn1 =
      '11' or e10pn1 = '12' or e10pn1 = '13' or e10pn1 = '14' or e10pn1 = '15' or
      e10pn1 = '16' or e10pn1 = '17' or e10pn1 = '19' or e10pn1 = '20' or e10pn1 =
      '21' or e10pn1 = '23' or e10pn1 = '24' or e10pn1 = '25' or e10pn1 = '34' or
      e10pn1 = '38' then N = 'N3';
      if e10pn1 = '98' or e10pn1 = '99' then N = 'NX';
RUN;
```

M-STAGE CLASSIFICATION

The M-stage classification was performed in the same way as the T-stage classification since the variables to determine M-stage were changed over time. From 1988 – 2003, metastases were determined through the E10EX1 variable (this is the same variable used to determine the T-stage). If the E10EX1 variable = 85, then it was considered to have metastasis. Otherwise, there was no metastasis. From 2004 – present, a new variable, named 'CSMET1', was created to determine the M-stage. A frequency was performed on both variables to determine all possible values.

```
PROC FREQ DATA = final_cohort;
    TABLE e10ex1 csmet1;
RUN;
```

A description of this variable is shown in Table 8.

Table 8. M-stage classification for CSMET1 variable (2004+)

CSEX1	Description of the variable	M Classification
00, 10, 11, 30	No; Distant lymph node	M0
40, 50, 98	Distant Metastases	M1
99	Unknown	Unknown

Finally, a new variable, named 'M', was created to describe the M-stage classification for the entire dataset.

```
DATA final_cohort;
SET final_cohort;
      if e10ex1 = '00' or e10ex1 = '01' or e10ex1 = '03' or e10ex1 = '06' or e10ex1
      ='10' or e10ex1 = '11' or e10ex1 = '12' or e10ex1 = '13' or e10ex1 = '14' or
      e10ex1 = '15' or e10ex1 = '16' or e10ex1 = '20' or e10ex1 = '23' or e10ex1 =
      '30' or e10ex1 = '34' or e10ex1 = '40' or e10ex1 = '42' or e10ex1 = '45' or
      e10ex1 = '55' or e10ex1 = '60' or e10ex1 = '65' or e10ex1 = '70' or e10ex1 =
      '80' then M = 'M0';
      if e10ex1 = '85' then M = 'M1';
      if e10ex1 = '99' then M = 'Unknown';

      if csmet1 = '00' or csmet1 = '10' or csmet1 = '11' or csmet1 = '30' then M =
      'M0';
      if csmet1 = '40' or csmet1 = '50' or csmet1 = '98' then M = 'M1';
      if csmet1 = '99' then M = 'UNKNOWN';
RUN;
```

Now that all of the variables needed for the 7th edition AJCC cancer staging were all uniformly classified, we performed the appropriate cancer staging; Adenocarcinoma, Squamous-Cell Carcinoma, or Other/Unknown staging. The DATA step for each type of cancer staging is shown below:

ADENOCARCINOMA STAGING

```
/* Adenocarcinoma Staging */
DATA final_cohort;
SET final_cohort;
      if group = 'Adenoca' and T = 'Tis' and (N = 'N0' or N = 'NX') and (M = 'M0' or
      M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
      grade1 = '6' or grade1 = '9') then adeno_stage = 'Zero';
      if group = 'Adenoca' and T = 'T1' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
      = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '6' or grade1 = '9') then
      adeno_stage = 'IA';
      if group = 'Adenoca' and T = 'T1' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
      = 'UN') and (grade1 = '3' or grade1 = '4') then adeno_stage = 'IB';
      if group = 'Adenoca' and T = 'T2' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
      = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '6' or grade1 = '9') then
      adeno_stage = 'IB';
      if group = 'Adenoca' and T = 'T2' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
      = 'UN') and (grade1 = '3' or grade1 = '4') then adeno_stage = 'IIA';
      if group = 'Adenoca' and T = 'T3' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
      = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
      grade1 = '6' or grade1 = '9') then adeno_stage = 'IIB';
      if group = 'Adenoca' and (T = 'T1' or T = 'T2') and N = 'N1' and (M = 'M0' or M
      = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
      grade1 = '6' or grade1 = '9') then adeno_stage = 'IIB';
      if group = 'Adenoca' and (T = 'T1' or T = 'T2') and N = 'N2' and (M = 'M0' or M
      = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
      grade1 = '6' or grade1 = '9') then adeno_stage = 'IIIA';
      if group = 'Adenoca' and T = 'T3' and N = 'N1' and (M = 'M0' or M = 'UN') and
      (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or grade1 = '6'
      or grade1 = '9') then adeno_stage = 'IIIA';
      if group = 'Adenoca' and T = 'T4' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
      = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
      grade1 = '6' or grade1 = '9') then adeno_stage = 'IIIA';
      if group = 'Adenoca' and T = 'T3' and N = 'N2' and (M = 'M0' or M = 'UN') and
      (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or grade1 = '6'
      or grade1 = '9') then adeno_stage = 'IIIB';
```

```
        if group = 'Adenoca' and T = 'T4' and (N = 'N1' or N = 'N2') and (M = 'M0' or M
        = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
        grade1 = '6' or grade1 = '9') then adeno_stage = 'IIIC';
        if group = 'Adenoca' and T = 'T4' and N = 'N3' and (M = 'M0' or M = 'UN') and
        (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or grade1 = '6'
        or grade1 = '9') then adeno_stage = 'IIIC';
        if group = 'Adenoca' and (T = 'T1' or T = 'T2' or T = 'T3' or T = 'Unk') and N
        = 'N3' and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '3' or grade1 = '4' or grade1 = '6' or grade1 = '9') then adeno_stage = 'IIIC';
        if group = 'Adenoca' and (T = 'T1' or T ='T2' or T ='T3' or T = 'T4' or T =
        'Unk') and (N = 'N0' or N = 'NX' or N = 'N1' or N = 'N2' or N = 'N3') and M =
        'M1' and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
        grade1 = '6' or grade1 = '9') then adeno_stage = 'IV';
RUN;
```

OTHER/UNKNOWN STAGING

```
/* Other/Unknown Staging */
DATA final_cohort;
SET final_cohort;
        if (group = 'Other' or group = 'Unknown') and T = 'Tis' and (N = 'N0' or N =
        'NX') and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '3' or grade1 = '4' or grade1 = '6' or grade1 = '9') then other_stage = 'Zero';
        if (group = 'Other' or group = 'Unknown') and T = 'T1' and (N = 'N0' or N =
        'NX') and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '6' or grade1 = '9') then other_stage = 'IA';
        if (group = 'Other' or group = 'Unknown') and T = 'T1' and (N = 'N0' or N =
        'NX') and (M = 'M0' or M = 'UN') and (grade1 = '3' or grade1 = '4') then
        other_stage = 'IB';
        if (group = 'Other' or group = 'Unknown') and T = 'T2' and (N = 'N0' or N =
        'NX') and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '6' or grade1 = '9') then other_stage = 'IB';
        if (group = 'Other' or group = 'Unknown') and T = 'T2' and (N = 'N0' or N =
        'NX') and (M = 'M0' or M = 'UN') and (grade1 = '3' or grade1 = '4') then
        other_stage = 'IIA';
        if (group = 'Other' or group = 'Unknown') and T = 'T3' and (N = 'N0' or N =
        'NX') and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '3' or grade1 = '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IIB';
        if (group = 'Other' or group = 'Unknown') and (T = 'T1' or T = 'T2') and N =
        'N1' and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '3' or grade1 = '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IIB';
        if (group = 'Other' or group = 'Unknown') and (T = 'T1' or T = 'T2') and N =
        'N2' and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '3' or grade1 = '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IIIA';
        if (group = 'Other' or group = 'Unknown') and T = 'T3' and N = 'N1' and (M =
        'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 =
        '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IIIA';
        if (group = 'Other' or group = 'Unknown') and T = 'T4' and (N = 'N0' or N =
        'NX') and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '3' or grade1 = '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IIIA';
        if (group = 'Other' or group = 'Unknown') and T = 'T3' and N = 'N2' and (M =
        'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 =
        '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IIIB';
        if (group = 'Other' or group = 'Unknown') and T = 'T4' and (N = 'N1' or N =
        'N2') and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '3' or grade1 = '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IIIC';
        if (group = 'Other' or group = 'Unknown') and T = 'T4' and N = 'N3' and (M =
        'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 =
        '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IIIC';
        if (group = 'Other' or group = 'Unknown') and (T = 'T1' or T = 'T2' or T = 'T3'
        or T = 'Unk') and N = 'N3' and (M = 'M0' or M = 'UN') and (grade1 = '1' or
        grade1 = '2' or grade1 = '3' or grade1 = '4' or grade1 = '6' or grade1 = '9')
        then other_stage = 'IIIC';
```

```
        if (group = 'Other' or group = 'Unknown') and (T = 'T1' or T ='T2' or T ='T3'
        or T = 'T4' or T = 'Unk') and (N = 'N0' or N = 'NX' or N = 'N1' or N = 'N2' or
        N = 'N3') and M = 'M1' and (grade1 = '1' or grade1 = '2' or grade1 = '3' or
        grade1 = '4' or grade1 = '6' or grade1 = '9') then other_stage = 'IV';
RUN;
```

SQUAMOUS-CELL CARCINOMA STAGING

```
/* Squamous-Cell Carcinoma Staging */
DATA final_cohort;
SET final_cohort;
        if group = 'Squamou' and T = 'Tis' and (N = 'N0' or N = 'NX') and (M = 'M0' or
        M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
        grade1 = '6' or grade1 = '9') then squamous_stage = 'Zero';
        if group = 'Squamou' and T = 'T1' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
        = 'UN') and (grade1 = '1' or grade1 = '6' or grade1 = '9') then squamous_stage
        = 'IA';
        if group = 'Squamou' and T = 'T1' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
        = 'UN') and (grade1 = '2' or grade1 = '3' or grade1 = '4') then squamous_stage
        = 'IB';
        if group = 'Squamou' and (T = 'T2' or T = 'T3') and (N = 'N0' or N = 'NX') and
        (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '6' or grade1 = '9') and
        (location = 'Lower' or location = 'Other') then squamous_stage = 'IB';
        if group = 'Squamou' and (T = 'T2' or T = 'T3') and (N = 'N0' or N = 'NX') and
        (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '6' or grade1 = '9') and
        (location = 'Upper' or location = 'Middl') then squamous_stage = 'IIA';
        if group = 'Squamou' and (T = 'T2' or T = 'T3') and (N = 'N0' or N = 'NX') and
        (M = 'M0' or M = 'UN') and (grade1 = '2' or grade1 = '3' or grade1 = '4') and
        (location = 'Lower' or location = 'Other') then squamous_stage = 'IIA';
        if group = 'Squamou' and (T = 'T2' or T = 'T3') and (N = 'N0' or N = 'NX') and
        (M = 'M0' or M = 'UN') and (grade1 = '2' or grade1 = '3' or grade1 = '4') and
        (location = 'Upper' or location = 'Middl') then squamous_stage = 'IIB';
        if group = 'Squamou' and (T = 'T1' or T = 'T2') and N = 'N1' and (M = 'M0' or M
        = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
        grade1 = '6' or grade1 = '9') then squamous_stage = 'IIB';
        if group = 'Squamou' and (T = 'T1' or T = 'T2') and N = 'N2' and (M = 'M0' or M
        = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
        grade1 = '6' or grade1 = '9') then squamous_stage = 'IIIA';
        if group = 'Squamou' and T ='T3' and N = 'N1' and (M = 'M0' or M = 'UN') and
        (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or grade1 = '6'
        or grade1 = '9') then squamous_stage = 'IIIA';
        if group = 'Squamou' and T = 'T4' and (N = 'N0' or N = 'NX') and (M = 'M0' or M
        = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
        grade1 = '6' or grade1 = '9') then squamous_stage = 'IIIA';
        if group = 'Squamou' and T = 'T3' and N = 'N2' and (M = 'M0' or M = 'UN') and
        (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or grade1 = '6'
        or grade1 = '9') then squamous_stage = 'IIIB';
        if group = 'Squamou' and T = 'T4' and (N = 'N1' or N = 'N2') and (M = 'M0' or M
        = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
        grade1 = '6' or grade1 = '9') then squamous_stage = 'IIIC';
        if group = 'Squamou' and T = 'T4' and N = 'N3' and (M = 'M0' or M = 'UN') and
        (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or grade1 = '6'
        or grade1 = '9') then squamous_stage = 'IIIC';
        if group = 'Squamou' and (T = 'T1' or T = 'T2' or T = 'T3' or T = 'Unk') and N
        = 'N3' and (M = 'M0' or M = 'UN') and (grade1 = '1' or grade1 = '2' or grade1 =
        '3' or grade1 = '4' or grade1 = '6' or grade1 = '9') then squamous_stage =
        'IIIC';
        if group = 'Squamou' and (T = 'T1' or T = 'T2' or T = 'T3' or T = 'T4' or T =
        'Unk') and (N = 'NX' or N = 'N0' or N = 'N1' or N = 'N2' or N = 'N3') and M =
        'M1' and (grade1 = '1' or grade1 = '2' or grade1 = '3' or grade1 = '4' or
        grade1 = '6' or grade1 = '9') then squamous_stage = 'IV';
RUN;
```

The above DATA steps are derived from Tables 1 and 2. Once all variables of interest were classified correctly, it was a matter of following Tables 1 and 2 to determine the most recent cancer staging.

After each type of cancer staging variable was developed, it was decided that a final cancer staging variable should be created, named 'final_stage'. The purpose of this variable would be to have a final cancer stage of each patient, no matter the type of cancer staging.

```
DATA final_cohort;
SET final_cohort;
      if other_stage = . and squamous_stage = . then final_stage = adeno_stage;
      else if adeno_stage = . and squamous_stage = . then final_stage = other_stage;
      else if adeno_stage = . and other_stage = . then final_stage = squamous_stage;
      else final_stage = .;
RUN;
```

The final frequency step below gives us the final cancer staging.

```
      PROC FREQ DATA = final_cohort;
          TABLE final_stage;
      RUN;
```

The output will look like this:

final_stage

final_stage	Frequency	Percent	Cumulative Frequency	Cumulative Percent
IA	308	18.17	308	18.17
IB	281	16.58	589	34.75
IIA	123	7.26	712	42.01
IIB	400	23.60	1112	65.60
IIIA	249	14.69	1361	80.29
IIIB	115	6.78	1476	87.08
IIIC	118	6.96	1594	94.04
IV	43	2.54	1637	96.58
Zero	58	3.42	1695	100.00

Frequency Missing = 39

It is important to note that there were only 39 patients with missing data for the cancer staging. One of the reasons why we decided to create our own cancer staging methodology was due to the fact that the other possible variables in the SEER-Medicare data set had a large number of missing values.

CONCLUSION

There is no doubt that cancer staging can be a tedious process, but it can be very important when dealing with certain data sets. The SEER-Medicare data set currently uses the AJCC 6th edition for cancer staging. With the many changes that come over time, it is important to be able to present data in most recent times.

Although esophageal cancer was used for this paper, you should be able to use the same concepts presented above to stage any type of cancer. The only real differences will occur in how the AJCC 7th edition staging manual classifies the cancer of interest. Once you have tables similar to Tables 1 and 2, then you can use the same algorithms described above. Also, it is important to make sure that some of the variables used to obtain the final staging are changed to fit the type of cancer of interest. Learning the methodology and algorithms explained above can save a

significant amount of time when it comes to performing the 7th edition AJCC cancer staging for your SEER-Medicare data set.

REFERENCES

1. Rice, Thomas W., Blackstone MD, Eugene H. and Rusch MD, Valerie W. (2010). "7th Edition of the AJCC Cancer Staging Manual: Esophagus and Esophagogastric Junction." Annals of Surgical Oncology. Pgs. 1721 – 1724. NewYork: Springer-Verlag.
2. http://healthservices.cancer.gov/seermedicare/

ACKNOWLEDGMENTS

I would like to acknowledge Kevin Ward from Emory University for his help and expertise in working with SEER-Medicare data. I would also like to acknowledge Dr. Theresa W. Gillespie from Emory University for her expertise on research and methodology.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Johnita L. Byrd, MS
Enterprise: ICF International
Address: 3 Corporate Blvd., NE #370
City, State ZIP: Atlanta, GA 30329
Work Phone: 404-718-8531
E-mail: Johnita.Byrd@icfi.com