

X



Memorial Sloan Kettering
Cancer Center

[Make an Appointment](#)
[Back](#)

[About MSK](#) [Cancer Treatment](#)
[Epidemiology & Population Sciences](#)
[Learn About Cancer & Treatment](#)

ABOUT US

[Our mission, vision & core values](#)

[Leadership](#)

[History](#)

[Inclusion & belonging](#)

[Annual report](#)

[Give to MSK](#)

FOR THE MEDIA

Problem: Health services researchers and epidemiologists are interested in secular trends in cancer incidence and survival. Analysis of secular trends requires application of consistent staging algorithms across time periods. However, this is made difficult because staging algorithms change over time.

Problem for Colon Cancer Researchers: The most consistent staging algorithm in use by the SEER registries is SEER historic staging system which categorizes patients as local, regional or advanced. The SEER historic staging system has the principal advantage of being recorded consistently across all time periods. Unfortunately, however, the historic staging system is a suboptimal strategy for colon cancer because it does not map easily to the staging schema that is commonly used by clinicians. The system most commonly used today is the AJCC system (now in its 6th edition) which relies on the TNM system.

The problem with SEER historic stage is that it lumps together node negative (Stage II) and node positive (Stage III) colon cancer patients together. This is problematic because clinical trials and practice guidelines support adjuvant chemotherapy for stage III patients but not for stage II patients where the data to support chemotherapy is equivocal. Therefore, distinguishing between these two groups is essential and investigators are dissatisfied with SEER historic stage to describe colon cancer.

Motivation for Constructing AJCC Staging System Across Time Periods: Analyses of secular trends in health care delivery or the study of technology diffusion such as the use of screening colonoscopy requires accurate and reliable information about stage-specific trends in incidence and survival. The desire to consistently stage colon cancer patients from 1970 to the present using a single schema motivated this project.

Unfortunately, AJCC staging was not recorded by the registrars prior to 1988. However, investigators as part of the cancer intervention and surveillance modeling network (CISNET) colorectal group have gone back to construct staging information for those patients who had cancer directed surgery. This is possible because colon cancer staging relies on assessment of regional lymph nodes. Because the number and involvement of nodes has been recorded consistently, the individual components of information can be used to reconstruct AJCC stage for those patients who had definitive surgery. This technical report serves to outline the approach to accomplish this and to remind users of several important points that should be considered in any analyses that rely on these data sets.

i) The AJCC algorithm for CRC has itself changed over time.

ii) The AJCC algorithm is available at <http://training.seer.cancer.gov/staging/systems/ajcc/>, and the most recent comparison guide is AJCC Cancer Staging Manual, Fifth versus Sixth Edition.

Although the current system in use is the sixth edition. The sixth edition essential subdivides the 5th edition into smaller more refined prognostic groups. The editions of AJCC map neatly from one to the other. In this report, we supply steps for application of the AJCC 5th edition to cohorts from time periods prior to 1988 when AJCC was not recorded. It is important to note that as long as tumor size/depth of invasion, nodal information and metastasis information are recorded, AJCC staging schema can be followed.

iii) Because of the time delay in reporting statistics (e.g., 2004 data becomes public available in April 2007), AJCC 5th edition is used instead of the latest version (AJCC 6th edition). In addition, the AJCC 6th edition differs primarily in that it adds additional sub-categories (IIIA, IIIB, IIIC)

Methods for Constructing AJCC Stage Prior to 1988: In order to describe long term trends in stage at diagnosis, disease incidence, and cancer survival, CISNET investigators have reviewed SEER historic staging manuals and applied contemporary staging algorithms (AJCC 5th edition) across time. This is possible for colon cancer for patients who have undergone cancer directed surgery. Over 95% of stage I-III colon cancer patients undergo cancer directed surgery. In the absence of apparent metastatic disease (stage IV-advanced) colon cancer staging requires pathologic assessment of regional lymph nodes. Often clinical investigators are interested in defining cohorts of patients with stage II (node negative) and stage III (node positive) and potentially stage I tumors that required surgical excision. Investigators want pathologically staged colon cancer patients. By combining the cancer site specific surgery codes and the staging codes, we have provided investigators with a way to quickly and efficiently identify cohorts of patients who have pathologically staged colorectal cancer. To provide data users with the potential to compare cohorts across time periods, we have used consistent algorithms to apply the AJCC 5 coding schema to patients who were either: 1) not staged by tumor registrars using the AJCC algorithms; 2) staged by AJCC algorithms prior to the 5th edition. This approach enables investigators to look at secular trends in incidence/mortality/survival using a consistent system over time. Investigators are cautioned in using these data and must recognize the approximate nature of these estimates.

v) While colon cancer is quite straightforward, it is imperative for investigators to recognize the complexities related to rectosigmoid and rectal cancer patients. Rectal cancer patients are increasingly treated with preoperative chemotherapy and radiation. As a result, the pathologic stage recorded does not necessarily correspond to the disease severity at diagnosis. Rectosigmoid patients may have cancers that are either above the peritoneal reflection and therefore true colon cancers or below the peritoneal reflection and therefore true rectal cancers. This distinction is not recorded by SEER. For the purpose of these analyses, rectosigmoid cancers are considered with rectal cancers. However, investigators studying rectal cancer should note that some patients with Rectosigmoid (RS) primaries would not clinically be considered to have rectal cancer.

Colon Cancer Staging Variable for SEER Data (1975-2003): A complex user-defined variable was defined to show colorectal cancer rates by AJCC stage, 5th Edition. The variable was created for use with the [SEER Limited-Use Data](#) within the [SEER*Stat software](#) (software link available soon) in order to apply AJCC 5th Edition stages to colon cases for years when this staging schema was not previously recorded. This will allow investigators to compare cases across time periods using a staging system commonly used in clinical practice.

A SEER*Stat frequency matrix file is provided here to demonstrate how the variable can be used, and how it was created. In the matrix we show malignant colon and rectal cancer by the defined AJCC stage variable, primary site subdivisions, and year of diagnosis within the SEER 17 registries for 1973-2004.

The AJCC stage variable was created by merging several variables in SEER*Stat in order to show statistics by AJCC stages (I, II, III, IV, unstaged) so that colon cancer is staged consistently for all years in the analysis. In order to do so, stage for

- 1975-1982 was based on a combination of the 13-digit extent of disease fields,
- 1983-1987 was based on a combination of the 4-digit extent of disease fields,
- 1988-2003 was based on AJCC stage, 3rd edition.

The following standard SEER variables were used to define the complex variable:

- Year of diagnosis;
- Site recode with Kaposi and mesothelioma
- Expanded EOD(5) - CP57 (1973-1982);
- Expanded EOD(7) - CP59 (1973-1982);
- Expanded EOD(12) - CP64 (1973-1982);
- Expanded EOD(13) - CP65 (1973-1982);
- Expanded EOD(10) - CP62 (1973-1982);
- SEER historic stage A;
- EOD 4 - extent (1983-1987);
- EOD 4 - nodes (1983-1987);
- AJCC stage 3rd edition (1988-2003);
- 2-Digit NS EOD part 1 (1973-1982).

How to Use this Variable: The following example demonstrates how this variable can be used.

Download and open the following, [SEER*Stat Frequency Matrix](#).

You must have SEER*Stat and access to the 1973-2004 SEER Limited-Use data to open the file.

If you look at the matrix, you'll notice that:

- The frequencies are displayed by AJCC Stage in the column dimension and Year of Diagnosis in the row dimension.
- There are no values for stage for the years 1973-74 and 2004. This variable is defined to use for cases diagnosed between 1975-2003.

To view the session information (selections made to define the parameters of the analysis), select Print Preview from the File menu. A summary of the selections, as well as the user-defined variable definitions can be viewed and printed with the frequency table.

To save the user-defined variable for use in your own SEER*Stat sessions:

- From the Matrix menu in SEER*Stat, select Retrieve Session.
- From the File menu, select Dictionary....
- On the Dictionary window, use the "+" to expand the Merged folder.
- Double-click on the "AJCC stage (Colon, Deb Schrag 1975-03)" variable to open the Edit Merged Variable window.
- Check the Save to Dictionary box and then click OK.
- The variable will be saved for use in other SEER*Stat sessions.

Additional SEER*Stat Selection for Histologies: On the selection tab in SEER*Stat, the selection of Site recode with Kaposi and mesothelioma = 'Colon and Rectum' excludes cancers outside the colon and rectum and also excludes some histologically-based cancers (leukemias, lymphomas, ...). See [SEER Site recode variable definition](#) for more information. An additional case selection was made on histology to include only those cell types that clinicians may consider to be colorectal cancer. These selections include: Histologic Type ICD-O-3 = 8000-8001,8010,8020,8140,8210-8211,8220-8221,8260-8263,8480-8482,8490.

Definitions of the Variable Groupings

Stage I:

```
((Race, Sex, Year Dx, Registry, County.Year of diagnosis) = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'  
AND {Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = 1-3  
AND {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = 0, ' ', 'Blank(s)'  
AND {Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 0  
AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = 0  
AND ({Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = 0  
OR ({Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = ' ', 'Blank(s)'  
AND {Stage.SEER historic stage A} = 'Localized'))  
OR ((Race, Sex, Year Dx, Registry, County.Year of diagnosis) = '1983', '1984', '1985', '1986', '1987'  
AND ({Extent of Disease.EOD 4 - extent (1983-1987)} = 1  
AND {Extent of Disease.EOD 4 - nodes (1983-1987)} = 0,9  
OR ({Extent of Disease.EOD 4 - extent (1983-1987)} = 2  
AND {Extent of Disease.EOD 4 - nodes (1983-1987)} = 0))  
OR ((Race, Sex, Year Dx, Registry, County.Year of diagnosis) = '1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '1996', '1997', '1998',  
'1999', '2000', '2001', '2002', '2003'  
AND {Stage.AJCC stage 3rd edition (1988-2003)} = 'Stage I')
```

Stage II:

```
((Race, Sex, Year Dx, Registry, County.Year of diagnosis) = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'  
AND ({Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = 0  
AND {Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 0  
AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = 0  
AND ({Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = 4,6-9, '&'  
AND {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = 0, ' ', 'Blank(s)'  
OR ({Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = 1-9, '&'))  
OR ({Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = ' ', 'Blank(s)'  
AND {Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 0  
AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = 0  
AND {Stage.SEER historic stage A} = 'Localized'  
AND ({Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = 4,6-9, '&')
```

AND {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = 0, '–', 'Blank(s)'
 OR {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = 1-9, '&'
 OR {Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = 5
 AND {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = '–', 'Blank(s)')))
 OR {Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = '–', 'Blank(s)'
 AND {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = '–', 'Blank(s)'
 AND {Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = '–', 'Blank(s)'
 AND {Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 'Blank(s)'
 AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = '–', 'Blank(s)'
 AND {Extent of Disease.2-Digit NS EOD part 1 (1973-1982)} = 'Regional, direct extension only'
 AND {Stage.SEER historic stage A} = 'Regional')))
 OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984', '1985', '1986', '1987'
 AND {Extent of Disease.EOD 4 - extent (1983-1987)} = 4-7
 AND {Extent of Disease.EOD 4 - nodes (1983-1987)} = 0
 OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '1996', '1997', '1998',
 '1999', '2000', '2001', '2002', '2003'
 AND {Stage.AJCC stage 3rd edition (1988-2003)} = 'Stage II')

Stage III:

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'
 AND ((({Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = 1
 AND {Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 0
 AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = 0
 AND ({Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = 1-9, '&', '–', 'Blank(s)')
 OR ({Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = 0
 AND {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = 2-9, '&')))
 OR ({Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = '–', 'Blank(s)'
 AND {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = '–', 'Blank(s)'
 AND {Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = '–', 'Blank(s)'
 AND {Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 'Blank(s)'
 AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = '–', 'Blank(s)'
 AND {Extent of Disease.2-Digit NS EOD part 1 (1973-1982)} = 'Regional, nodes only', 'Regional, direct extension and nodes'
 AND {Stage.SEER historic stage A} = 'Regional')))
 OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984', '1985', '1986', '1987'
 AND {Extent of Disease.EOD 4 - extent (1983-1987)} = 0-7
 AND {Extent of Disease.EOD 4 - nodes (1983-1987)} = 1,8
 OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '1996', '1997', '1998',
 '1999', '2000', '2001', '2002', '2003'
 AND {Stage.AJCC stage 3rd edition (1988-2003)} = 'Stage III')

Stage IV:

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'
 AND ({Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 1
 AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = 0, '–', 'Blank(s)')
 OR {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = 1-9, '&'
 OR {Stage.SEER historic stage A} = 'Distant'))
 OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984', '1985', '1986', '1987'
 AND ({Extent of Disease.EOD 4 - extent (1983-1987)} = 8
 OR {Extent of Disease.EOD 4 - nodes (1983-1987)} = 7)
 OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '1996', '1997', '1998',
 '1999', '2000', '2001', '2002', '2003'
 AND {Stage.AJCC stage 3rd edition (1988-2003)} = 'Stage IV')

Unstaged:

({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1975', '1976', '1977', '1978', '1979', '1980', '1981', '1982'

AND {Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = '–', 'Blank(s)'
AND {Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 0
AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = 0
AND {Stage.SEER historic stage A} = 'Unstaged'
OR ({Extent of Disease.Expanded EOD(5) - CP57 (1973-1982)} = '–', 'Blank(s)'
AND {Extent of Disease.Expanded EOD(7) - CP59 (1973-1982)} = '–', 'Blank(s)'
AND {Extent of Disease.Expanded EOD(10) - CP62 (1973-1982)} = '–', 'Blank(s)'
AND {Extent of Disease.Expanded EOD(12) - CP64 (1973-1982)} = 'Blank(s)'
AND {Extent of Disease.Expanded EOD(13) - CP65 (1973-1982)} = '–', 'Blank(s)'
AND ({Extent of Disease.2-Digit NS EOD part 1 (1973-1982)} = 'Non-localized, NOS'
AND {Stage.SEER historic stage A} = 'Unstaged'
OR ({Extent of Disease.2-Digit NS EOD part 1 (1973-1982)} = 'Localized'
AND {Stage.SEER historic stage A} = 'Localized'
OR ({Extent of Disease.2-Digit NS EOD part 1 (1973-1982)} = 'Unknown', 'Blank(s)'
AND {Stage.SEER historic stage A} = 'Unstaged'
OR ({Extent of Disease.2-Digit NS EOD part 1 (1973-1982)} = 'Regional, NOS'
AND {Stage.SEER historic stage A} = 'Regional'))))
OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1983', '1984', '1985', '1986', '1987'
AND ({Extent of Disease.EOD 4 - extent (1983-1987)} = 2-7
AND {Extent of Disease.EOD 4 - nodes (1983-1987)} = 9
OR ({Extent of Disease.EOD 4 - extent (1983-1987)} = 3
AND {Extent of Disease.EOD 4 - nodes (1983-1987)} = 0
OR ({Extent of Disease.EOD 4 - extent (1983-1987)} = 9
AND {Extent of Disease.EOD 4 - nodes (1983-1987)} = 0-6,8-9
OR ({Extent of Disease.EOD 4 - extent (1983-1987)} = 1-7,9
AND {Extent of Disease.EOD 4 - nodes (1983-1987)} = 5-6)))
OR ({Race, Sex, Year Dx, Registry, County.Year of diagnosis} = '1988', '1989', '1990', '1991', '1992', '1993', '1994', '1995', '1996', '1997', '1998',
'1999', '2000', '2001', '2002', '2003'
AND {Stage.AJCC stage 3rd edition (1988-2003)} = 'Unknown')

© 2026 Memorial Sloan Kettering Cancer Center