PharmaSUG China 2019 - Paper DM-042

Medical Review Support in Clinical Trial: Method of Staging Lung Cancer Automatically According to TNM Staging System Using SAS Software

Pu Huang, Beigene, Inc

ABSTRACT

In many oncology clinical trials, cancer stage of patient is classified according to TNM Staging System. Only patients with specified cancer stage are qualified to enter the clinical trial. Cancer stage is usually judged by investigators during the screening period. The judged stage must match the results of patients' screening tumour assessment.

This paper will present how to stage lung cancer with patients' tumour assessment results using SAS Software, by applying TNM Staging System from American Joint Committee Cancer 7th edition of Cancer Staging Manual. It can support medical review by picking up the potential data issues during lung cancer staging and make sure all enrolled patients are classified into correct cancer stage.

INTRODUCTION

Cancer staging is the process of determining how much cancer is in the body and where it is located. Staging describes the severity of an individual's cancer based on the magnitude of the original (primary) tumor as well as on the extent cancer has spread in the body.

In most cases, the stage is based on four main factors:

- Location of the primary (original) tumor.
- Tumor size and extent of tumors.
- Lymph node involvement (whether or not the cancer has spread to the nearby lymph nodes).
- Presence or absence of distant metastasis (whether or not the cancer has spread to distant areas of the body).

Investigators collect data about a cancer to determine its stage. This information comes from the various tests used to identify staging in different types of cancer. In this paper, lung cancer stage is derived using SAS Software with collected information. Physical examinations and imaging tests such as CT scans, and MRI scans can show the location of the cancer, the size of the tumor, and whether the cancer has spread. Laboratory tests, pathology reports and surgical reports can also be used to collect cancer information.

TNM STAGING SYSTEM

The TNM Staging System was developed and is maintained by the AJCC and the Union for International Cancer Control (UICC). It is the most commonly used staging system by medical professionals around the world. The TNM classification system was developed as a tool for doctors to stage different types of cancer based on certain, standardized criteria.

The TNM Staging System is based on the extent of the tumor (T), the extent of spread to the lymph nodes (N), and the presence of metastasis (M). It can be separated into two steps. First, lung cancer is classified using TNM classification rules defined in Table 1. Seventh tumour, node, metastasis classification of lung cancer: January 2010.

T: Tumour	
TX	Primary tumour cannot be assessed, or tumour proven by the presence of malignant cells in sputum or bronchial washings but not visualized by imaging or bronchoscopy
T0	No evidence of primary tumour

Tis	Carcinoma in situ
T1	Tumour < 3 cm in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the lobar bronchus (i.e., not in the main bronchus)
T1a	Tumour < 2 cm in greatest dimension
T1b	Tumour > 2 cm but < 3 cm in greatest dimension
T2	Tumour > 3 cm but < 7 cm or tumour with any of the following features (T2 tumours with these features are classified T2a if < 5 cm):
	Involves main bronchus, > 2 cm distal to the carina
	Invades visceral pleura
	Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung
T2a	Tumour > 3 cm but < 5 cm in greatest dimension
T2b	Tumour > 5 cm but < 7 cm in greatest dimension
T3	Tumour > 7 cm or one that directly invades any of the following:
	Chest wall (including superior sulcus tumours), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium
	Tumour in the main bronchus < 2 cm distal to the carina but without involvement of the carina
	Associated atelectasis or obstructive pneumonitis of the entire lung
	Separate tumour nodule(s) in the same lobe
T4	Tumour of any size that invades any of the following:
	Mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina
	Separate tumour nodule(s) in a different ipsilateral lobe
N: Nodes	
NX	Regional lymph nodes cannot be assessed
N0	No regional lymph node metastasis
N1	Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension
N2	Metastasis in ipsilateral mediastinal and/or subcarinal lymph node(s)
N3	Metastasis in contralateral mediastinal, contralateral hilar, ipsilateral or contralateral scalene, or supraclavicular lymph node(s)
M: Metastases	
MX	Distant metastasis cannot be assessed
M0	No distant metastasis
M1	Distant metastasis
M1a	Separate tumour nodule(s) in a contralateral lobe
	tumour with pleural nodules or malignant pleural/ pericardial effusion
M1b	Distant metastasis

Table 1. Seventh tumour, node, metastasis classification of lung cancer: January 2010

Second, lung cancer is staged based on TNM classification results. The rules are shown in Table 2. Staging of lung cancer based on the new tumour, node, metastasis classification.

	N0	N1	N2	N3
T1a	Ⅰ A	Ⅱ A	Ⅲ A	Ⅲ B
T1b	Ⅰ A	Ⅱ A	Ⅲ A	Ⅲ B
T2a	Ⅰ B	Ⅱ A	Ⅲ A	Ⅲ B
T2b	Ⅱ A	Ⅱ B	Ⅲ A	Ⅲ B
T3	Ⅱ B	Ⅲ A	Ⅲ A	Ⅲ B
T3	Ⅱ B	Ⅲ A	Ⅲ A	Ⅲ B
T3	Ⅱ B	Ⅲ A	Ⅲ A	Ⅲ B

T4	III A	III A	III B	III B
T4	III A	III A	III B	III B
M1a	IV	IV	IV	IV
M1a	IV	IV	IV	IV
M1b	IV	IV	IV	IV

Table 2. Staging of lung cancer based on the new tumour, node, metastasis classification

TNM classification for lung cancer could be interpreted as the rules below for SAS coding. The rules are shown in Figure 1. TNM Classification Rules for Lung Cancer.

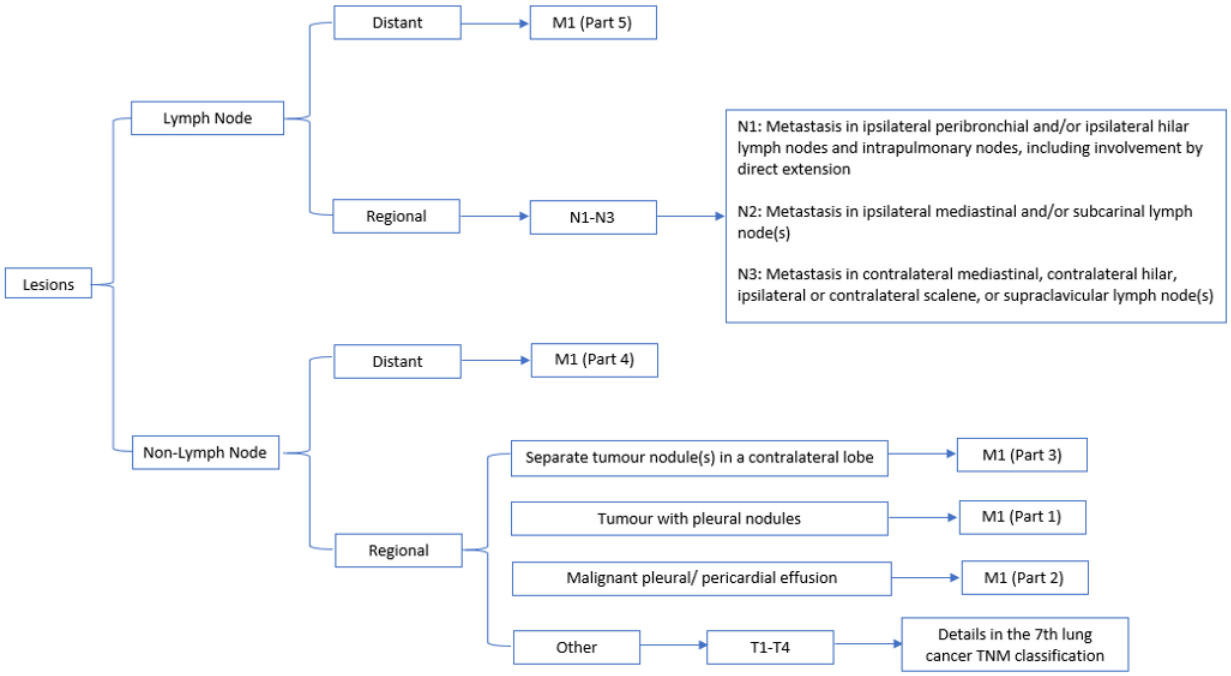

Figure 1. TNM Classification Rules for Lung Cancer

PROCESS

Medical review happens when study is going on, so the raw data in database used to stage lung cancer is always not well cleaned. If any potential data issue exists, it should be queried and corrected.

PREPARATION

Combine target and non-target lesions collected during the screening period together into one dataset. The dataset contains one record per patient per lesion. Tumour information may be collected in several variables, such as tumour location/other tumour location/tumour description etc. It would be better to derive one new variable ‘LINE’ which concatenates all tumour information variables with comma and removes all blanks for each lesion. The ‘LINE’ variable is like a pool and we can capture key words from it when classifying tumour.

SEPARATE LYMPH NODE LESIONS FROM LESIONS

If a lesion belongs to the two situations below, it could be regarded as lymph node lesion and variable lymphfl=’Y’, otherwise lymphfl=’N’.

Situation 1: Value of tumour location is lymph node.

Situation 2: Value of tumour location is not lymph node, but keyword ‘lymph’ exists in new concatenated variable ‘LINE’. Always this lesion belongs to lymph node lesion. Tumour location should be recorded as

lymph node and it is actually a potential data issue. But there are still exceptions. if ‘by’ or ‘with’ are before keyword ‘lymph’, perhaps it describes a non-lymph node lesion fused with lymph node. In this case, it should be regarded as non-lymph node lesion.

When data is well cleaned, only Situation 1 will exist for lymph node lesions.

DISTINGUISH BETWEEN REGIONAL AND DISTANT LYMPH NODE LESIONS

Select records with lymphfl=’Y’. The values of tumour location are ‘lymph node’, so we can only get the location of lymph node lesions from the newly derived variable ‘LINE’. Because the extra-thoracic and regional locations include so many anatomical terminologies, it is a good option for us to generate several location macro variables for convenient use and update.

Distant Lymph Node Lesion: Variable ‘LINE’ contains extra-thoracic location, such as liver/brain/pancreas/etc., except for scalene or supraclavicular location.

Regional Lymph Node Lesion: Lesions do not satisfy the Distant Lymph Node Lesion definition and meet one of the conditions below.

1. Variable ‘LINE’ contains thoracic location, such as lung/trachea/bronchus/heart/etc.
2. Variable ‘LINE’ contains scalene or supraclavicular location.
3. Variable ‘LINE’ contains regional lymph node group number of lung cancer, such as group one/2R/etc.

Always there are some lymph node lesions which could not be classified into either of the Distant/Regional Lymph Node Lesions above. Maybe the locations collected are not clear enough to judge which kind of lesion they belong to. If so, query needs to be raised. Also, we probably missed some location keywords when define Distant/Regional Lymph Node Lesion. We could train our SAS code by adding these locations in location macro variables to resolve this issue.

DISTINGUISH BETWEEN PRIMARY TUMOUR AND DISTANT METASTASIS FOR NON-LYMPH NODE LESIONS

Select records with lymphfl=’N’. We can get the location of lesions from tumour location variable or the newly derived variable ‘LINE’.

Non-Lymph Node Distant Metastasis: Lesions meet one of the conditions below.

1. Location of this lesion is ‘Pleura’.
2. ‘Pleural/ pericardial effusion’ is recorded in tumour location or could be found in variable ‘LINE’.
3. For lesion in lung area with lymphfl=’N’, we can derive which lung sides the lesions exist. If both left and right lungs of a patient have non-lymph node lesions, one of them should be regarded as distant metastasis.
4. For lesions with lymphfl=’N’, Variable ‘LINE’ contains extra-thoracic location.

Primary (Original) Tumour Lesion: Lesions do not satisfy the Non-Lymph Node Distant Metastasis definition and Lesions’ ‘LINE’ variable contain thoracic location, such as lung/trachea/heart/etc.

Always there are some non-lymph node lesions which could not be classified into either of the two groups above. Maybe the locations collected are not clear enough to judge which kind of lesion they belong to. If so, query needs to be raised. Also, we probably missed some location keywords when define the two groups. We could train our SAS code by adding these locations in location macro variables to resolve this issue.

METASTASIS CLASSIFICATION (M)

If any target or non-target lesion of a patent meets one of the parts below, M=’M1’ for this patient. Otherwise, M=’M0’.

M1 (Part 1) - Tumour with pleural nodules. Lesion meets Condition 1 of Non-Lymph Node Distant Metastasis definition.

M1 (Part 2) - Malignant pleural/ pericardial effusion. Lesion meets Condition 2 of Non-Lymph Node Distant Metastasis definition.

M1 (Part 3) - Separate tumour nodule(s) in a contralateral lobe. Lesion meets Condition 3 of Non-Lymph Node Distant Metastasis definition.

M1 (Part 4) - Distant metastasis of non-lymph node lesions. Lesion meets Condition 4 of Non-Lymph Node Distant Metastasis definition.

M1 (Part 5) - Distant metastasis of lymph node lesions. Lesion meets the definition of Distant Lymph Node Lesion.

NODE CLASSIFICATION (N)

Select records meeting the definition of Regional Lymph Node Lesion. Because M='M1' for patient with non-lymph node lesions in both lungs and the stage is IV, we only need to calculate variable N for the patients with M='M0'. Non-lymph node lesions of these patients exist in one-side lung. Firstly, the side of non-lymph node lesions should be captured. Then, derive N for each record based on the rules below. Choose the biggest N within a patient as the value of N for this patient.

N3 (Part 1) - Metastasis in contralateral mediastinal, contralateral hilar lymph nodes. Lesions meet one of the conditions below.

1. Variable 'LINE' clearly records that lymph nodes exist in both lungs.
2. Variable 'LINE' records that there is lymph node lesion in contralateral mediastinal, hilar or lung location. It also needs to count in the opposite regional lymph nodes recorded as group number, such as 2R/4R/etc.

N3 (Part 2) - Metastasis in ipsilateral or contralateral scalene, or supraclavicular lymph nodes. Lesions meet one of the conditions below.

1. Variable 'LINE' contains scalene or supraclavicular location.
2. Variable 'LINE' contains regional lymph node Group 1 (including 1L/1R) of lung cancer.

N2 - Metastasis in ipsilateral mediastinal and/or subcarinal lymph nodes. Lesions do not satisfy the N3 conditions and meet one of the conditions below.

1. Variable 'LINE' contains mediastinal or subcarinal location.
2. Variable 'LINE' contains regional lymph node Group 2-9 of lung cancer.

N1 - Metastasis in ipsilateral peribronchial and/or ipsilateral hilar lymph nodes and intrapulmonary nodes, including involvement by direct extension. Lesions do not satisfy the N2/N3 conditions and meet one of the conditions below.

1. Variable 'LINE' contains lung or hilar location.
2. Variable 'LINE' contains regional lymph node Group 10-14 of lung cancer.

TUMOR CLASSIFICATION (T)

Select records meeting the definition of Primary (Original) Tumour Lesion. Because M='M1' for patient with non-lymph node lesions in both lungs and the stage is IV, we only need to calculate variable T for the patients with M='M0'. Firstly, the lobes(upper/middle/lower) that non-lymph node lesions locate in should be captured. Then, derive T for each record based on the rules below. Choose the biggest T within a patient as the value of T for this patient.

T4 - Tumour of any size that invades any of the following: Mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina. Separate tumour nodule(s) in a different ipsilateral lobe. Variables T equal T4 for patients meeting one of the conditions below.

1. Variable ‘LINE’ of any non-lymph node lesion contains one of these locations: Mediastinum, heart, great vessels, trachea, recurrent laryngeal nerve, esophagus, vertebral body, carina.
2. Two or more lobes in the same lung side are affected by non-lymph node lesions.

T3 - Tumour > 7 cm or one that directly invades any of the following: Chest wall (including superior sulcus tumours), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium. Tumour in the main bronchus < 2 cm distal to the carina but without involvement of the carina. Associated atelectasis or obstructive pneumonitis of the entire lung. Separate tumour nodule(s) in the same lobe. Variables T equal T3 for patients who do not satisfy T4 condition and meet one of the conditions below.

1. Biggest diameter of non-lymph node lesions > 7 cm.
2. Variable ‘LINE’ of any non-lymph node lesion contains one of these locations: Chest wall (including superior sulcus tumours), diaphragm, phrenic nerve, mediastinal pleura, parietal pericardium.
3. Variable ‘LINE’ of any non-lymph node lesion includes the information related to ‘Tumour in the main bronchus < 2 cm distal to the carina but without involvement of the carina’.
4. Variable ‘LINE’ of any non-lymph node lesion includes the information related to ‘Associated atelectasis or obstructive pneumonitis of the entire lung’.
5. Two or more non-lymph node lesions are in the same lobe.

T2 - Tumour > 3 cm but <= 7 cm or tumour with any of the following features (T2 tumours with these features are classified T2a if < 5 cm): Involves main bronchus, > 2 cm distal to the carina; Invades visceral pleura; Associated with atelectasis or obstructive pneumonitis that extends to the hilar region but does not involve the entire lung. T2a: Tumour > 3 cm but <= 5 cm in greatest dimension. T2b: Tumour > 5 cm but <= 7 cm in greatest dimension. For patients who do not satisfy T4/T3, T equals T2 for lesions with the features above. Besides, T2 lesions are classified as T2a if <= 5cm or if the size cannot be determined and T2b if > 5 cm but <=7 cm.

T1: Tumour < 3 cm in greatest dimension, surrounded by lung or visceral pleura, without bronchoscopic evidence of invasion more proximal than the lobar bronchus (i.e., not in the main bronchus). T1a: Tumour < 2 cm in greatest dimension. T1b: Tumour > 2 cm but < 3 cm in greatest dimension. For patients who do not satisfy T4/T3, T equals T1 for lesions with the features above and does not meet T2. Besides, T1 lesions are classified as T1a if <= 2cm or if the size cannot be determined and T1b if > 2 cm but <=3 cm.

CALCULATING LUNG CANCER STAGE

Calculated lung cancer stage can be derived easily with the TNM classification above and Table 2 rules. For the patients with different collected stage and calculated stage, data checking should be done to find the reason.

At the same time, we could do more checks against RICIST guidelines. For example, when more than one measurable lesion is present at baseline, all lesions up to a maximum of five lesions total (and a maximum of two lesions per organ) representative of all involved organs should be identified as target lesions and will be recorded and measured at baseline. If number of target lesions exceeds the limit, data query should be raised.

CONCLUSION

Lots of data is collected during clinical trials and data quality is very critical. The method in this paper could support medical review by staging lung cancer automatically and improving the efficiency and accuracy of data issue check. Not only lung cancer, but also other cancer disease using TNM staging system could be staged with similar method. SAS code needs to be modified based on the corresponding classification rules.

While the method of staging lung cancer using SAS Software is developed to aid medical review in defining a patient's cancer state, it should not be considered a standalone review method. Physicians always are expected to go through the checking results and seek out newer information that might impact the diagnostic recommendations contained within a guideline.

REFERENCES

Website American Joint Committee on Cancer. 2019. “What is Cancer Staging?”. Available at: http://cancerstaging.org/references-tools/Pages/What-is-Cancer-Staging.aspx.

Journal article Saeed Mirsadraee, etc. 2012. “The 7th lung cancer TNM classification and staging system: Review of the changes and implications” *World Journal of Radiology,* 4(4):128–134. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3351680/

Journal article E.A. Eisenhauera, etc. 2009. “New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1)” *European Journal of Cancer,* 45:228–247. Available at: https://ctep.cancer.gov/protocolDevelopment/docs/recist_guideline.pdf

ACKNOWLEDGMENTS

Really appreciate the professional advice from medical monitoring members and other statistical programmers in the company. Their help makes this method more scientific and useful. Thank you very much!

CONTACT INFORMATION

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

Pu Huang
Beigene, Inc.
Email: williamhp0322@163.com
Wechat: williamsh233